Diplomarbeit

Evaluating the Impact of Human-to-Robot Communication Channels on Human-Robot Task Takeovers

ausgeführt um Zwecke der Erlangung des akademischen Grades

Diplom-Ingenieur

eingereicht an der

Technischen Universität Wien
Fakultät für Maschinenwesen und Betriebswissenschaften

unter der Leitung von:
Univ.-Prof. Dr.-Ing Sebastian Schlund
(Institut für Managementwissenschaften, Forschungsbereich: Mensch-Maschine-Interaktion)

und
Dipl.-Ing. David Kostolani
(Institut für Managementwissenschaften, Forschungsbereich: Mensch-Maschine-Interaktion)

von
Bernd Alexander Brenter BSc.

Matrikelnummer: 1908547

Wien, Februar 2025

____

Vorname Nachname

Ich habe zur Kenntnis genommen, dass ich zur Drucklegung meiner Arbeit unter der Bezeichnung

Evaluating the Impact of Human-to-Robot Communication Channels on Human-Robot Task Takeovers

nur mit Bewilligung der Prüfungskommission berechtigt bin.

Eidesstattliche Erklärung

Ich erkläre an Eides statt, dass die vorliegende Arbeit nach den anerkannten Grundsätzen für wissenschaftliche Abhandlungen von mir selbstständig erstellt wurde. Alle verwendeten Hilfsmittel, insbesondere die zugrunde gelegte Literatur, sind in dieser Arbeit genannt und aufgelistet. Die aus den Quellen wörtlich entnommenen Stellen, sind als solche kenntlich gemacht.

Das Thema dieser Arbeit wurde von mir bisher weder im In- noch Ausland einem:r Beurteiler:in zur Begutachtung in irgendeiner Form als Prüfungsarbeit vorgelegt. Diese Arbeit stimmt mit der von den Begutachter:innen beurteilten Arbeit überein. Ich nehme zur Kenntnis, dass die vorgelegte Arbeit mit geeigneten und dem derzeitigen Stand der Technik entsprechenden Mitteln (Plagiat-Erkennungssoftware) elektronisch-technisch überprüft wird. Dies stellt einerseits sicher, dass bei der Erstellung der vorgelegten Arbeit die hohen Qualitätsvorgaben im Rahmen der geltenden Regeln zur Sicherung guter wissenschaftlicher Praxis "Code of Conduct" an der TU Wien eingehalten wurden. Zum anderen werden durch einen Abgleich mit anderen studentischen Abschlussarbeiten Verletzungen meines persönlichen Urheberrechts vermieden

Wien, Februar 2025

Vorname Nachname

Abstract

The division of labor in human-robot teams is a major research topic in Human-Robot Collaboration (HRC). Effective collaboration requires seamless task transitions, especially when a human takes over a task from a robot. However, current industry standards are constrained by communication channels, like buttons, and complex recovery procedures. Even with recent advancements in adaptive and communication-free task-allocation methods, pre-defined task boundaries still limit when and how a human can take over a robot's task. To overcome these limitations, two key components are: (1) a task model to recover from interruptions, and (2) methods for communicating them.

To investigate how interruptions can be handled and how different Human to Robot Communication (HTRC) channels affect task takeovers, this thesis employs a mixed research approach, combining a literature review with a user study. The review explores the fundamentals of human-robot interactions, communication, and current task allocation techniques, and reveals a lack of research on task takeovers, in particular on how to handle and communicate them. To quantitatively evaluate the effects of different HTRC channels on team dynamics, a prototype setup was developed, and a user study was conducted using three different communication channels: (a) *baseline*, mediated via the push of a button in a Graphical User Interface (GUI), (b) *explicit*, facilitated by physically touching the robot, and (c) *implicit*, based on human actions.

The findings suggest noteworthy differences in how HTRC channels affect the perceived fluency, likability, and the human-robot bond during task takeovers. In contrast, the results indicate no effect between HTRC channels for the trust in the robot, robot traits like perceived intelligence or commitment to the task, and the goal alignment between human and robot. Therefore, the results have implications for the design of future collaborative robots, as they reveal the importance of employing communication channels that minimize disruption to the human's workflow while fostering a team-dynamic comparable to human teams.

Kurzfassung

Die Arbeitsteilung in Mensch-Roboter-Teams ist ein zentrales Forschungsthema in der Mensch-Roboter-Kollaboration. Eine effektive Zusammenarbeit erfordert nahtlose Übergänge zwischen Aufgaben, insbesondere wenn ein Mensch die Aufgabe eines Roboters übernimmt. Aktuelle Industriestandards sind jedoch durch Kommunikationskanäle, wie physische Tasten und komplexe Prozeduren zur Wiederaufnahme des Betriebs eingeschränkt. Selbst mit den neuesten Algorithmen zur adaptiven Arbeitsteilung schränken vordefinierte Grenzen immer noch ein, wann und wie ein Mensch Aufgaben vom Roboter übernehmen kann. Um diese Einschränkungen zu überwinden, sind zwei Komponenten essenziell. (1) Eine interne Repräsentation, die es dem Roboter ermöglicht sich von Unterbrechungen zu erholen. (2) Methoden zur Kommunikation von Unterbrechungen.

Um zu untersuchen, wie Unterbrechungen bewältigt werden und wie sich verschiedene Kommunikationskanäle auf das Team auswirken, kombiniert diese Arbeit eine Literaturrecherche mit einer Studie. Die Recherche analysiert sowohl die Grundlagen der Mensch-Roboter-Interaktion, als auch der Kommunikation und dynamische Arbeitsteilung. Dabei identifiziert die Arbeit eine Forschungslücke bei der Handhabung und Kommunikation von Unterbrechungen. Zur Schließung dieser Lücke und zur Bewertung des Einflusses verschiedener Kommunikationskanäle auf die Teamdynamik wurde eine Studie mit drei Kanälen durchgeführt: (a) eine Basisbedingung, bei der Unterbrechungen durch eine Taste ausgelöst werden, (b) eine haptische-explizite Methode, bei der der Roboter physisch berührt wird, und (c) eine implizite Methode, die sich aus menschlichen Aktionen ableitet.

Die Ergebnisse der Untersuchung zeigen, dass der Kommunikationskanal (a) den Aufgabenfluss, (b) die Sympathie sowie (c) die Mensch-Roboter-Bindung erheblich beeinflussen kann. Hingegen wirkt er sich weder auf (a) das Vertrauen in den Roboter, (b) dessen wahrgenommene Intelligenz noch auf (c) die Zielausrichtung des Teams aus. Diese Erkenntnisse sind entscheidend für das Design zukünftiger kollaborativer Roboter, da sie die Wahl von Kommunikationskanäle betonen, die eine menschenähnliche Teamdynamik fördern.

Contents

Acronyms

AMM Agent Markov Model
ANN Artificial Neural Network
AOG AND/OR Graph
AR Augmented Reality
ASR Automatic Speech Recognition

BCI Brain Computer Interface
BT Behavior Tree

CAD Computer Aided Design
CNN Convolutional Neural Network

DAG Directed Acyclic Graph
DNN Deep Neural Network
DOF Degrees of Freedom
DTW Dynamic Time Warping

EEG electroencephalograph
EOL end-of-life

GMM Gaussian Mixture Modeling
GMR Gaussian Mixture Regression
GUI Graphical User Interface

HCI Human-Computer Interaction
HMI Human-Machine Interaction
HMM Hidden Markov Model
HMS Human-Machine System
HRC Human-Robot Collaboration
HRI Human-Robot Interaction
HTN Hierarchical Task Network
HTRC Human to Robot Communication

IMU Inertial Measurment Unit

KNN K-Nearest Neighbours

LSTM Long Short-Term Memory

MABA-MABA Men Are Better At - Machines Are Better At

MDP Markov Decision Process

ML Machine Learning

MMDP Multi-Agent Markov Decision Process

NPC Non-Player Character

NUI Natural User Interface

POMDP Partially Observable Markov Decision Process

RM-ANOVA Repeated Measurements Analysis of Variance

ROS Robot Operating System

RTHC Robot to Human Communication

SoM State of Mind

SSVEP steady-state visual evoked potential

SVM Support Vector Machine

ToM Theory of Mind

UI User Interface

VUI Voice User Interface

1 Introduction

1.1 Motivation and Problem Statement

Human-Robot Collaboration (HRC) is a field of research promising to democratize robotics and lower entry barriers by offering concepts enabling robots to flexibly, safely, and efficiently work alongside humans. To enhance human-robot flexibility, dynamic human-aware task-allocation methods [1]–[3], enabling robots to dynamically adapt to actions performed by a human, are currently being investigated in favor of traditional static task-assignment approaches such as MABA-MABA (Men Are Better At - Machines Are Better At) [4].

The general direction of research is going toward machine learning-enabled methods to handle offline as well as online task assignments. Zhang et al. [2] proposed an approach utilizing fusion-based spiking neural networks to process signals generated from robot poses, human behavior, and product states to determine the optimal moment for robotic assistance. Ramachandruni et al. [3] and Cheng et al. [1] both introduced adaptive user-aware collaboration frameworks based on HTN (Hierarchical Task Network), an artificial intelligence technique comprised of an initial state, a task network, and domain knowledge [5]. Unlike traditional leader-follower models, their methods enable a robot to adapt to human actions in real time without predefined communication modalities. These methods excel at handling task-level transitions, like dynamically adapting the task sequence if the human switched tasks. Nevertheless, they are limited by clear task boundaries and do not consider interruptions to an ongoing task, as they typically assume uninterrupted processes and optimal human behavior, making human intervention basically only possible at said boundaries or via interrupting the robot's programming. However, in addition to knowing when and what to do, effective collaboration relies on smooth transitions not only between but also within tasks. Particularly, when it comes to the handling of task takeovers, where a human momentarily steps in and takes over a task originally assigned to the robot.

Moreover, human-in-the-loop decision-making introduces uncertainty [6] and inappropriate robot behavior can disrupt the workflow and lead to dissatisfaction among human operators [2]. To increase certainty, communication between humans and robots is deemed vital for decision-making processes in HRI [7]. However, considerations in regard to communication are generally excluded from task allocation frameworks, as they aim to be communication-free [1], [3]. The need for communication within human-robot teams has also been pointed out by others such as Salehzadeh et al. [8] and Hellstrom et al. [9], who both argue that communication can lower uncertainty for real-world robot applications.

To address the identified gap, this thesis explores existing literature on task allocation methods and how they integrate Human to Robot Communication (HTRC). In addition, this thesis utilizes a user study to evaluate the effects of HTRC on human-robot collaboration during task takeovers. The results provide new insights into how the utilized communication channel affects the perceived team dynamics such as team fluency and trust in the robot.

1.2 Research Question and Research Objective

Based on the discussed problem landscape, it becomes clear that one key obstacle to fluent task takeovers is the lack of effective communication strategies. Consequently, the research objective of this thesis is to explore the impact of various channels of communication. Therefore, it is guided by the following research questions:

Q1.) Which communication theories can be used to model a takeover request?

Q2.) Which channels of communication can be utilized to communicate a task-takeover request?

Q3.) What are current task allocation techniques, and how do they integrate task-takeovers and communication?

Q4.) To what extent do different HTRC channels impact team dynamics, such as perceived team fluency and trust, during task-takeovers?

1.3 Methodology and Expected Outcome

As a first step, a semi-structured literature review explores the theoretical foundations of human-machine interaction as well as human-robot collaboration. Academic search engines are utilized for building a foundation to review and answer *Q1*, *Q2*, and *Q3*. In order to find relevant state-of-the-art literature, search terms such as [[”robot-to-human” OR [”human” AND ”robot”]] AND ”communication”] as well as [”human” AND ”robot” AND ”task” AND [”allocation” OR ”request” or ”takeover”]] are used. Based on the initial screening of available literature, forward and backward searches are applied to find additional sources.

For answering *Q4*, a prototype setup exploring different communication modalities is developed, and a user study is conducted to evaluate the impact of the individual modalities on human-robot task takeovers. The prototype setup includes a cobot with 6-DOF (Degrees of Freedom), a stereo-vision camera, and is built using ROS (Robot Operating System). The user study primarily utilizes quantitative evaluation metrics for directly and indirectly assessing perceived team fluency and trust in the robot. The study is concluded by performing a statistical analysis of the gathered data and a discussion of the results. By evaluating different channels in the context of task takeovers, this thesis tries to identify channels of communication that balance intuitive understandability with technological feasibility. It thereby contributes to the understanding of how HRC can be enhanced and potentially guides the design of future robots and their interaction protocols.

2 Research Background

The following chapter first lays out the general objectives and research directions of Human-Machine Interaction (HMI) and Human-Robot Interaction (HRI) (section 2.1). Followed by a thorough discussion of how the terms human (section 2.2.1), robot (section 2.2.2), and interaction (section 2.2.3) are conceptualized in previous literature. The subsequent section dives into Human-Robot Collaboration (HRC) (section 2.3). The final part of this chapter examines selected communication theories and their relevancy for HRI (section 2.4) as well as important concepts of communication such as the communicated message (section 2.5) and the utilized communication channel (section 2.6).

2.1 Human-Robot Interaction

Human-Machine Interaction (HMI) is an interdisciplinary field of research, at its core focused on the development, design, and evaluation of interactions between humans and machines. Even though humanity has been using machines for millennia, the investigation of design principles such as comfort, safety, performance, and aesthetics have only been of interest since the mid-20$^{\text{th}}$ century [10]. Before that, machines were hardly ever adjusted to humans, instead, people had to adapt to machines in order to operate them and avoid being injured or even killed in the process [11]. Human-related factors like cognition, emotion, society, and physiology, are even more recent topics with the first dedicated journal, called "International Journal of Man-machine Studies", published in 1969 [10]. For the last four decades, most newly engineered human-machine interactions have been software-based [10]. Along with this dominance of Human-Computer Interaction (HCI) within research, human cognition was at the center of attention [10]. Unlike the 1980s, when computers were considered specialized machines, today they are commonly seen as mediators, enabling human-machine interactions to the point where they often go unnoticed [10].

Human-Robot Interaction (HRI) is a cross-disciplinary sub-area of research within Human-Machine Interaction (HMI), focused on various aspects such as robot behavior, robot autonomy, social robotics, robot ethics, trust in robots, and human factors associated with robot design [12]. By understanding the dynamics of human-robot interactions, researchers aim to create robots that can be seamlessly integrated into various domains such as manufacturing, healthcare, entertainment, education, and everyday life [12]. The major research communities within HRI each contribute different yet relevant expertise. Roboticists, for example, provide the physical hardware, controlling algorithms, and robot performance metrics [13]. In contrast to that, HCI researchers contribute performance metrics in regard to the interaction, as well as methodologies and guidelines for performing usability studies [13]. Cognitive scientists, on the other hand, provide human performance measurement metrics and methods to model the human agent in HRI [13]. Despite major advancements in the last decades, robot applications are still not capable of interacting on human-like levels, hence most literature sees a successful human-robot interaction in mutual adaptation and successfully achieving a shared goal [8].

Besides the many research-related facets of HRI, Burke et al. [13] also stress the importance of looking at human-robot relations from multiple viewpoints. The three major relationship taxonomies mentioned are:

(1) **Numeric relation**, depicting the ratio between humans and robots

Humans	**Robots**
One person	One robot
One person	Many robots
Many people	One robot
Teams of people	Teams of robots

Table 2.1: Numeric relation = ratio of humans-to-robots (adapted from [13])

(2) **Spatial relation**, enabling a view on who is located where. For example: (a) *Remote* - the physical workspaces are separated. (b) *Beside* - human and robot work as peers in one workspace. (c) *Robo-immersed* - the human teleoperates the robot and sees the world through the robot's eyes. (d) *Inside* - the human is designing the robot and, therefore, conceptionally within the robot (see Table 2.2).

(3) **Authorial relation**, expressing how control is allocated in the interaction. For

Role	Human's POV	Spatial Relationship
Commander	God's-eye	Remote
Peer	Bystander	Beside
Teleoperator	Robot's eye	"robo-immersion"
Developer	Homunculus	Inside

Table 2.2: Spatial relation = who is located where (adapted from [13])

instance (see Table 2.3): (a) Human as *Supervisor*, directing what the robot does in a grander scheme of things. (b) Human as *Operator*, controlling how the robot does a specific task. (c) Human as *Peer*, directly working with the robot. (d) Human as *Bystander*, only sharing the environment without any direct interaction. [13]

Authority Relationship	Function	Context Required
Supervisor	Commands "what"	Tactical situation
Operator	Commands "how"	Detailed perception
Peer	Cross-cueing	Shared environment, functions
Bystander	Interacts	Shares environment

Table 2.3: Authorial relation = who is in control (adapted from [13])

2.2 Defining Human, Robot, and Interaction in HRI

2.2.1 The Human

When considering the human element, research has found numerous ways of characterization. Two often cited models in defining humans for HMI are based on early HCI research by Kantowitz and Sorkin [14] as well as Card et al. [15]. The "human processor" by Card et al. [15] depicts the cognitive functions of the human mind using an analogy of a computer system. They integrate components such as short-term memory, long-term storage, and processing speeds to explain how humans perceive, process, and respond to information [15]. Kantowitz and Sorkin's model [14], on the other hand, depicts the human already integrated into a Human-Machine System (HMS) (see figure 2.1). In their model, the human is represented by a brain, sensors, and responders, and the machine is depicted via its state, controls, and displays [14]. Based on their model, interaction takes place at the interface between human and machine.

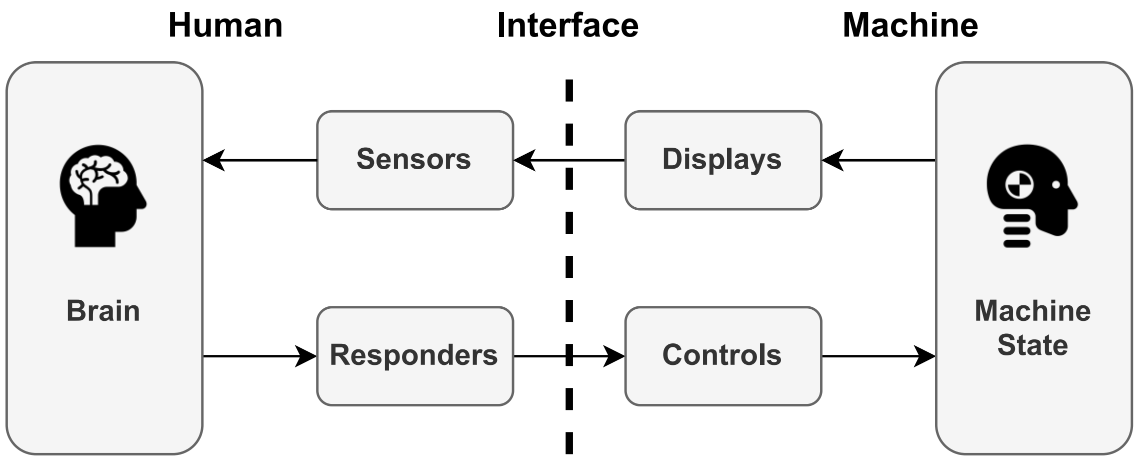


Figure 2.1: Human Machine Interaction Model (based on [14])

2.2.2 The Robot (Machine)

Similarly to the term “human”, research has not agreed on a single definition for what to consider as “machine”. Usually, when talking about a machine in HMI, a man-made technological artifact, for example, a car or a robot, but most commonly, a computer is meant [16]. In contrast to HMI’s broader field of interest, HRI is only studying interactions with robots. The fictional word *“robot”* was first introduced in 1920 by Czech author Karel Capek in his book *“Rossum’s Universal Robot (R.U.R.)”* [17], portraying a society with automated workers and class conflicts. Since then it has been reused by many science fiction authors and inspired research, leading to the first industrial robot *“Unimate”* by George Devol and Joseph Engelberger in 1961 [18], and ultimately to the vast variety of modern robotics we know today, such as industrial, service, social, humanoid, medical, and collaborative robots, depicted in figure 2.2. Among these, social robots are the most researched when it comes to interactions with humans [19].

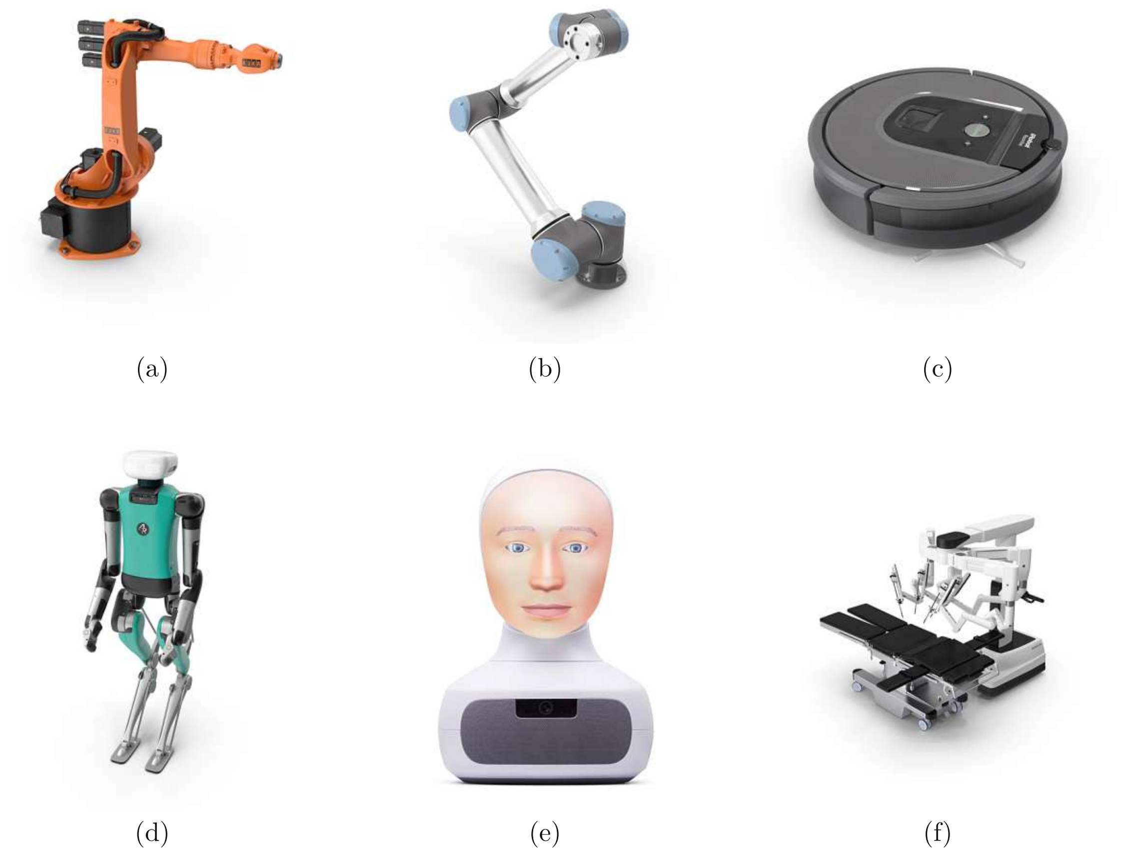


Figure 2.2: (a) Industrial Robot by Kuka[1], (b) Collaborative Robot by Universal Robots[2], (c) Robot Vacuum by iRobot[3], (d) Humanoid Robot by Agility Robotics[4], (e) Social Robot by Furhat Robotics[5], (f) Surgical Robot by Intuitive Surgical[6]

2.2.3 The Interaction

Despite being the core focus of HMI and HRI respectively, the meaning of the term "interaction" is not characterized intuitively [20]. To some degree, the difficulty of defining interaction seems to stem from HMI being an interdisciplinary endeavor, as it includes research fields such as informatics, computer science, sociology, game theory,

[1] www.pixelsquid.com/png/kuka-kr-16-3-6-axis-robot-arm-3301356004083504972
[2] www.pixelsquid.com/png/universal-robots-ur10e-2656715789318493388
[3] www.pixelsquid.com/png/irobot-roomba960-robotic-vacuum-cleaner-2535027442859382238
[4] www.pixelsquid.com/png/digit-robot-3254797900609230193
[5] www.esa.int/ESA_Multimedia/Images/2019/04/Furhat_robot
[6] www.pixelsquid.com/png/surgical-robotic-system-da-vinci-si-with-operating-table-2839187826094905088

psychology, philosophy, cognitive science, media studies, and communication science [16]. Each discipline contributes its own unique perspective, making a singular, universally accepted definition not only challenging but also potentially restrictive to some degree [16]. In addition, its vagueness becomes even more obvious when one tries to interpret the commonly used definitions published in dictionaries. The Collins-Dictionary, for example, lists interaction as (1) "a mutual or reciprocal action or influence" [1], but also as (2) "the transfer of energy between elementary particles, between a particle and a field, or between fields"[2], which relates to physical phenomena such as electromagnetic-, strong-, weak-, gravitational-interaction. The Cambridge-Dictionary, on the other hand, defines interaction as (3) "an occasion when two or more people or things communicate with or react to each other" [3]. Taking a closer look at these three characterizations, the 2^{nd} definition can be discarded for evident reasons, as it constitutes several entirely unrelated physical phenomena.

Utilizing a combination of the remaining two definitions, [16] suggests to interpret "interaction" as a four-dimensional concept, respectively: (1) *subjects*, (2) *mode*, (3) *purpose*, and (4) *context*. The first dimension, *subjects*, is derived from the question of who is involved in the interaction [16]. The second dimension, *mode*, refers to the method or way in which these subjects interact with each other [16]. The third dimension, *purpose*, is concerned with the reasons behind the interaction [16]. And finally, the fourth dimension, *context*, examines the environment or circumstances in which the interaction takes place [16].

Along the line of these four dimensions, literature on defining "interaction" often also discusses it in relation to the terms (1) "interactivity", (2) "interactability", and (3) "interactiveness" [20]. (1) *Interactivity* is commonly used synonymously with interaction but also to describe ongoing interaction [20]. (2) *Interactability*, on the other hand, is denoted as an artifact's or system's capability to be interacted with [20]. Last but not least, (3) *Interactiveness* is often described as an artifact's or system's way of how it engages people to interact with it, for example by lure, insistence, or necessity [20].

[1] www.collinsdictionary.com/de/worterbuch/englisch/interaction
[2] www.collinsdictionary.com/de/worterbuch/englisch/interaction
[3] www.dictionary.cambridge.org/de/worterbuch/englisch/interaction

HRI specific Conceptualization of Interaction

Besides the just discussed general HMI approach of characterizing interaction, previous HRI research has introduced four distinguishable ways of conceptualizing interaction:

(1) **Sending and Receiving Signals**: Commonly seen as the simplest way of characterization, sending and receiving of signals or cues, views interaction as a series of turn-taking with disentangled discrete events [21]. While signals are by design informative, cues are often not explicitly meant to be [21]. In addition to distinguishing between cues and signals, literature suggests the use of taxonomies differentiating on a human-and-artificial likeness spectrum [22].

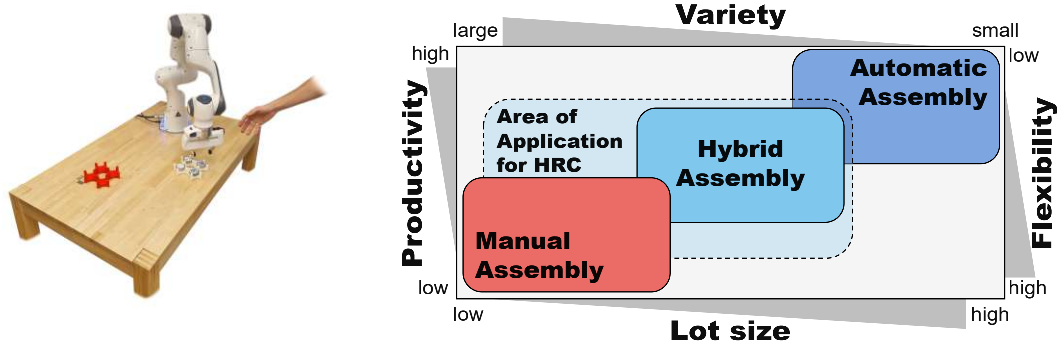
(2) **Communicative Action**: Communicative action, conceptualizes interaction as chained events of communication where each interactor has a State of Mind (SoM) and builds a mental model of their counterpart's SoM [9]. Interaction can then be seen as the intention to influence these mental models by increasing the other's knowledge of the own SoM [23]. Following this definition of interaction by Hellström and Bensch [23], offers the benefit of a turn-taking algorithmic approach to interaction. Another very important concept in this regard is the Theory of Mind (ToM) [24], first introduced by Premack and Woodruff in 1978 [25]. ToM is seen as the ability to reason about, try to explain, and predict actions of intelligent agents, this includes seeing oneself and others as intentional entities [24].

(3) **Joint Action**: Expanding on the turn-taking concept of communicative action, joint action incorporates continuous temporal and spatial coordination as important aspects of co-performing a task [26]. Following this characterization of interaction, concepts such as Theory of Mind provide tool sets to view both interactors as actively involved in constantly sharing verbal as well as nonverbal information necessary for co-constructing meaning and coordinating collective actions [26].

(4) **Dynamic System**: The characterization as a dynamic system combines ideas from the other three into a continuous multimodal self-organizing process of coordination between agents [27]. In contrast to simpler turn-taking concepts of interaction, the meaning of individual interactions can no longer be derived without the knowledge of the broader context in which these interactions occur [21].

2.3 Human-Robot Collaboration

Human-Robot Collaboration (HRC) is a sub-field of research within HRI focusing on the study of cooperation, communication, and the division of labor between humans and robots, primarily within manufacturing and assembly-related, tabletop scenarios where humans and collaborative robots work in close proximity with shared workspaces and shared tasks (see figure 2.3a) [28], [29]. With growing product individualization [30] and a shortage of skilled labor [31], real-world applications of HRC are becoming increasingly sought after due to increasing complexity of modern production and assembly [30].



(a) (b)

Figure 2.3: (a) Collaborative Robot in Tabletop Scenario, (b) HRC - Area of Application in Assembly (based on [32]),

The term cobot, short for collaborative robot, was first introduced by Colgate et al. [33] in 1996 when they were tasked to find ways of making robots adequately safe to work alongside humans without the need for safety fences. Since then cobots have promised to democratize the use of robotics and lower entry barriers by offering affordability, flexibility, safety, productivity, and ease of use [30]. The vital increase in safety necessary for enabling cobots to work close to humans is achieved by their lightweight design in combination with collision sensing technologies such as torque monitoring or collision-sensitive robot enclosures [28]. In addition to being flexibly relocated as they are lighter than traditional industrial robots, they are flexible in regard to quick changes in task assignments, due to easy programmability even by non-expert users [28]. The proclaimed productivity increases are primarily attributed to the combination of human reasoning and decision-making as well as robot precision, repeatability, and durability [28]. Cobots can, therefore, broaden the scope of robotic automation to include smaller lot sizes at higher productivity and flexibility paired with larger variety (see figure 2.3b) [32].

Previous literature introduced several ways to classify human-robot collaboration. In addition to the already discussed (see section 2.1) numeric, spatial, and authorial relation introduced by Burke et al. [13], other scholars such as Bauer et al. [34] or Müller et al. [35] also distinguish between temporal relations and others like El Zaatari et al. [36] based on task-related interdependencies. Although the classifications of [34] and [36] are further discussed in section 2.3.1, it should be mentioned that there are several others very similar to their approaches.

Another crucial aspect of HRC is task allocation, which focuses on the assignment of sub-tasks between human and cobot. Generally speaking, task assignments can be pre-planned and static or dynamically allocated during runtime [37]. Within task allocation, the ability to handle task-takeovers and to transfer tasks between agents, outlined in section 2.3.4 and section 3.2, plays a vital role towards increased flexibility in human-robot-teams.

2.3.1 Defining Collaboration

Similar to the definition of interaction (see section 2.2.3), defining collaboration can be done in many ways. When one only considers dyadic HRC, often cited approaches are similar to Bauer et al.'s [34] classification based on spatial and temporal relation and El Zaatari et al.'s [36] categorization based on task dependencies and process intersections between cobot and operator.

Degrees of Collaboration based on Spatial and Temporal Separation

Based on Bauer and colleagues' distinction [34], the degree of HRC can be classified as either with or without temporal and spatial separation, their resulting types of collaboration are shown in figure 2.4 and best described as follows:

(1) **Cell**: The Cell is used to depict the case where no collaboration is possible, as human and robot are physically separated, e.g. via a fence, and work entirely disconnected from each other [34].

(2) **Coexistence**: Coexistence, is similar to a cell in regards to spatial and temporal separation as both still work independently of each other [34]. However, in contrast to a cell, the robot is no longer encapsulated in a protective fence but is equipped with safety concepts to ensure that it does not injure the human when a collision occurs [35].

(3) **Synchronized**: In synchronized collaboration, human and robot are no longer spatially separated as they now share a single workspace [34]. However, they are still temporally separated and take turns working on the same workpiece [34]. Usually, each agent has to accomplish a different set of interrelated sub-tasks depending on the prior output of tasks performed by the other agent [35].

(4) **Cooperation**: In cooperative settings, human and robot are not temporally nor spatially separated as they now simultaneously work in the same workspace, tough on independent tasks with different components [34]. Therefore, the cobot must be equipped with some sort of spatial and task awareness to ensure collision-free fluent teamwork [35].

(5) **Collaboration**: Collaborative settings are the pinnacle of complexity within HRC. There is no temporal or spatial separation, as human and robot interactively work in a shared workspace, on a shared workpiece, towards a shared goal [34]. Their individual actions are only possible through the other's simultaneous actions [36]. To accomplish such a unison, the cobot must be able to understand the task requirements and comprehend the intentions of his human colleague [36].

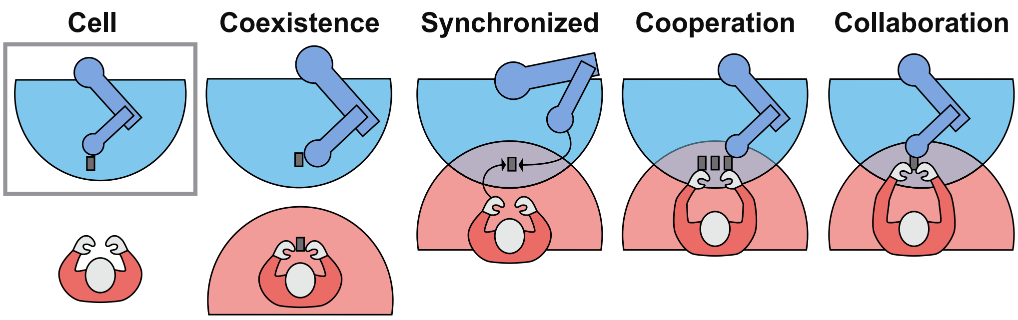


Figure 2.4: Degrees of Human Robot Collaboration (based on [34])

Degrees of Collaboration based on Task Interdependencies

Although following an approach, such as Bauer et al.'s [34], to distinguish between different degrees of HRC allows for a better understanding of especially safety-related requirements, their distinction does not explicitly take into account if and how tasks and processes assigned to cobot or human depend on each other. In contrast to that, Zaatari et al. [36] classify as either with or without dependencies and shared workpieces. Their resulting classification, shown in figure 2.5, distinguishes four degrees of collaboration:

(1) **Independent**: In independent HRC, which is comparable to Bauer et al.'s [34] classification of coexistence, human and cobot independently work on their individual workpieces [36]. Safety is achieved by utilizing the cobot's internal safety features such as torque monitoring [36].

(2) **Simultaneous**: HRC can be classified as simultaneous if cobot and human work on the same workpiece, yet on separate processes [36]. Consequently, their main dependency is the shared workpiece, but they are neither task nor time-dependent from each other [36]. Linking it to Bauer et al. [34], simultaneous HRC would roughly translate to *cooperation*. Therefore, the cobot must also be equipped with spatial and task-awareness techniques.

(3) **Sequential**: In sequential HRC, human and cobot still perform individual processes, but time is introduced as the primary dependency between them [36]. The respective processes have to be executed in sequence and outputs from one agent can be seen as inputs for the other's subsequent process [36]. Thus, depending on the overall process structure, sequential HRC is comparable to Bauer et al.'s [34] classification of either synchronized or cooperative HRC.

(4) **Supportive**: Similar to Bauer et al.'s [34] classification of collaboration, supportive HRC requires human and cobot to work on a single process in a shared workspace, with a shared workpiece, without temporal or spatial separation [36]. Consequently, the main dependencies between both agents are the joint actions toward a common goal [36]. Therefore, the cobot must again be capable of understanding human intentions and task requirements [36].

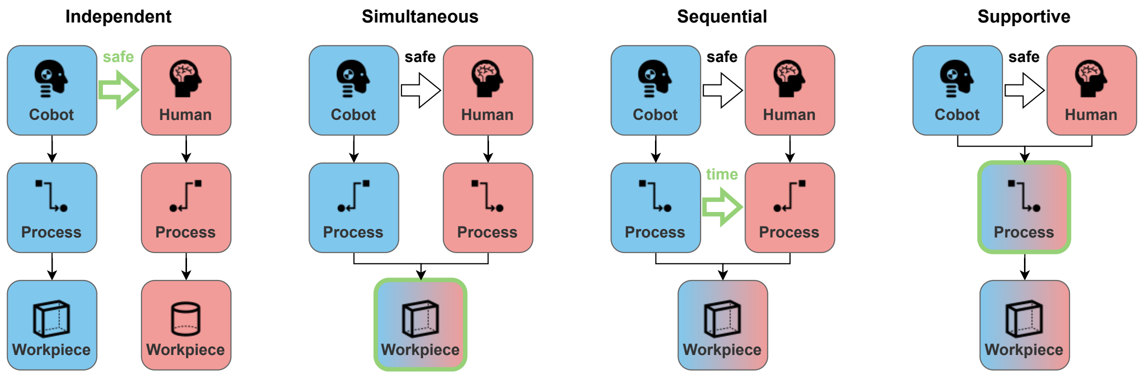


Figure 2.5: Degrees of Human Robot Collaboration (based on [36])

2.3.2 Task allocation in HRC

In Human-Robot Collaboration, both humans and robots need to be aware of their own as well as their teammate's current and future actions - this problem is commonly referred to as task allocation [38]. In addition to the consideration of who does what and in which order, it is important to assess when the allocation of tasks takes place [7] and who is given the authority to allocate (see section 2.3.3). Typically, a given overall process sequence contains information in regard to the required execution order which can either be fully sequential or allow parallel execution of sub-tasks [32]. Additionally, to maximize objectives such as time, quality, ergonomics, and flexibility [32], conflicts like collisions, obstructions, or reaching for the same object need to be avoided in order to exploit all advantages and the full potential of human-robot teams [38]. Commonly, these human and robot characteristics are used to formulate some sort of optimization problem, where a cost function encoded with all crucial aspects has to be minimized.

Previous literature reports two major approaches each having two subcategories, namely (1) *static/offline* allocation with (a) *suitability assessment* or (b) *simulation-supported* allocation, and (2) *dynamic/online* allocation, as either (a) *reactive* or (b) *proactive* (see figure 2.6) [32].

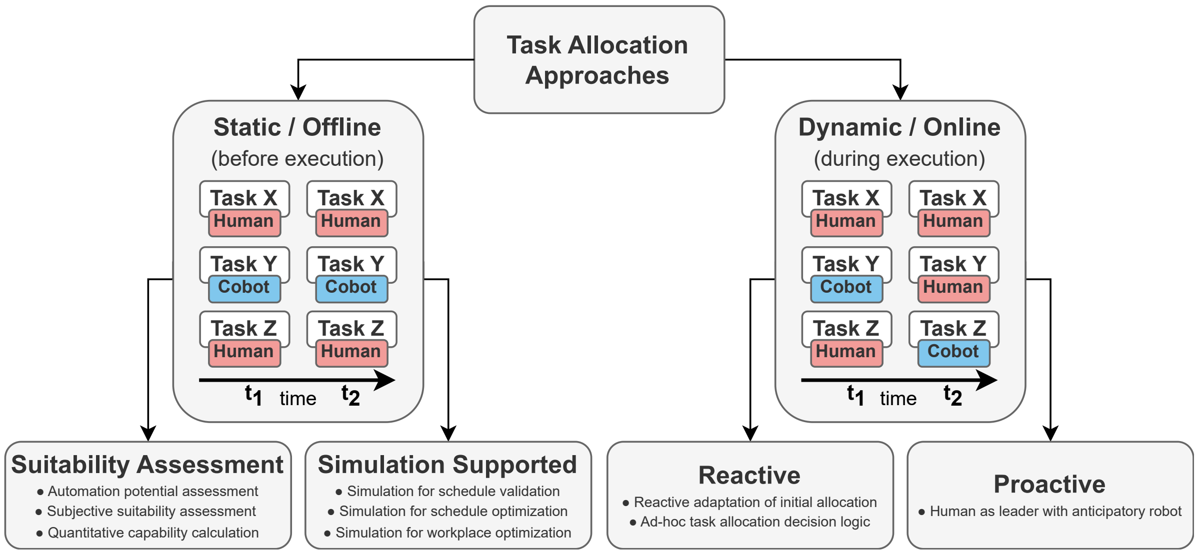


Figure 2.6: Categorization of Task Allocation (based on [32])

Static (Offline) Task Allocation

The static view regards all design choices as pre-plannable [37]. Hence, task assignment is performed prior to task execution, making it unable to cope with unpredictable circumstances as it cannot be changed at runtime and is therefore static over time [37]. In general, task assignment prior to execution can be further distinguished by two subcategories:

(a) **Suitability Assessment**: Allocation based on suitability assessments, assigns tasks to humans and robots based on individual suitability given a task's requirements [32]. Additionally, a reasonability check can be applied to assess the overall appropriateness of the allocation [32]. In principle, the most straightforward approaches are very similar to Fitts' [4] MABA-MABA (Men Are Better At - Machines Are Better At) approach from the 1950s, as they simply assign sub-tasks based on automation potential and cost to automate [32]. Hence, the human is often forced to do leftover tasks that can not be handled by the cobot or are to expensive to automate [37]. To overcome the inherent subjectivity associated with these MABA-MABA-based approaches, a combination of criteria catalogs and quantitative rating systems have been proposed, for example, by Schmidbauer et al. [39], Ranz et al. [40], or Müller et al. [35]. However, even these more analytical approaches are not further validated prior to execution [32].

(b) **Simulation Supported**: In contrast to that simulation-supported, capability-based task allocation methods try to validate their assignment by running simulations prior to task execution [32]. Therefore, these methods can evaluate and optimize multiple task assignments based on the interaction dynamics of humans and robots [41]. Additionally, these approaches can be used to optimize the physical workplace layout for a given process [32].

However, even though these static allocation methods, can reduce some human-related errors and decrease the cognitive load while executing a task, research criticizes and questions the static view for driving a decline in human skills, knowledge, and agency [42], while also causing distrust in automation, decreased situational awareness, automation-induced errors, and ultimately job dissatisfaction [37].

Dynamic (Online) Task Allocation

In order to tackle the limitations and challenges of static task allocation, research has been investigating dynamic (online) allocation of tasks [37]. At its core, dynamic allocation tries to use performance metrics, human attributes, and system conditions to dynamically create contingency plans for emerging or unpredictable circumstances [37]. In accordance with [32], dynamic task allocation can be categorized into two subcategories:

(a) **Reactive / Ad-hoc Allocation**: Reactive approaches either involve (i) a pre-computed plan of tasks already assigned to the human or the robot that gets monitored and adjusted on an individual task level or by invoking a re-planing of all pending tasks, or (ii) an online ad-hoc decision logic assigning each task based on e.g. resource availability, resource capability, or a balanced workload to reduce to the potential of human fatigue [32]. However, it's important to note that these sub-approaches are usually blended together and overlap, especially when an ad-hoc logic is used to reassign tasks in response to deviations from the original plan [43].

(b) **Proactive Allocation**: Proactive approaches, on the other hand, are inspired by human teamwork, where each individual can anticipate and proactively support the actions of their teammate by recognizing the need for assistance before it is even requested [44]. Therefore, proactive allocation empowers humans to be the team leader, as the system proactively determines how to assist, anticipating the human's needs and taking the initiative to perform preparatory work in parallel [32].

2.3.3 Authority and Agency in HRC

Within automation, in general, and HRC in particular, the authority to allocate functions within the team can be distinguished into three types: In (a) *Adaptable Allocation*, also called user-driven allocation, only the human is given the authority to adapt the task distribution [37]. (b) *Adaptive Allocation*, also known as system-driven allocation, if the authority to adapt task assignments belongs to the system [37]. Last but not least, (c) *Hybrid Allocation* when human and cobot share the authority to allocate [37].

Comparing these types of authority coupled with human agency within human-robot teams, previous empirical work remains inconclusive, as reported results paint a spectrum of preferences based on multiple criteria. On the one hand, the findings by Gombolay et al. [42] indicate that humans prefer adaptive system-driven allocation merely due to a higher degree of efficiency. Schulz et al. [45], however, suggest it may be preferable to only use an interaction style that allows for robot-led interactions when dealing with higher cognitive load actions and shift to human-led interactions for joint actions. In contrast to that, the conclusions by Tausch et al. [7], imply that humans prefer to have agency over the allocation of tasks. It is also worth mentioning, that while dynamic techniques have been found to increase acceptance, the level of trust, and situational awareness of the human worker when compared to static approaches [46], they can also distill distrust in automation when the system is perceived as unreliable [37].

2.3.4 Task Takeovers

Lending from other collaborative human-machine systems like autonomous driving, the concept of task takeovers refers to the dynamic transfer of control between agents, e.g. between humans and vehicles [47]. Depending on the specific situation and task requirements, this transfer can occur in both directions, allowing humans to hand over or request control from the machine and vice versa [47]. However, to the best of the author's knowledge, there are little to no publications in the HRC domain on the concept of mid-task interruptions and task takeovers, respectively. Even though research in this area could especially be of interest to human-robot teams in collaborative scenarios, where human-in-the-loop decision-making introduces uncertainty [6] and inappropriate robot behavior can disrupt the workflow and lead to dissatisfaction among human operators [2]. Figure 2.7 shows how a simple task takeover can be conceptualized: After the human requests to take over a task, the robot must interrupt what it is currently doing and transfer the task to the human. Once the human has completed the interrupted task, the robot can be triggered to resume its remaining tasks.

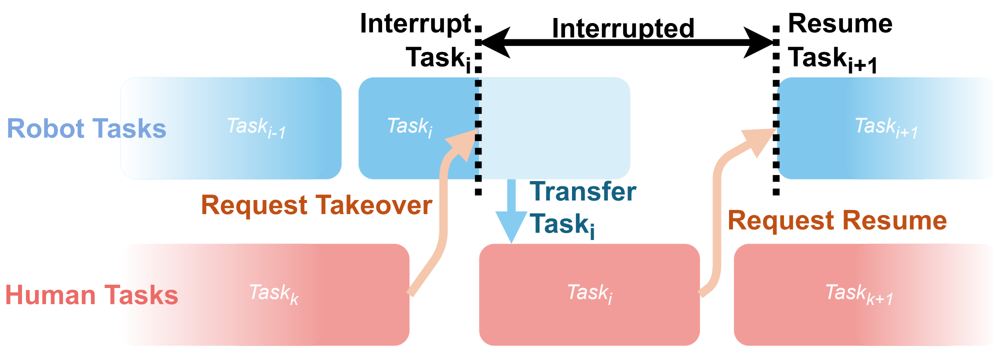


Figure 2.7: Concept of a Task Takeover)

2.4 Theories of Communication

A primary challenge in HRI is to design robots that can effectively communicate with humans [13], [48], [49]. Yet, regardless of being a core enabler of purposeful human-robot interaction [8], literature struggles to agree on a single definition, as it "is an umbrella term for different perspectives and constructs" [50, p. 3].

Often regarded as fundamental work in this regard are Watzlawick's five axioms on human communication [51], discussed in section 2.4.1. Additionally, two frequently used types of communication frameworks are of transmissional and transactional nature [21]. Transmissional models of communication, depict communicating as a linear one-directional act of transmitting a message from an information source to a destination [50]. Transactional models, on the other hand, try to capture the dynamic, irreversible, and continuous characteristics, by introducing concepts such as feedback loops for establishing shared awareness in order to achieve a common goal and reduce misunderstandings [21]. The general concepts, as well as respective advantages and drawbacks of these two approaches, will be discussed in section 2.4.2 (transmissional) and 2.4.3 (transactional).

Despite several publicly available frameworks, which were specifically developed for application in HRI and in combination with the fact that robots are currently only capable of imitating, replicating, or approximating human communication [52], Krämer et al. [24] argue that human will always look for familiarities with previous experiences. Hence, every theory trying to depict human-robot communication should be grounded in human-human communication theories [21], [24]. The main advantage of using human-human communication theories is a huge corpus of already available insights into human expectations [21]. Potential drawbacks, however, include not being able to fulfill human expectations, due to expectations exceeding achievable humanlike robot behaviors [53] and restricted design options due to focusing on said human likeness [54].

2.4.1 Watzlawick's Five Axioms

One of the most cited theories on communication are Watzlawick's five axioms, which sum up to: (1) One cannot not communicate, (2) every communication includes content and relationships, (3) communication is punctuated, (4) communication is digital (verbal) and analog (nonverbal), (5) communication is either complementary or symmetric [51]. Relating the five axioms to human-robot interactions and consequently to human-robot communication, [50] sees the key takeaway of the first axiom in recognizing that communicating with a human involves unconscious processes. This means that any action or behavior, or even the lack thereof, can potentially be interpreted, hence every design decision intentionally or not can influence success in HRI [50]. As humans tend to attribute minds to almost anything they encounter [9], the notion of the second axiom highlighting how relationships (e.g friends vs strangers) shape the communication content and the third axiom emphasizing how different parties can construct diverging meanings out of the flow of communication [51], are vital in acknowledging how humans can understand robots. Utilizing the fourth axiom which tries to separate the spoken word from how it is said and how other non-verbal cues can influence communication [51], research can examine how different forms and modalities of human-robot communication can be categorized [55]. The fifth and final axiom relates to power dynamics [51]. For human-human communication, symmetric would mean that both parties behave as equals [51]. Using [13]'s authorial relation viewpoint (see 2.1) this would roughly translate to human and robot as peers. In this context of power, complementary would mean that there is unequal power, for human-human this could be boss-employee, parent-child, teacher-student and so on [51]. For human-robot on the other hand, again using [13]'s third perspective on HRI, this translates to scenarios with humans as supervisors or operators.

2.4.2 Linear (transmissional) Frameworks of Communication

As HMI and HRI, in particular, are rather recent areas of research (see section 2.1), commonly used frameworks to describe unidirectional communication are adapted from research on human-to-human communication as well as research on telecommunication from the early 20$^{\text{th}}$ century. Three of the most influential publications in this regard are “A Mathematical Theory of Communication” by Shannon [56], “The Structure and Function of Communication in Society” by Lasswell [57] both published in 1948 and Berlo's “SMCR” [58] model published in 1960.

Shannon's mathematical model [56], depicted in figure 2.8, initially designed to understand and enhance telecommunication message transmission, views communication as a unidirectional linear process, where messages flow from an information source (left) to a destination (right). In total, [56]'s schematic communication system consists of six essential parts: (1) The *information source* to produce a single message or a set of multiple messages that are supposed to be transmitted to the destination. (2) The *transmitter* to transform messages into transmittable signals. (3) The *channel* as a medium to transmit the signals. (4) The *noise source*, influencing the channel, potentially altering the transmitted signals and resulting in diverging signals at the transmitter and receiver. (5) The *receiver* to reconstruct messages from the received signals. (6) The *destination* a person (or thing), who is supposed to receive the messages. The general schematics of source, transmitter, channel, noise, receiver, and destination are helpful analogies for human-robot communication and can aid in understanding how for example a robot's intended and unintended communication cues can influence successful communication with a human.

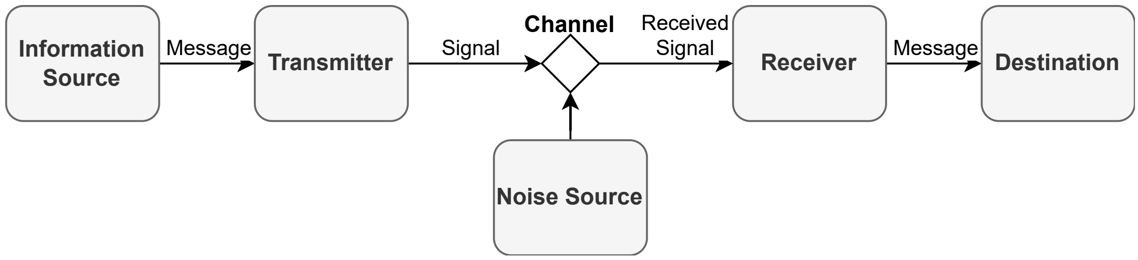


Figure 2.8: Schematic diagram of transmissions (based on [56])

In contrast to that, Lasswell's model of communication within society (figure 2.9), while also depicting communication as a unidirectional linear process, focuses on answering the five questions [57]: (1) "Who?" to identify the person (or thing) initiating the communication. (2) "Says What?" to extract the topic/content of the communicated message. (3) "In Which way?" to answer how the message is conveyed. (4) "To Whom?" to identify the intended recipient of the message. (5) "With What Effect?" to determine if and how the message has observably affected/manipulated the receiver. Based on these five questions covering the entire process, from source and receiver characteristics, over manipulable variables like behavioural cues, to dependent variables such as the effect of the communication, Laswell's 5Ws are still a valuable framework in designing human-robot communication [50]. Especially, as Lasswell's theory implies that communication must be planned and designed before it actually takes place, which is even more true for man-made systems where a designer must in advance consider how the design will implicitly as well as explicitly influence communication [50]. However, following this notion of planning and designing communication, Fritjns et al. [21] argue that interpreting human-robot communication merely as communication with the designer is not sufficient as there will always be an element of unpredictability and situational dependency.

Figure 2.9: Laswell's 5W communication model (based on [57])

Berlo's "SMCR" [58] model of communication, depicted in figure 2.10, aims to extend and combine the models of Shannon and Lasswell by discussing key ingredients shaping communication. At its core, it follows the same concept of linearity with messages flowing from source to receiver [58]. For Berlo, the four main components of communication are: (1) The *Source*, which can either be one individual or a group of individuals, is characterized by the five features (a) *communication skills*, (b) *attitudes*, (c) *knowledge*, (d) *social system*, and (e) *culture* [58]. (2) The *Message* as a physical product produced by the source, e.g., written text, speech, or artwork [58]. Each message can be analyzed based on its elements or its structure and has features such as (a) *code* - which can be a group of structured symbols with meaning, (b) *content*, i.e., the expressed information, and (c) the *treatment* correlating to the decision of which code to use, which content to express, and how to express it [58]. (3) The *Channel*, consisting of a vehicle, its carrier, and docks to load and unload, as an analogy of traveling between shores, e.g in oral communication sound waves are the vehicle, the air is the carrier and hearing and speaking are the corresponding docks [58]. (4) The *Receiver*, defined with identical skills and features as the source. Additionally, these four basic components are aided by (5) an *Encoder* and (6) a *Decoder* to transform ideas into messages and messages back to ideas.

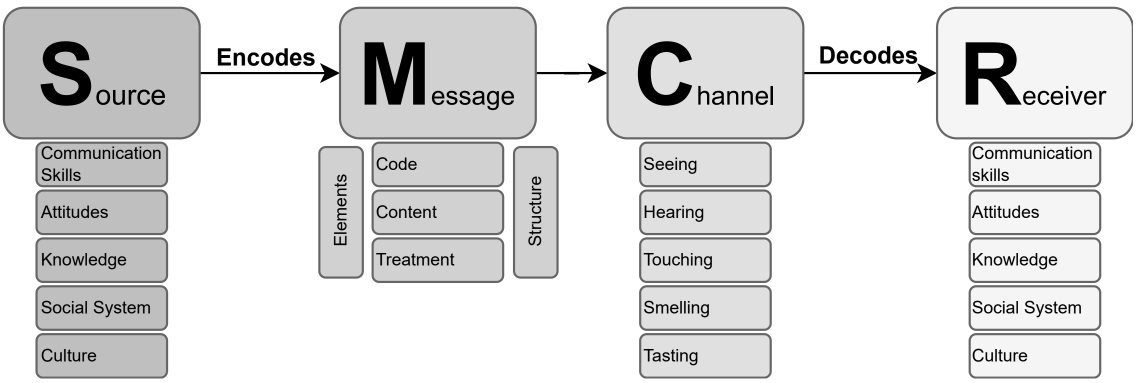


Figure 2.10: SMCR communication model (based on [58])

Due to the rigidity and linearity, some researchers are criticizing that the mutual and dynamic aspect of generating shared meaning via communication is not covered by these transitional models [59]. In addition, it has also been pointed out, that treating information as a carryable physical object fails to account for the crucial role of context in shaping how a message is understood, as the same words can have different meanings depending on the situation and background knowledge of the sender and receiver [21]. However, according to [50] the argument of missing mutuality is primarily valid for complex human-to-human communication and less for communication between robots and humans, where the linearity helps in systematizing and understanding the underlying communication-processes.

2.4.3 Transactional Frameworks of Communication

Analogous to transmissional frameworks, most transactional communication theories are grounded in research on human-to-human communication. However, in contrast to the linearity, they introduce feedback loops from receiver to sender, often even going as far as depicting both involved agents as communicators who continuously send and receive messages. With the main focus on identifying relevant components and factors influencing communication instead of defining the underlying technical process, these frameworks try to see communication as a process of achieving mutual understanding [60] and establish shared meaning with less uncertainty [61]. Most literature viewing communication as a transaction can be linked back to Barnlund's "Transactional Model of Communication" [61] published in 1970 and Kincaid's "Convergence Model of Communication" [60] published in 1979.

Barnlund [61], characterizes communication as a complex, multifaceted and evolving process. According to him, it is continuous, dynamic and circular, i.e. messages flow back and forth in an ongoing, interconnected cycle [61]. Hence, meaning is not directly linked to a perfect reconstruction of the message content itself, like in Shannon's mathematical model [56], but rather to the individual process of constructing meaning based on multiple factors such as behavioural, verbal, nonverbal, public and private cues as well as prior experiences and social, cultural, relational, physical, and psychological context [61]. He also suggests that each event of communication is unique and once a message is communicated, it cannot be taken back, underscoring the irreversible nature of interactions [61]. Figure 2.11 tries to depict this continuous circularity with its numerous elements influencing the creation of shared meaning.

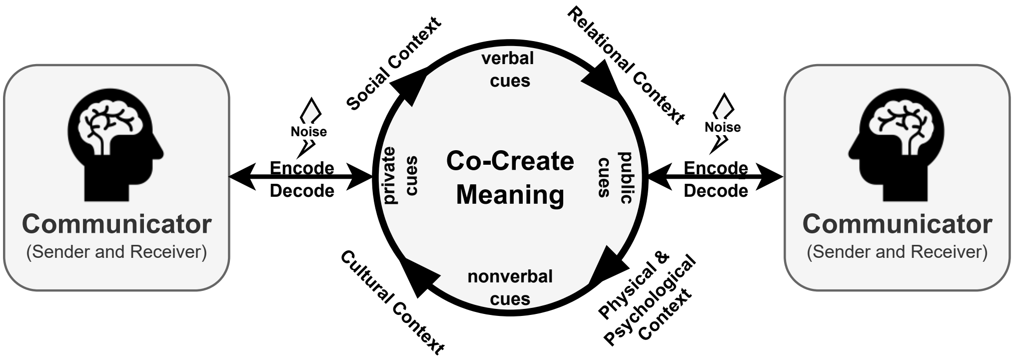


Figure 2.11: Transactional Communication Model (based on [61])

Comparable to Barnlund's approach, Kincaid's convergence model of communication [60], depicted in figure 2.12, interprets communication as a dynamic process aimed at mutual understanding and mutual agreement through the creation and interpretation of information. Meaning is no longer attributed to the message, instead meaning has to be actively negotiated and worked out by the communicating agents [60]. However, this also implies that while the understandings of involved agents might align over time, perfect convergence is unattainable due to each participant's unique prior experiences [60]. According to Kincaid, human communication occurs within each communicator's social, physical, and psychological realities and mutual action, agreement, and understanding are primarily shaped by the individual perception, interpretation, understanding, belief, and action [60].

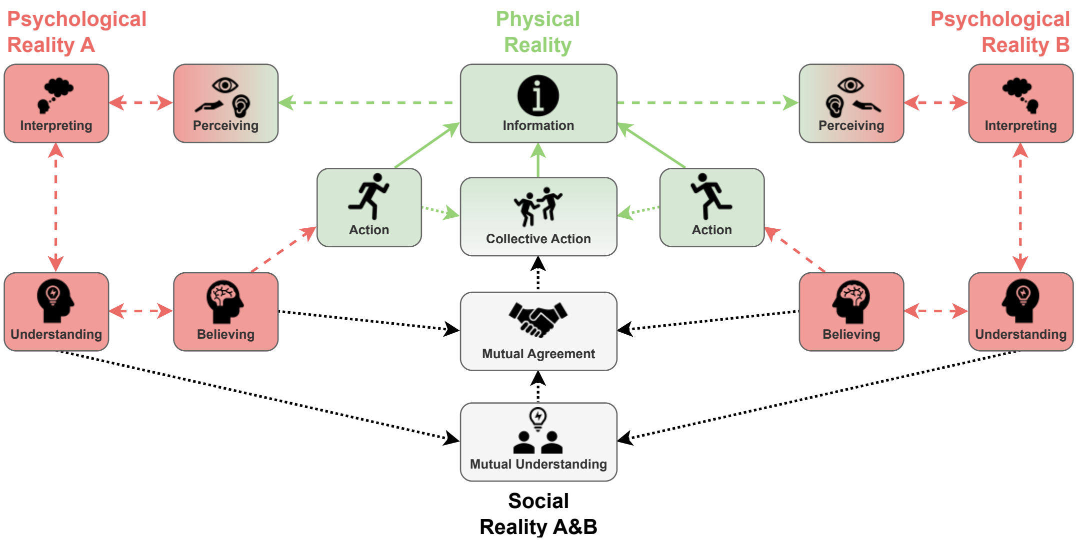


Figure 2.12: Convergence Model of Communication (based on [60])

Although transactional frameworks have been instrumental in understanding human-to-human communication, literature notes objections in regard to their application to other forms of communication [50]. These types of models have also been criticized for their lack of explanation of how meaning is actually produced [62]. Furthermore, their focus on continuous circular mechanisms, which frequently rely on shared background knowledge and mutual understanding, is less applicable to robots that do not possess the same conversational understanding and social cues as humans [24]. This asymmetry in agent capabilities is central to the proposal of HRI specific communication models discussed in section 2.4.4.

2.4.4 HRI Frameworks of Communication

To tackle the debated key issues of classical human-to-human centered communication frameworks in regard to human and robot interactions, research has been discussing simple adaptations to existing models and developing new HRI specific models for the last decades. Classical approaches such as Osgood's and Schramm's model of circular communication, Berlo's SMCR model, and Barnlund's transactional model have often been used as a basis for further exploration [63], [64]. At the core of most HRI specific communication models, such as Banks' and De Graaf's [65] "Agent-Agnostic Transmission Model" and Visser et al.'s [66] "Theory of Longitudinal Trust Calibration in Human–Robot Teams", the reciprocal aspect of communication is depicted by viewing humans and robots as different yet interrelated entities both capable of equally contributing towards a common goal.

In contrast to these symmetric approaches, Frijns et al.'s [21] "Asymmetric MODel of ALterity in Human-Robot Interaction (AMODAL-HRI)" argues that, with currently available technologies, communication between human and robot can not be symmetrical due to agent inherent capabilities and processes. They build their model on well-established theories such as Kincaid's convergence model of communication [60] and link it to Hellström and Bensch's [23] concept of mutual understanding and the Theory of Mind (ToM). Additionally, they incorporate robot-specific aspects such as the robot perception process as described by Christensen and Hager in the "Handbook of Robotics" [67]. Thus their resulting model, depicted in figure 2.13 distinguishes:

(1) The *Human* with psychological capabilities such as perceiving, analyzing and interpreting information, and choosing actions [21]. (2) The *Situation*, which encom-

passes the physical environment and elements of the interaction such as interfaces and interaction outcomes like the situational common ground, which accumulates over time [21]. (3) The *Robotic System* in terms of its sensors and actuators, but also its information handling and behavioral capabilities [21]. Following the general approach of trying to model similarities between the asymmetric agents, another important aspect of their framework is the inclusion of short-term and long-term memory for the human, and respectively a buffer and database for the robot [21]. These long and short term storages are the foundation for arriving at common ground, as they hold the humans and robots (although arguably artificial) beliefs, mental models, situational awareness, inferred intent, and goals [21].

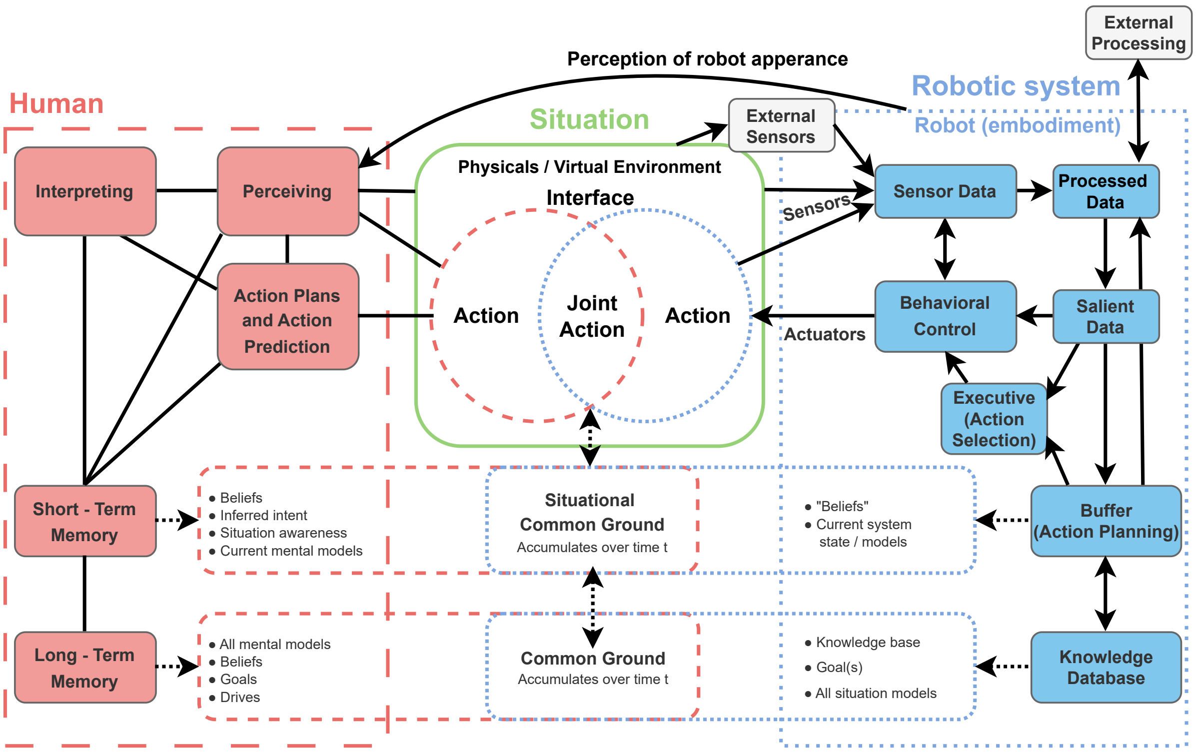


Figure 2.13: Asymmetric MODel of ALterity in Human-Robot Interaction (based on [21])

Despite the benefits of using feedback loops [21], [66], concepts of mutual understanding [21], [23], and dynamic agent roles [65], the applicability of these approaches in HRI is mostly challenged by the difficulty to perceive and predict human behavior, attitudes and internal states of mind [24], [50]. Therefore, it seems reasonable to follow Kunold's and Onnasch's approach [50] of discretizing human-robot communication acts. This discretization allows for a better understanding of how each individual act of communication influences the Human-Machine System (HMS) and helps to identify applicable robot design options [50]. Their framework builds on Lasswell's theory of mass communication and extends it by introducing context-related key ingredients necessary for answering his five questions in regard to HRI (see figure 2.14). In a way their ingredients are comparable to how Berlo's "SMCR" model introduced the essential features for human-to-human communication. Accordingly, they try to answer Lasswel's 5Ws as follows:

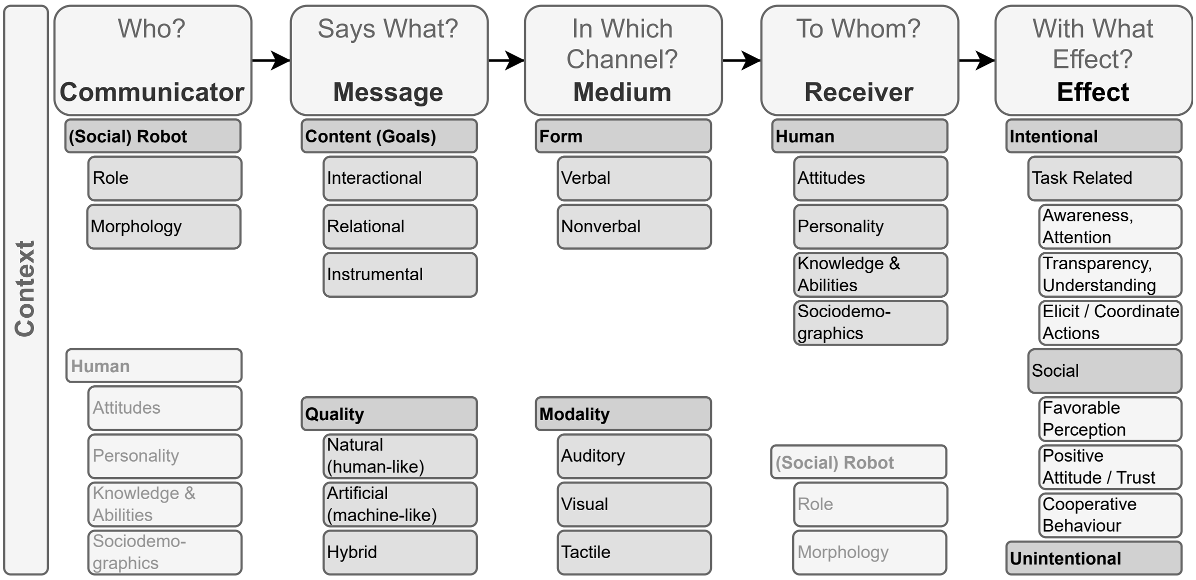


Figure 2.14: Linear HRI - Framework (based on [50])

(1) **"Who?"** - The *Source* of information as either (a) a (social) robot, which can be characterized by its role and morphology, or (b) a human, with attitudes, a personality, personal knowledge and abilities, and a sociodemographic background such as age, gender, education, ethnicity [50].

(2) **"Says What?"** - The *Message* of a given communicative act, can be put into context based on (a) its content and (b) quality [50] - a detailed discussion of the characteristics of a message can be found in section 2.5.

(3) **"In Which Channel"** - The *Channel* as a combination of (a) communication form and (b) output modality [50], further elaboration can be found in section 2.6.

(4) **"To Whom?"** - The *Receiver* again as either (a) a (social) robot characterized by role and morphology, or (b) a human with the already mentioned human traits [50].

(5) **"With What Effect"** - The *Effect* of communicative actions can either be (a) intentional and related to the task or social aspects, e.g., to raise awareness or shift attention, to coordinate actions, to enhance transparency and understanding, or to affect trust [50]. Alternatively, the effect can be (b) unintentional, e.g. caused by a random movement that gets wrongfully interpreted [50], linking the framework to Watzlawick's first axiom "One cannot not communicate" [51].

2.5 Communication - The Message

Despite their varying understanding of human-robot communication, most of the theories discussed in section 2.4, see the message and respectively its meaning as a central element of communication. Expanding on Lasswell's theory, the answer to his second question "Says What?" [57], discussing what the content of a message conveys, can be elaborated through (1) *interpersonal message-centered goals* [68] and additionally, in the context of HRI it can also be categorized by (2) *the quality* of how the message is communicated [69] as well as (3) *six distinct message types* [55].

Taking a closer look at (1), the three major message-centered goals proposed by Burleson et al. [68] are: (a) *Interactional goals*, in order to control the flow of communication, e.g. to initiate or end a conversation, adapt messages for an audience, or shape the impression one has on others [68]. (b) *Relational goals*, where tactics such as humor, politeness, or flattery can be used to build, maintain, or restore a relationship [68]. (c) *Instrumental goals*, covering all functional aspects such as asking for additional information or the fulfillment of a task [68].

When talking about (2), the quality of a communicated message, research proposes a spectrum of three approaches for considering how a robot could be communicating messages: (a) The *natural* approach, suggests enabling robots to use similar communication patterns to humans, in order to improve relatability and effectiveness [69]. In contrast, (b) the *artificial* approach advocates for creating robots that communicate in a way clearly distinguishable from humans [69]. (c) The *none of the above* approach involves mixed techniques and suggests integrating elements of human likeness and robot characteristics [69].

Following the intent of (3), Gustavsson et al.'s [55] messages types, the messages required for HRI can be categorized based on their content as follows: (a) *Command* *messages*, to communicate what to do next, for example, start or stop commands [55]. (b) *Data messages*, to inform either agent about things such as dimensions, quantity, speed, system pressure, etc. [55]. (c) *Highlighting messages*, to reference objects and their location in the physical world, in order to perform tasks where for example, the human asks the robot to get a specific object [55]. (d) *Demonstration messages*, to illustrate the chronological and real world positional information required for the execution of longer work-flows, such as product assemblies [55]. (e) *Guidance messages*, to demonstrate to the robot how it should move for a specific task by physically moving it to the appropriate real world poses [55]. (f) *Option messages*, to suggest context dependant options to the human, e.g. to change the current operating mode [55].

2.6 Communication Channels (= Form + Modality)

Besides the message, the majority of theories examined in section 2.4 include the communication channel as an essential element, which can be further distinguished by (1) *Form* and (2) *Modality* [8]. When talking about, (1) the form of communication, most literature distinguishes between (a) verbal and (b) nonverbal communication [8]. Following this distinction, verbal communication utilizes written or spoken words to share messages [50]. In contrast, nonverbal communication refers to information being transmitted via facial expressions, body language, gestures, and other physical or visual cues instead of words [50]. In regard to (2), visual, auditive, and tactile are typically mentioned as the most common modalities in HRI [19]. Some exemplary combinations of form and modality and their respective, channels of communication are shown in table 2.4.

	Modality		
Form	*Auditive*	*Visual*	*Tactile*
Verbal	Speech	Text	Braille dots
Nonverbal	Primitive Sounds	Pictures, Facial Expressions	Vibrations

Table 2.4: Example combinations of form and modality

In addition to characterizing the communication channel by form and modality, it can also be useful to distinguish between (1) *explicit* and (2) *implicit* communication [8]. Typically, explicit communication refers to an intentional combination of a specific form and modality, to reduce the possibility of misinterpretations and ambiguities. In contrast, implicit communication usually refers to nonverbal cues like pitch and tone in a verbal message or visual cues like facial expression, gaze, and body language.

Drawing on Frijns et al.'s [21] claim of inherent agent asymmetry, coupled with Kantowitz and Sorkin's model of Human-Machine Interaction [14] (see figure 2.1), it is reasonable to differentiate between Human to Robot Communication (HTRC) and Robot to Human Communication (RTHC) in addition to categorizing by form and modality. In a broader sense, Human to Robot Communication can be seen as an enabling factor to empower humans to operate the robot [70]. Despite the design notion of human-centered design, focusing on making robots adapt to humans, the majority of HTRC to date is mediated by physical interfaces [55]. Section 2.6.1 and 2.6.2 further discuss the modalities for HTRC and RTHC channels.

The combination of HTRC and RTHC channels is typically defined as part of the User Interface (UI) [70]. The most common UIs to date are Graphical User Interface (GUI), Voice User Interface (VUI), and most recently Natural User Interface (NUI) [70]. GUIs typically rely on a combination of verbal as well as nonverbal visual RTHC [70]. VUIs, on the other hand, solely rely on bidirectional auditive communication [70]. While GUIs are based on metaphors, such as the office metaphor with a computer desktop and a trash can for deleting objects, NUIs do not need metaphors anymore, as they try to follow the notion of human-centered design and rely on direct input, such as gestures or facial expressions [70].

2.6.1 Human to Robot Communication Channels

Human-to-human communication is complex and multimodal as it utilizes most human responders and can include verbal as well as nonverbal aspects [71]. Therefore, it involves channels such as speech, gestures, facial expressions, an body language [71]. Humans can practically respond with almost any singular muscle or group of muscles, e.g. use a finger or hand to touch something or point somewhere, raise an eyebrow to express disbelief, walk or run via a coordinated movement of upper and lower limbs, or produce sounds via synchronized muscle contractions affecting the lung, vocal chords, tongue and lips [72]. Hence, to enable natural and intuitive interaction, robots need to be able to perceive and interpret these signals and cues of human speech, gaze, facial expression, and body language in order to respond appropriately [71]. There are numerous technologies that can be utilized in Human to Robot Communication channels [70]. These technologies can be roughly grouped into six subgroups each meeting different requirements in regard to wearability, coverage, the necessity for active involvement of at least one of the user's hands, and the complexity of the transmittable message:

(1) **Simple haptic control elements**, such as levers, joysticks, and mouse pads are usually not wearable and have limited coverage as the require physical contact with the user's hands [55]. In regard to message complexity, they are typically used for sending simpler command messages [55].

(2) **Buttons**, e.g real or virtual keyboards are mostly also not directly wearable, however, trough modern touch screens and for example smartphones or smartwatches soft-buttons have also become wearable [55]. Nevertheless, real and virtual buttons and keyboards typically require the use of hands [55]. Message complexity is also limited to simpler types, with the ability to send command messages and, when using a keyboard with multiple buttons, more complex data messages [55].

(3) **Microphones**, for speech recognition allow for hands-free HTRC and can either be wearable or stationary [55]. Hence the coverage is either limited by the microphone's range to detect sound or the microphone's wireless connection to the robot's control unit [55]. The transmittable message complexity includes command messages and more complex data messages

(4) **Cameras and motion sensors**, can be utilized for hands-free recognition of gaze and facial expressions but also for detecting gestures, body language, or human intent [55]. Although there are individual use cases with wearable cameras, it is by far more common to have a fixed camera position and, therefore, limited coverage based on the camera's field of view [55]. In contrast to that, motion sensors are by design meant to be worn and can be less restricting when it comes to coverage [55]. Message complexity can range from simple command messages to highlighting messages, or even demonstration messages.

(5) **Guiding devices**, to allow for manual and direct movement of the robot, require physical contact and can be used to transmit guidance messages to the robot [55].

(6) **Brain Computer Interfaces**, are the latest addition of wearable technologies enabling hands free HTRC [73]. However, these systems are still an area of active research [73], but in theory, they should be able to transmit most of the message types

discussed in section 2.5.

2.6.2 Robot to Human Communication Channels

All modalities used in RTHC need to transmit information capable of being perceived by at least one human sensation: (1) sight, (2) touch, (3) hearing, (4) taste, (5) smell, (6) temperature, (7) acceleration, (8) pain or (9) the position of body parts [14], [72], [74]. A key feature shared among all human sensors is the interplay of sensing physical phenomena such as light rays, physical contact, sound waves, flavors, odors, etc. with the respective sensory organ and the necessary processing of nerve signals in the brain in order to perceive and develop a mental model of awareness [72]. Without claims to completeness, table 2.5 depicts a list of human sensory modalities and their respective organs.

Sensory System	Sensation	Organ
Visual	Color, Brightness	Eye
Tactile	Pressure, Touch, Vibration	Skin
Auditory	Pitch, Loudness	Inner Ear
Gustatory	Taste	Surface of the Tongue
Olfactory	Smell	Nasal Cavity
Thermal	Temperature	Skin
Vestibular	Linear and Angular Acceleration	Middle Ear
Pain Perception	Pain	All Free Nerve Endings
Kinesthetic	Position of Body Parts	Muscles and Spine

Table 2.5: Overview of sensory modalities (based on [74])

To produce one of these signals detectable by a human sensor, multiple technologies can be utilized. Similar to HTRC, these technologies meet different requirements in regard to wearability, coverage, usage of hands, and the transmittable message type:

(1) **Simple Indicators**, such as light beacons do not require the human to hold a device but require a direct line of sight to be perceivable [55]. In contrast to that, haptic alerts need physical contact in order to work. Whereas simple analog indicators, like pressure gauges, are also hands-free and again limited by an unobscured line of sight [70]. For light beacons and haptic alerts, the transmittable message complexity is mostly limited to safety-critical command messages, simple highlighting messages like in pick-by-light scenarios, or option messages [55]. Simple analog indicators, on the other hand, are typically used for sending data messages.

(2) **Speakers and Headphones**, enable hands-free RTHC and can be worn or have a fixed location [55]. Therefore, the limiting factor in coverage is mostly the ambient noise level. Message complexity can range from simple command messages to data messages and, to some degree, even highlighting messages and option messages are possible [55]

(3) **Graphical Displays**, are the most common technology used to facilitate RTHC [70]. With the exception of mobile devices, they do not require to be worn by the user and allow for hands-free perception. Coverage is again limited to a direct line of sight. In regard to message complexity, graphical displays are the most versatile technology for RTHC, as they are capable of basically covering all message types discussed in section 2.5.

(4) **Augmented Reality (AR) Devices** are a special group of displays capable of integrating virtual elements into real environments [73]. Depending on whether AR is achieved via spatial monitors, spatial projectors, hand-held devices, or head-mounted devices, humans can be required to wear something or use their hands to be able to perceive [55]. When it comes to message complexity, AR is equally versatile as regular displays, also covering all message types.

(5) **The Robot Embodiment**, as communication technology, can be interpreted following Watzlawick's [51] first axiom and extends beyond the concept of Gustavsson et al.'s [55] HRI message types as the robot's intentional or unintentional behavior, appearance, and presence alone can inherently communicate messages such as joy, frustration, or urgency.

3 State of the Art

In this chapter, the state of the art in key areas essential for seamless handling of mid-task interruptions in HRC are discussed in greater detail. Generally speaking, there are several fundamental requirements to enable these kinds of interruptions and task takeovers: (1) The interrupting agent must be able to explicitly or implicitly communicate the intent to takeover, (2) the interrupted agent must be able to recognize the intent to takeover (3) the task must be transferred from interrupted to interrupter, (4) the interrupted must be able to recover from the interruption.

To address these aforementioned requirements, the first subchapter 3.1, explores different methods of how tasks are currently dynamically allocated between humans and robots, with a special focus on (a) the respective general concepts, (b) how these techniques integrate the handling of mid-task interruptions, and (c) if and how the inclusion of communication techniques is accomplished. This is followed by the second subchapter 3.2, examining various explicit as well as implicit methods for transferring tasks between humans and robots.

Methodology

To compile a list of relevant literature a semi-structured literature review, following the suggestions by Carrera [75], was employed. Utilizing scientific databases such as ACM Digital Library[1], IEEE Xplore[2], and SCOPUS[3], an initial set of 542 articles was identified using the following set of keywords: **("task allocat*" OR "task plan*) AND "human" AND "robot" AND "collabo*"**. After removing duplicates, 480 articles remained. In order to perform a suggested initial screening based on the title and abstract [75], the following inclusion and exclusion criteria are applied to the remaining articles:

[1] dl.acm.org
[2] ieeexplore.ieee.org/
[3] scopus.com/

- Only full-text articles are considered for review.
- Only articles accessible via open access or through the TU Wien library are included.
- Only papers specifically addressing tabletop HRC are included.
- The numerical relation (see section **??**) between human and robot must be one on one.
- Papers focusing primarily on reviews as well as proceedings are excluded.
- Only studies with a focus on dynamic allocation during runtime are included.

The initial screening narrowed the list of applicable papers down to 31. Following a more thorough reading process paired with forward and backward search to find additional literature especially relevant to section 3.2, 26 key articles are selected for in-depth analysis, representing the major types of task allocation and communication techniques relevant to this study.

3.1 Dynamic Task Allocation: Readiness for Mid-Task Interruptions

In this section, the state-of-the-art of dynamic task allocation is explored with respect to how different allocation strategies can support the requirements discussed in the introduction of this chapter and enable fluid responses to dynamic changes within collaborative tasks. Based on the in-depth analysis of the selected articles, four commonly used dynamic task allocation techniques can be identified: (1) *AND/OR Graphs* to create a hierarchical representation of tasks with alternative execution paths (section 3.1.1), (2) *Behavior Trees* as modular and reactive behavior model (section 3.1.2), (3) *Hierarchical Task Networks* to define decomposable action sequences with specified constraints and ordering (section 3.1.3), and (4) *Markov Decision Processes* as a probabilistic framework for decision-making under uncertainty (section 3.1.4). Given the unpredictable nature of collaborative work, understanding these techniques is essential for developing systems that can effectively handle mid-task interruptions, allowing for seamless transitions and adaptations when unexpected events or changes in human intentions occur.

3.1.1 AND/OR Graphs

The concept of graph theory has been researched as a fundamental problem-solving process since the 17$^{\text{th}}$ century, with early renowned contributors such as Euler, Leibniz, Cauchy, and many more [76]. Within graph theory, AND/OR Graphs (AOG) are a special group of Directed Acyclic Graphs (DAG) for representing and solving problems with complex dependencies [77]. Utilizing AOGs, solving a problem (e.g., task allocation) can, therefore, be viewed as searching a path of minimal cost within the graph [77]. In general AND/OR Graph (AOG) based algorithms can be used for static as well as dynamic task allocation with the main differentiation of which nodes are included and when the graph is queried to find the optimal path. Therefore, most static as well as dynamic techniques include (1) the *Task Representation Layer* (see figure 3.1) for semantically formalizing the problem at hand and (2) the *Task Planning Layer* to perform the necessary calculation to find a preferably optimal solution:

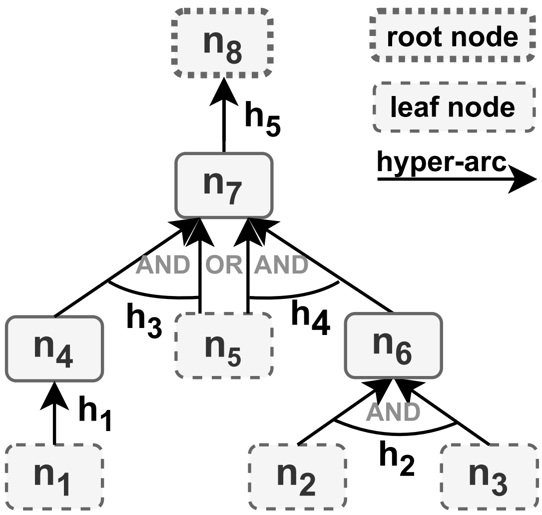


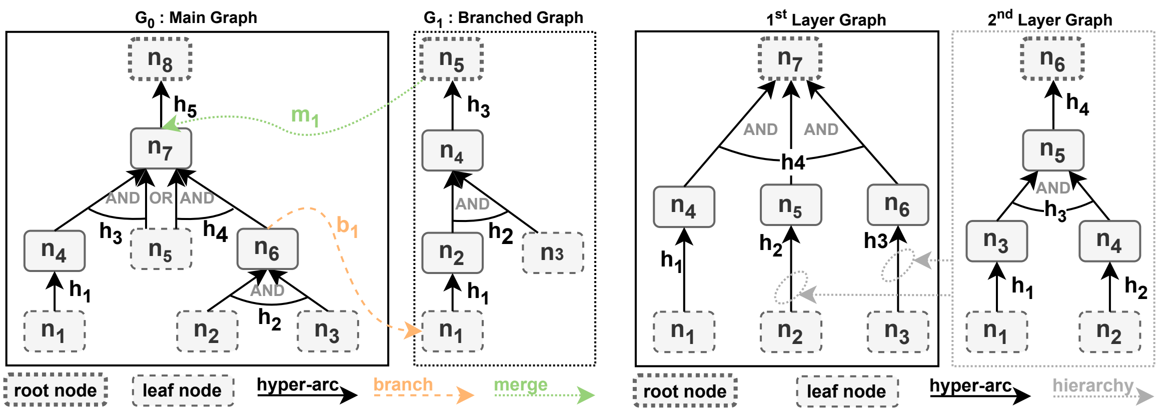
Figure 3.1: AND/OR Graph Representation (based on [78])

(1) **Task Representation Layer**: To create a semantic representation of the hierarchical dependencies within a collaborative task, a predefined description is used to construct a directed acyclic state-based graph [77]. Mathematically, a graph $\mathcal{G}$ can be defined as tuple $\langle \mathcal{N}, \mathcal{H} \rangle$, where each node $n \in \mathcal{N}$ denotes a potential tasks state and each hyper-arc $h \in \mathcal{H}$ a transition between states [43]. Additionally, when trying to solve the task allocation problem, each hyper-arc is associated with a sequence of required actions $A(h_i) = (a_j, ..., a_k) \in A$ [43]. To model a logic AND, a hyper-arc can be used to connect multiple child nodes $\mathcal{N}_C \subset \mathcal{N}$ to a single parent node $n_P \in \mathcal{N}$ [43]. In contrast to

that, a logic OR is modeled by having multiple hyper-arcs act on a single parent node [43]. Within the graph $\mathcal{G}$, multiple paths $P_i = \{n_i, ..., n_j, h_l, ..., h_m\}$ can be defined and associated with the individual cost of nodes and arcs $c(P_i) = \sum_{k_1=i}^{j} c(n_{k_1}) + \sum_{k_2=l}^{m} c(h_{k_2})$ to connect the root node $n_r \in \mathcal{N}$ with a set of leaf nodes $\mathcal{N}_L \subset \mathcal{N}$ [43]. For dynamic algorithms, the initialization marks all actions $A(h_i) \in A$ as *undone* and all nodes (states) $n \in \mathcal{N}$ together with all hyper-arcs $h \in \mathcal{H}$ as *unsolved* [43].

(2) **Task Planning Layer**: To solve the problem of task allocation, some AOG based methods such as proposed by Knepper et al. [79] and Johannsmeier et al. [80] employ heuristic search algorithms like A* [81] or AO* [82] to directly find the most suitable path based on the associated costs. While A* is guaranteed to find the optimal solution, AO* and similar algorithms compromise optimality to reduce the required computational complexity [79]. In contrast to that, the frameworks by Darvish et al. [77], [83] and Karami et al. [78] use a breadth-first graph traversal search for selecting branches to simulate and compare them via a utility function, for example, execution time. For dynamic approaches, once either the human or robot executes an action a_i its status is changed to *done*, and after all associated actions $A(h_i)$ are *done* in the correct order, a hyper-arc h_i is marked as *solved* [43]. The respective sets of solved nodes $\mathcal{N}_s$ and solved arcs $\mathcal{H}_s$ can then be used to query the AND/OR graph $\mathcal{G}$ to find sets of currently feasible nodes $\mathcal{N}_f$ and hyper-arcs $\mathcal{H}_f$ which can then be used to solve the task allocation problem [43].

To tackle some of the restrictions in regard to computational complexity and flexibility as well as scalability requirements, more recent expansion of the representation layer of AND/OR Graphs are hierarchical AOGs by Darvish et al. [83] and branched AOGs by Karami et al. [78]. Conceptionally, hierarchical and branched AOGs both aim at reducing the number of relevant nodes and hyper-arcs for online decision-making, however, they achieve this in fundamentally different ways. Branched AOGs reduce the effective number of nodes and arcs by introducing additional graphs, which are only reached and subsequently exited if a set of branching and merging conditions are met (see figure 3.2a) [78]. In contrast to that, hierarchical AOGs are basically multi-layer AOGs where lower level graphs have to be completed in order to be able to execute the next higher level graph (see figure 3.2b) [83]. Each AOG-layer can be used to represent a different level of granularity within a given process, for example, when assembling a kitchen cabinet, the first layer could be the assembly of multiple cabinets, the second layer could be the assembly of a single cabinet and a third layer the preparation of each cabinet wall with support brackets [83].



(a) (b)

Figure 3.2: (a) Branched AND/OR Graph (based on [78]), (b) Hierarchical AND/OR Graph (based on [83])

While AND/OR Graph based algorithms are capable of online re-planning in case of unexpected failures during the plan execution and aim to decrease the cognitive load on the human [83], they have been criticized for assuming optimal human collaboration behavior and limiting human collaborators to passive roles [3]. Additionally, to the author's best knowledge, there have not been any publications utilizing AOGs to handle mid-task interruptions, although concepts such as Karami et al.'s branched AOGs [78] could potentially be adapted to accommodate the required functionality. While some of the examined AOG based allocation techniques limited communication to be mediated via a GUI, others used human activity recognition to facilitate a gesture-based communication

3.1.2 Behavior Trees

Initially developed to enhance the modularity and scalability of Non-Player Characters (NPCs) in game design, recent research has transferred the concept of Behavior Trees (BTs) to robotic applications [84]. Not unlike AND/OR Graphs, BTs are represented as directed rooted trees (see figure 3.3), a special form of Directed Acyclic Graphs [85]. However, their nodes and arcs do not represent potential task states and state transitions, but rather control and execution functionalities. Using the common terminology of parent and child nodes, every control node must have at least one child node [85]. The root control node is the only node without a parent, and leaf nodes without children are called execution nodes [85]. The control nodes can further be classified as (a) *Sequence*, (b) *Fallback*, (c) *Parallel*, or (d) *Decorator* nodes, and the execution nodes as (a) *Action*, or (b) *Condition* nodes [85]. Utilizing these simple structural elements allows for straightforward code reuse, incremental functionality design, and efficient testing [85].

The execution of a BT is initiated by the root control node sending signals at a specific frequency, these signals are propagated over the control nodes and trigger the so-called ticks at the receiving child nodes, which immediately return either (a) *Success*, (b) *Failure*, or (c) *Running*, depending on their current status [85]. It's worth mentioning that a tick of a child node is only executed if it receives a signal [85]. Table 3.1 shows the different node types and how they can be evaluated during a tick.

Node type	**Symbol**	**Success If**	**Failure If**	**Running If**
Sequence	$\boxed{\rightarrow}$	all children succeed	one child fails	one child is running
Fallback	$\boxed{?}$	one child succeeds	all children fail	one child is running
Parallel	$\boxed{\rightrightarrows}$	M children succeed	N-M children fail	neither succeeded nor failed
Decorator	$\diamond$	custom	custom	custom
Action	$\boxed{\textit{text}}$	completed	failed to complete	currently executing
Condition	$\boxed{\textit{text}}$	evaluates to true	evaluates to false	Never

Table 3.1: Node Types of a Behavior Tree (adapted from [85])

In order to adapt the concept of Behavior Trees to Human-Robot Collaboration and make the robot adapt to human presence and intent, Fusaro et al. [84] extend the standard definition of BTs with three custom nodes: The (1) *Reactive Fallback* node marked with $_R?$ is used to reactively select tasks which the human has not chosen to execute. The custom decorator (2) *KeepRunningUntilSuccess* denoted with $\circlearrowright$ keeps an action running until it succeeds [84]. Additionally, they introduce a slightly adapted sequence control node (3) *Sequence-Cost* indicated by $\rightarrow^{\$}$ to adapt the tick sequence of its child nodes online, based on a cost-utility function [84]. To calculate the cost assigned to each child of a $\rightarrow^{\$}$ node, they use a triple of (a) duration index, (b) ergonomics index, and (c) travel distance index [84].

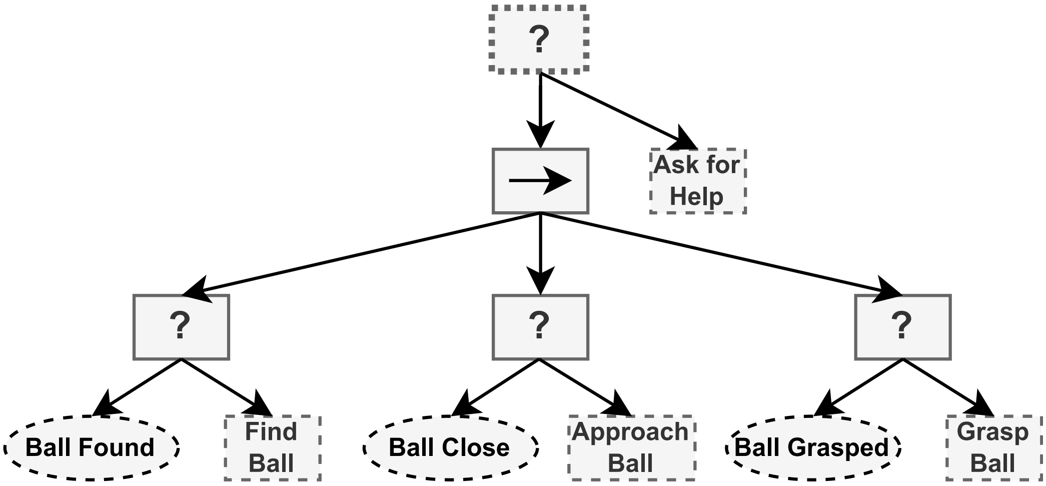


Figure 3.3: Example Behaviour Tree Representation (based on [85])

Despite using the term "reactive" to describe their method of task planning, [84]'s proposed Behavior Tree based algorithm can at best be classified as low-level ad-hoc decision logic to plan the optimal order of robot assigned steps without taking any hierarchical assembly order into account. Their implementation allows the user to actively select the step they want to accomplish via a smartphone app and enables the robot to "reactively" select its actions to minimize the cost-utility function. Hence, similar to the previously discussed AND/OR Graph based methods, the human is required to explicitly follow the chosen task in order for successful collaboration. Furthermore, Fusaro et al.'s framework [84] does only include human pose estimation as part of the ergonomics utility index and not to accommodate any higher level communication. Additionally, HTRC is only modeled via a smartphone as a mediator. However, by implementing more sophisticated intention recognition algorithms as custom control nodes, the concept of BTs could potentially be extended to include more complex HTRC as well as RTHC and the ability of task takeovers midway through execution.

3.1.3 Hierarchical Task Networks

Decomposing huge intangible concepts into smaller more manageable ideas is crucial for human understanding. While this decomposition can be achieved via several strategies, hierarchies are commonly used due to their easy interpretability for humans [5]. Hierarchical Task Networks (HTNs) offer a powerful way to represent and plan complex tasks by breaking them down into smaller, more manageable subtasks [5]. Unlike state-based methods such as AND/OR graphs (see section 3.1.1), HTNs are action-based and decompose tasks into a network of nodes $n \in \mathcal{N}$, with each node representing either a primitive node $n_x \in \mathcal{N}_x$ associated with a directly executable action $a \in \mathcal{A}$ or a compound node that can be decomposed even further [3]. However, similar to AOGs and BTs, the parent and child terminology can be used to define the root node without a parent to represent the entire task and leaf nodes without children as the directly executable atomic actions [1]. The remaining nodes incorporate ordering constraints, dictating how these primitive actions and compound tasks relate to each other during execution [5]. Figure 3.4 depicts an exemplary hierarchical task network for the assembly of a desktop PC.

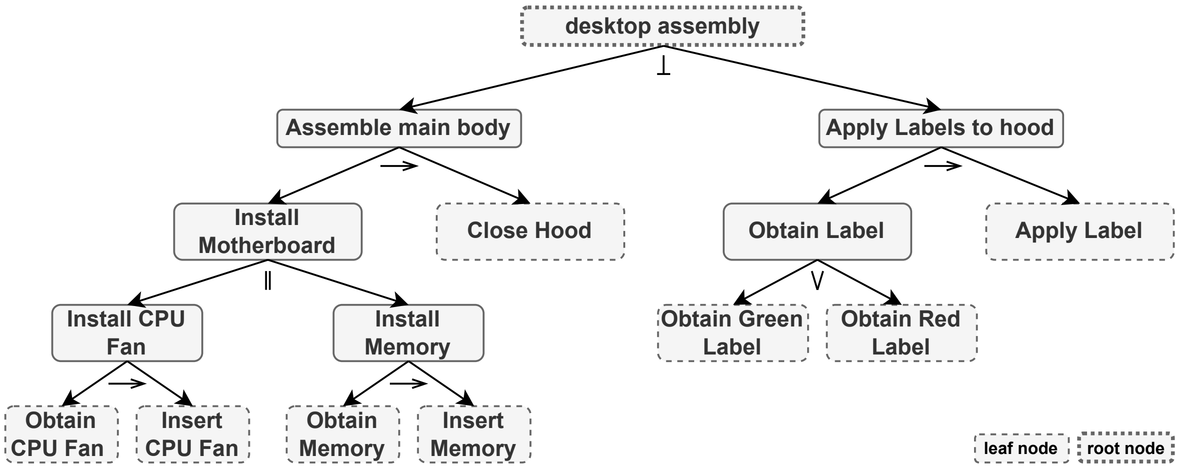


Figure 3.4: Example Hierarchical Task Network (based on [1])

The two most basic types of ordering constraints are, (1) *Fully Ordered*, where child nodes must be run based on a fixed execution order, and (2) *Partially Ordered*, where child nodes can be executed in an arbitrary sequence even in parallel [3]. Expanding on these two basic types, the four most common types of ordering constraint introduced in literature, although not all are used in every publication, are (1) *Sequential*, (2) *Independent*, (3) *Parallel*, and (4) *Alternative* [1], [86], [87].

(1) **Sequential**: Commonly marked with $\rightarrow$, sequential ordering is equivalent to a fully ordered constraint, as it enforces a strict sequence on the execution of its child nodes [1]. For instance, in a desktop PC assembly task, the subtask "Install CPU fan" might have child nodes "Obtain CPU fan" and "Insert CPU fan" which must be executed in that precise order [1].

(2) **Independent**: Denoted with $\perp$, nodes with independent ordering are a subtype of partially ordered constraints, as the execution of their child nodes can be of arbitrary order but conditionally not in parallel [1]. For example, sticking to the PC assembly

use case, although the subtasks "Apply labels to hood" and "Assemble main body" can be performed in any sequence, independently from each other, the atomic tasks "Apply lable" and "Close hood" cannot be executed simultaneously [1].

(3) **Parallel**: Parallel ordering constraints, typically marked with $\|$, are conceptually a special type of independent nodes and allow simultaneous execution of child nodes as long as additional constraints, like resource availability, are met [1]. In case of resource unavailability, they can be performed in any order [1]. Again utilizing the example of the assembly of a desktop PC, "Install CPU fan," "Install memory," and "Tape cables" can be independently installed in any given order or even at the same time.

(4) **Alternative**: In order to extend the functionality of a HTNs representation capabilities, it can be extended by an alternative operator, denoted with $\vee$, to model disjunct subtasks [87]. Within the PC assembly example, this could translate to individualizing the label color within the "Optain Label" subtask.

With humans typically tending to follow hierarchical structures when executing a task, recently several papers adopting Hierarchical Task Networks to solve the problem of task allocation for tabletop human-robot collaboration have been published, for example by *Cheng et al.* [1], and *Ramachandruni et al.* [3]. In 2021, Cheng et al. [1] introduced their communication-free HTN based approach, which prioritizes robot actions parallel to human actions, minimizing potential spatial interference. The time optimization-based planner explicitly leverages parallelism extracted from the hierarchical task model, leading to smoother, conflict-free collaboration [1]. Additionally, they propose a set of algorithms to automatically construct the sequential and parallel hierarchical task model from demonstration, simplifying the system setup procedure [1]. However, as their planner is encoded with the assumption that humans will always complete sequential actions consecutively [1], flexibility in scenarios where humans might want to switch between parallel subtasks based on their preferences or task constraints is limited, and mid-task interruptions are not accounted for. Moreover, their approach is limited by the confinement of the planning horizon to only these parallel subtasks and to task time as a single-objective optimization.

Regarding human-robot interactions, their approach can be used in two modes, namely (a) *command* and (b) *automation* [1]. Unfortunately, the paper does not further specify how the commands are communicated in *command* mode. When in *automation* mode, their approach is truly free of explicit RTHC and only uses a vision-based nonverbal HTRC channel, to implicitly communicate what the human is currently doing by recognizing human actions. However, by including more channels, Cheng et al.'s dynamic task allocation technique could potentially benefit from increased capabilities when it comes to recovering from task interruptions.

Similar to Cheng et al. [1], Ramachandruni et al.'s "UHTP: User-Aware Hierarchical Task Planning Framework" [3], enables robots to adapt to human actions in real-time while maintaining the worker's autonomy without explicit communication. They extend classical HTN by adding collaborative elements such as agent assignments for atomic action nodes ($a_r \in \mathcal{A}_r$, $a_h \in \mathcal{A}_h$) and decision nodes $n_d \in \mathcal{N}_d$ with probabilities to model actions executable by either agent [3]. In contrast to [1]'s limited planning horizon, [3]'s framework optimizes task completion time by minimizing the cost of a combined plan over the entire remaining task horizon. In order to enable mutual adaption and generate valid robot plans, the UHTP algorithms continuously monitor human and robot activities to remove completed or invalid actions from the combined plan and reevaluate the remaining cost objective, until all nodes are removed from the HTN [3]. However, as they assume that individual primitive actions can only be performed by either agent and will always be complete once started [3], their approach is not capable of collaboration as defined by [34] (see section 2.3), and also not of handling unexpected human actions outside the task model e.g task takeovers. Additionally, their assumption of flawless activity recognition might be unattainable in real-world scenarios with sensor noise and occlusions.

While UHTP is designed to be free of explicit (verbal) communication from either agent, it requires implicit HTRC of human actions via a vision-based feed-forward action classification network [3]. However, if the action classification is unable to predict the human's goal, verbal as well as non-verbal channels could be used to signal a lack of understanding or propose alternative actions. Additionally, communication could be beneficial for negotiating how to handle interruptions or unsuccessful completion of primitive tasks.

In contrast to most human-aware task planning approaches, assuming a completely controllable human-agent, Buisan et al. [88], model HRC as interaction between a controllable agent (robot) and an uncontrollable agent (human). Their proposed concept, depicted in figure 3.5, interconnects the two agent's individual beliefs, HTN models, triggers, agendas, and plans [88].

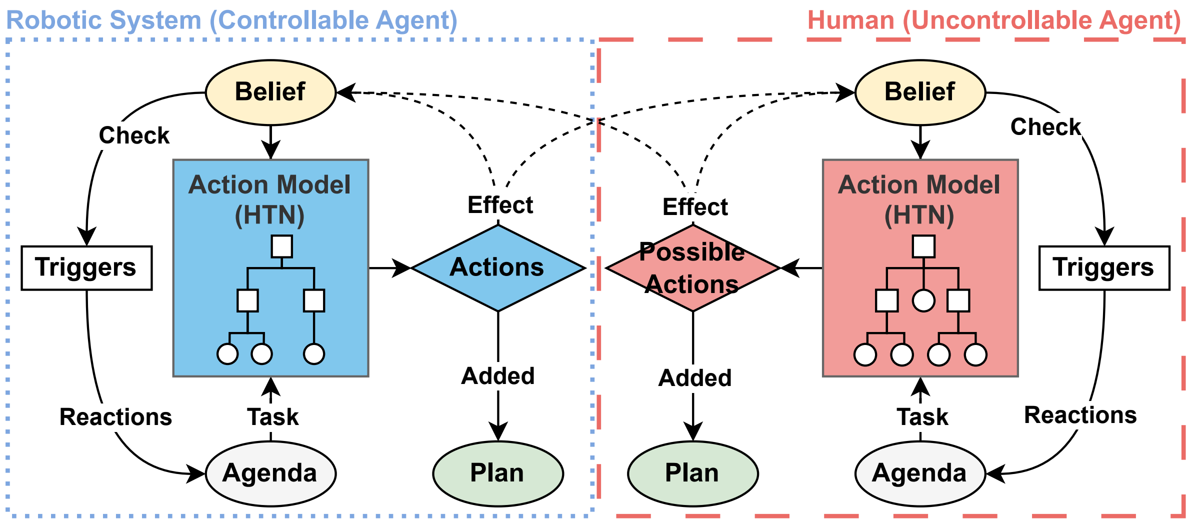


Figure 3.5: Multi-Agent Hierarchical Task Network exploration based on belief estimates and communication (adapted from [88])

Each agent is defined as the tuple $\alpha = \langle name_\alpha, \sigma_\alpha, \Lambda_\alpha, Tr_\alpha \rangle$, consisting of (a) the agents $name_\alpha$, (b) the state-tuple $\sigma_\alpha = \langle d_\alpha, \pi_\alpha, s_\alpha \rangle$ with (i) an individual agenda d_α, (ii) a partial plan π_α, and (iii) the agents belief s_α, combined with (c) a unique HTN action model $\Lambda_\alpha = \langle Op_\alpha, Ab_\alpha, Me_\alpha \rangle$ represented by (i) primitive tasks Op_α, (ii) abstract tasks Ab_α, and (iii) methods Me_α, and last but not least (d) agent inherent triggers Tr_α [88].

In regard to communication, they only mention that knowledge alignment is achieved via verbal communication between both agents [88]. However, they do not go into greater detail to discuss the involved communication channels in consideration of the utilized modality. Furthermore, they do not address the handling of task interruptions. Although they mention using trigger functions to represent agent reactions to specific situations [88], it is unclear how these triggers would manage interruptions to an ongoing task. In essence, while Buisan et al. [88] lay a foundation for human-aware task planning, the question of how communication and task interruptions can be handled requires further research to extend their framework.

3.1.4 Markov Decision Processes

Conceptionally, Markov Decision Processes (MDPs) are an extension of the stochastic concept of Markov Chains first introduced by Russian mathematician Andrey Markov in 1906 [89]. Whereas Markov Chains can be utilized as a stochastic representation to model a potential state-event-series by assigning a probability for each event that only depends on the previous state [89], MDPs can be used to enable decision-making and control functionality within such a time discrete stochastic process where the outcome is partially random and partially controllable [90], [91]. Early contributions to MDPs date back to the 1950s and 1960s with influential publications such as “A Markovian Decision Process” by Bellman [90] and Howard’s book on “Dynamic programming and Markov processes” [91]. Nowadays, MDPs are employed to solve dynamic problems in various domains such as manufacturing, robotics, economics, and automatic control.

Within the research on robot planning and decision-making, MDPs are often utilized to characterize tasks and capabilities due to their expressive graphical representation and arguably easy understandability [86]. Typically, a MDP framework is represented as an environment-oriented directed graph with a 4-tuple $\langle \mathcal{S}, \mathcal{A}, R, P \rangle$ [86]. Similar to AOGs, the domain knowledge and environment get encoded into nodes representing the possible states ($s \in \mathcal{S}$), and directed arcs (edges) connecting the nodes, which are labeled with actions ($a \in \mathcal{A}$) [86]. Additionally, each transition from state s to s', due to an action a, is assigned a reward $R = Pr(s'|s,a)$ for transitioning and a dynamic system probability $P = Pr(s'|s,a)$ indicating the likelihood of arriving at s' from s when a is executed [86]. Figure 3.6 depicts a simple MDP for a train that wants to quickly travel to its destination. Following the just introduced nomenclature this MDP consists of the three states $\mathcal{S} = \{cold, warm, overheated\}$, two actions $A = \{slow\,down, go\,faster\}$ and the associated probabilities P and rewards R.

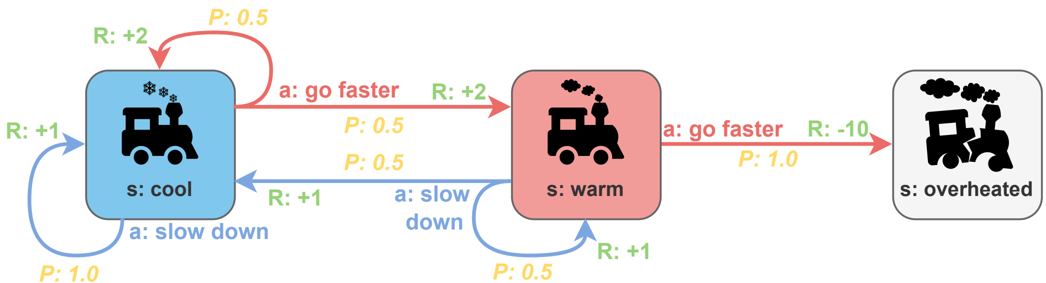


Figure 3.6: Example Markov Decision Process

Despite their general usefulness for dynamic decision making, a major drawback of standard MDPs is the limitation to observable states [92]. In order to extend the application of MDPs for deployments where the agent cannot directly observe the underlying state of the system, Partially Observable Markov Decision Processes (POMDPs) can be formulated. In contrast to directly observing the system’s states, POMDPs provide a framework to maintain an agent's belief over possible states by updating the agent belief based on probabilistic observations [93]. This belief can then be used to select actions that maximize expected rewards over time, considering the inherent system uncertainty. Therefore, POMDPs are no longer represented by a 4-tuple, but rather a 9-tuple $\langle \mathcal{S}, \mathcal{A}, \mathcal{O}, T, Z, R, b_0, h, \gamma \rangle$ [93]. Whereas states $\mathcal{S}$, actions $\mathcal{A}$, and rewards R are defined analog to MDPs, observations are denoted as $o \in \mathcal{O}$, the transition probability of reaching s' from s when executing a as $T = Pr(s'|s, a)$, the observation probability of o' in state s' as $Z = Pr(o'|a, s')$, the initial belief distribution as $b_0 = Pr(s_0)$, the finite or infinite planing horizon as h, and the discount factor accounting for how much previous rewards can influence the current time step as $\gamma \in [0, 1]$ [93].

To enable robotic systems to reason under uncertainty, Roncone et al. [87], utilize a combination of HTNs and atomic POMDPs to optimize task completion time by dynamically assigning roles. Their proposed technique is capable of automatically transforming each of the HTNs primitive tasks (leaf nodes) into a modular POMDP, which then exploits written or auditory verbal communication as a means to reduce uncertainty between both agents by informing about robot intent and asking for human intent [87]. Additionally, combining HTNs as high-level and POMDPs only for atomic task representations, improves scalability as the state space only grows linear [87]. However, their current method requires both agents to follow the agreed-upon role assignment without mid-action replanning [87], limiting flexibility in dynamic collaborative scenarios. Moreover, the assumption of a "blind" robot that can only rely on acts of verbal communication to coordinate with its human collaborator might increase the task execution time and limit the flexibility of real-world applications.

According to Unhelkar et al. [94], excessive communication can result in a decrease in human-robot team performance. Their framework, CommPlan, enables robots to make informed decisions regarding if, when, and how communication with its human partner is appropriate during human-robot collaboration in sequential tasks [94]. They utilize a combination of (1) a hybrid *model specification process*, where some parts are defined by the developer and others are learned, to describe the *Robot Decision-Making Model* represented by a POMDP and (2) a *execution-time POMDP solver* which is used to select the appropriate robotic actions and communication efforts [94]. The hybrid model specification of the robots POMDP includes several sub-modules such as (a) the *Task Model* described via a Multi-Agent Markov Decision Process (MMDP), (b) the *Communication Capabilities*, (c) a novel task- and context-specific non-linear *Commu-*

nication Cost Model, (d) a *Human Response Model* to estimate if and how the human responds to the robots query, and (e) a *Human Action-Selection Model* represented by a Agent Markov Model (AMM) to model sequential human decision making [94]. With the model specified, the robot utilizes its online POMDP solver to dynamically determine if, when, and what to communicate during task execution [94]. This involves updating its beliefs about the human and choosing actions and communications that maximize the team's overall reward [94]. Although Unhelkar et al. [94] demonstrate the effectiveness of their approach in a shared workspace task, they acknowledge that specifying accurate cost models for various communication types, particularly as communication options increase, can be challenging. Additionally, while their hybrid approach to modeling human behavior aims to address the complexity of human-robot interaction, the reliance on developer-specified probabilities introduces subjectivity and potential inaccuracies in predicting human responses to robot communication.

Like the previously discussed frameworks, CommPlan does not explicitly incorporate the possibility of mid-task interruptions. However, utilizing a similar approach to determining if, when, and what to communicate could be useful for handling task takeovers, as communication might not always be necessary. In regard to utilized communication channels, CommPlan is designed to be flexible. Although it is demonstrated using speech, the authors state that it can be equally suitable for other verbal modalities such as visual signals or text [94].

3.2 Review of Human to Robot Communication Channels

The following section will primarily discuss current state-of-the-art HTRC techniques that can be used to transmit a command-message in order to initiate a transfer of a task from a robot to a human. As to the author's best knowledge, there are no publications dealing with mid-task interruptions or task takeovers in tabletop HRC settings, hence, the discussed methods will mostly be related to ad-hoc task allocation techniques and concepts from other research domains, due to conceptual similarities and transferability to the problem at hand. Following the distinction between explicit vs implicit communication, discussed in section 2.6, it is again useful to distinguish between (1) *explicit*, where users actively invoke something, and (2) *implicit*, where the robot infers human intent based on a pre-defined heuristic and/or the user's actions [95].

3.2.1 Explicit

Considering the six technology groups introduced in section 2.6.1, explicit HTRC to facilitate a transfer of tasks can, in theory, be accomplished via all of them. Below is an overview of how current research incorporates these. Due to the already discussed limitation of available literature discussing task takeovers in HRC, it must be mentioned that some methods have not been validated for the proposed use case. Nonetheless, they can be considered a useful guideline during the implementation process.

Simple Haptic Control Elements and Buttons

Theoretically, simple elements such as joysticks, levers, (physical or virtual) buttons, and mouse pads can be used to invoke a mid-task interruption and transfer of task respectively. However, as most of these elements are typically stationary and not wearable (see section 2.6.1), using such an interface requires the human to first go to some sort of control panel, e.g. the teach pendant of the robot, to then be able to communicate the intent to take over. Hence, these elements are typically only considered for simple web interfaces on mobile devices or in virtual environments. For example, Roncone et al. [87] and Fusaro et al. [84] both utilize a web interface accessed from a tablet or smartphone to facilitate HTRC. As opposed to Mahadevan et al. [95], who propose a virtual-reality menu of soft buttons to allocate objects/tasks from the human to the robot. However, in contrast to research on driving-task-takeovers for autonomous vehicles, which typically include physical buttons on the steering wheel to actively request control over the vehicle at any given time [96], there seems to be no such equivalent to request mid-task-interruptions for tabletop HRC scenarios, except for pausing or emergency stopping the robot without context awareness for the robot.

Microphones

Unlike the just discussed simple elements, microphones have been frequently used for making human-robot collaboration more intuitive by enabling robots to understand and respond to explicit auditory communication, similar to how we humans interact with each other [97]. As the transmittable message complexity, discussed in section 2.5, can include command messages as well as more complex data messages, microphones in combination with Automatic Speech Recognition (ASR) have been utilized for HTRC to facilitate explicit ad-hoc allocation techniques where verbal commands such as “Can you grab object X at location Y" can be used to allocate tasks to the robot [95]. In addition to the already, discussed coverage and range of stationary vs wearable microphones (see section 2.6.1), ASR is primarily affected by (1) the *number of simultaneous speakers*, (2) the *vocabulary size* i.e. the number of recognizable words, and (3) the *spectral bandwidth* of trained vs production system [98]. With ASR being considered among the most complex problems of computer science[4], it is a typical example for a Machine Learning (ML) problem. ASR can in principle be solved using many well-established ML algorithms such as Artificial Neural Network (ANN), Convolutional Neural Network (CNN), or Deep Neural Network (DNN) to name only a few [99].

Reverting back to HRC research, Mahadevan et al. [95] for instance, use verbal commands, though in a virtual environment, as a baseline to compare other implicit and explicit allocation techniques. Similarly, Angleraud et al. [100] use speech recognition and natural language processing to generate and adapt action plans for joint human-robot tasks. However, to the author's best knowledge, current research has not addressed the possibility of a robot interrupting its task based on the auditory input it receives mid-task. Instead, the focus has been on designing systems that enable robots to understand human intent and respond accordingly, often in a structured and semi-pre-planned manner with distinct task boundaries.

Cameras and Motion Sensors

Similar to microphones, cameras and motion sensors are actively being investigated as means to enable explicit HTRC via gestures [97]. Utilizing the message complexity classification introduced in section 2.5, cameras and motion sensors are best suited to facilitate explicit communication for simple command messages through gestures or highlighting messages by estimating the object the human is pointing at [101]. In general, one can distinguish between (a) *body gestures*, (b) *hand and arm gestures*, and (c) *head and facial gestures*, all of which require different *sensory technology* and algorithms to *identify*, *track* and *classify* the perceived explicit gestures [102]. As with ASR, describing each sensory technology and methodology of gesture identification, tracking, and classification in greater detail would quickly turn into a thesis of its own. Hence, only a short summary is provided here:

From the 1980s to the 2000s, early research on sensors to enable recognizing gestures

[4] www.ibm.com/topics/speech-recognition

and human motions was primarily focused on glove-based systems [102]. With significant advancements in image processing, single cameras became more popular from roughly 1995 to 2005 [102]. Since the 2010s, most research has been done utilizing either depth sensors, wearables such as wristbands, and non-wearables such as radio frequency-based systems [102]. In order to distinguish and identify gestures either visual features, learning algorithms, or skeleton models (most common) are typically used [102]. For tracking gestures over time, early research relied on single-hypothesis algorithms, whereas more recent research typically involves multiple concurrent hypotheses or advanced techniques that directly integrate into the identification process [102]. Just like ASR, gesture classification is a typical example for a ML problem. Hence, in principle it can be solved using numerous ML algorithms including but not limited to K-Nearest Neighbours (KNN), Hidden Markov Model (HMM), Support Vector Machine (SVM), Dynamic Time Warping (DTW), Artificial Neural Network (ANN), Convolutional Neural Network (CNN), or deep learning [102].

Again focusing on HRC research, Mahadevan et al. [95], for example, propose multiple explicit gesture-based HTRC techniques for a virtual environment such as pushing an object toward the robot or into a predefined zone. In contrast to that, Gottardi et al. [101], suggest a real-world application of a depth camera paired with a CNN, to recognize and classify a predefined set of gestures such as clapping, raising one hand, or pointing at a desired location to invoke robot actions such as requesting the next step, triggering a workspace object detection, and stopping the robot. Likewise, Karami et al. [78] utilize a vision-based system based on OpenPose [103] to detect human gestures, in their case, raising either arm to create a branch to the AOG. These branches can then be used to update a goal pose through kinesthetic teaching or to repeat a process step [78].

While [78]'s and [101]'s approaches both provide a mechanism for humans to signal command requests to the robot, their current implementation does not include the capability to interrupt the robot at any given time. For instance, Karami et al.'s [78] branching of the main AOG only works at certain nodes of the main graph and Gottardi et al.'s [101] user commands are only processed after a primitive action has been completed.

Guiding Devices

As mentioned in section 2.3, cobots are by design equipped with collision sensing technologies such as collision-sensitive enclosures or torque monitoring. The torque monitoring capabilities are tightly integrated with the cobot's motion control algorithms, as they need to be able to detect deviations from the expected static and dynamic forces/torques with low latency to ensure safety [104]. The most common strategy called impedance control, simultaneously handles force and position control of the robot [104]. The primary benefits of using torque monitoring-enabled control strategies are (1) the increased safety due to reduced risks of injuries on collisions [104] and (2) the ability to directly guide the robot via physical contact between humans and robots during setup and programming [36]. However, the potential of utilizing these capabilities to facilitate explicit HTRC for deliberate human intervention during a task's execution remains largely unexplored.

Brain Computer Interfaces

Recent research advances have shown that Brain Computer Interfaces (BCIs) can be used to communicate explicit human intent to a robot [105]. The fundamental principle of BCI lies within the continuous interaction of brain cells and the change in electric potential they constantly induce in the scalp [106]. By utilizing a electroencephalograph (EEG), these changes in electric potential can be measured and transformed into signals which, in turn, can be analyzed to infer explicit human intent [106]. While there are many different types of BCIs, steady-state visual evoked potential (SSVEP) based BCIs are often used in HTRC, as they require less training and adaptation for individual users due to their relatively strong signal compared to other BCI methods [106]. The fundamentals of SSVEP are electrical signals produced by the brain in response to visual stimulation at specific frequencies [106]. For example, if a human is looking at an object flickering at 12 Hz, their EEG signals will peak at 12 Hz [106]. A BCI system can make use of this phenomenon by attaching different flickering frequencies to distinct objects, allowing the system to infer which object the human user wants to select by analyzing their EEG signals [106]. In this way, BCI can provide a way for humans to communicate simple command and highlighting messages without requiring physical movement.

Like most of the other discussed technologies, BCIs are currently underrepresented in HRC research when it comes to mid-task interruptions and transfer of task authority between human and robot. However, theoretically approaches such as SSVEP could be adapted to enable task takeovers.

3.2.2 Implicit

Even though explicit HTRC can reliably be used to establish joint actions, it usually entails a reduction in efficiency as it requires the human to interrupt its current task in order to explicitly communicate [8]. To address the need for flexibility without sacrificing efficiency, researchers are exploring implicit techniques inspired by coordination behavior observed in human collaboration [95]. Unlike explicit HTRC, implicit communication can not be facilitated by all technology groups discussed in section 2.6.1. This is primarily due to the style of interaction where, for example, simple haptic control elements, buttons, microphones, and guiding devices all require the user to actively interact. In contrast, as highlighted by a recent meta-analysis by Arents et al. [97], cameras and motion sensors are not only perfectly capable of inferring implicit HTRC based on the motion data they extract, they are also by far the most researched technology to enable implicit HTRC.

In general, the majority of motion-based techniques, have comparable requirements to the already discussed explicit gesture-based methods in regards to (1) *sensory technology*, (2) *identification*, (3) *tracking*, and (4) *classification* of human motion [102]. Hence, the utilized sensory technology is basically identical, with 3D cameras being the most common, followed by wearables such as Inertial Measurment Units (IMUs) to measure accelerations, forces, and other tactile feedback [97]. Previous literature on implicit HTRC methods for tabletop HRC scenarios can roughly be separated in research focused on (1) *action recognition* and (2) *action / intent prediction*. The following paragraphs aim to give an overview of the state-of-the-art in implicit action recognition and action/intent prediction algorithms proposed for tabletop HRC settings.

Action Recognition

In collaborative settings, human action recognition plays a crucial role in enabling implicit non-verbal communication between humans and robots [8]. By accurately interpreting human actions, robots can adapt their behavior, and contribute to a more seamless and efficient collaborative process [8]. Especially within HRC research on ad-hoc as well as proactive task allocation many techniques based on action recognition have been proposed.

Tuli et al. [107], for example, present a novel approach called HAROPP (Human Activity Recognition On Probabilistic Partition), which utilizes probabilistic spatial partitions based on continuous probability density estimates to recognize human activities. In other words, they propose to enhance activity recognition by fusing the spatial workplace layout information and human skeleton-based motion data gathered via an Inertial Measurment Unit (IMU) to map a human worker's real-time movements to a pre-defined library of workplace activities and partition the physical space into action sets [107]. According to the authors, utilizing this approach of partitioning the physical space is not only easier to implement but also provides better explainability, transparency, and interpretability of classification results when compared to other state-of-the-art action recognition techniques like deep learning [107]. Tuli et al.'s [107] approach could be adapted to handle task takeovers by using probabilistic spatial partitions to recognize when a human is performing an action within the robot's workspace. This information could then be used to trigger a change in the robot's behavior, such as pausing its current task or moving out of the way.

Still based on wearable IMUs, Darvish et al. [77] propose a human action recognition module that classifies human actions by comparing real-time IMU data from wearable sensors against a set of predefined gesture models [77]. These models are built offline using Gaussian Mixture Modeling (GMM) and Gaussian Mixture Regression (GMR) to represent expected movement patterns and their variations [77]. During online operation, the system calculates the statistical likelihood of the observed movement data matching each gesture model and the likelihood of the observed movement given the current state within the predefined AOG task model, ultimately selecting the model with the highest combined likelihood as the recognized action [77]. When the system recognizes a human action, it updates its internal state within the AOG to reflect the task's progress, using this information to guide the robot's subsequent actions [77]. Even though Darvish et al.'s approach is not primarily designed to handle mid-task interruptions, its perception pipeline could be adapted to recognize when a human operator deviates from the expected sequence of actions as defined in the AOG, resulting in the adaptation of the robots plan without explicit HTRC.

Taking an entirely different direction, Ramachandruni et al. [3] present an activity recognition module that continuously monitors the human body poses utilizing a depth camera. They achieve action classification through a feed-forward neural network trained on human activity data [3]. The model uses a five-layer architecture with dropout layers to classify individual frames of the task into six actions: grab parts, attach battery, attach shell, screw, place drill, and a null class for unknown poses [3]. By observing the human's activities, the robot can implicitly understand their actions and maintain a shared understanding of the task's progress in order to make decisions that support the human [3]. However, their perception pipeline is limited due to the assumption of flawless activity recognition and is not capable of handling unexpected human actions outside the task model. Despite that, the general approach of utilizing a feed-forward neural network to detect human actions could be adapted to also recognize motion sequences associated with the intent to takeover.

Human Intent Prediction

As briefly mentioned in the introduction of this section, another interesting approach to implicit HTRC is human intent prediction, where multiple intertwined algorithms try to anticipate future human trajectories and actions in order to proactively choose robot actions that best complement what the human is doing [3].

In order to drive their communication-free HTN task allocation model [1], Cheng et al. [108] propose a 3D-skeleton-based "recognition-then-prediction framework" [108] for anticipating human actions in a collaborative assembly task over longer time horizons. Their framework allows a robot to not only estimate the duration of both current and future actions, but also to predict a human worker's future motion by analyzing the human's past and current actions using an algorithm that considers spatial and hierarchical task information, the observed hand trajectory, and the history of human activities [108]. The human hand trajectories are formulated using a sigma-lognormal model, fitted to training data offline, and then adapted online, using the observed trajectory, human intentions, and target location, to account for variations in human behavior [108]. Finally, the updated sigma-lognormal model predicts the trajectory and estimates the action duration [108]. According to the authors, this predictive capability allows the robot to collaborate with the human worker in a safer and more efficient manner [108]. Even though Cheng et al. [108] do not mention if their method is capable of inferring the intent to interrupt a robot during a task, it seems perfectly capable of doing so with only minor modifications.

Recognizing that the efficiency of an assembly/disassembly process is significantly impacted by the movement of the human worker, Lee et al. [109] propose a task planning algorithm that leverages human intent/action prediction to facilitate implicit communication between human operators and robots in a collaborative setting. Due to numerous potential scenarios based on the current relative position of the human, robot, and parts, predicting future human movement is non-trivial [110]. In order to address these numerous possibilities, the authors leverage a Convolutional Neural Network (CNN) trained on a dataset of images representing various scenarios [110]. These training images capture information about the positions of completed tasks, unfinished tasks, the robot, and the task assigned to the human operator, allowing the CNN to learn spatial relationships between these elements [110]. The proposed CNN extracts features from these images, reduces their dimensionality, and processes them to generate a probability distribution over eight possible movement directions [110]. This distribution is then used to predict the most likely direction the human operator will move in after completing a task, enabling the system to anticipate the operator's actions and optimize task allocation accordingly [110]. Similar to other discussed sources, Lee et al. [110] do not directly address how their approach to implicit communication could be used to handle task takeovers or mid-task interruptions. However, their method of predicting the human operator's movements could potentially be adapted to predict the intent to interrupt based on the human's movement.

Following a completely different approach, Zhang et al. [2] suggest a fusion-based spiking neural network to classify human actions and environmental cues to infer implicit communication. Their proposed model integrates multi-channel inputs such as real-time data on human behavior, robot posture, and the state of the assembly task, each encoded as spiking signals to predict a human's need for collaborative action [2]. This prediction then triggers the robot to undertake the necessary collaborative action, like retrieving and handing over the required tool [2]. The study emphasizes the significance of considering temporal aspects and the interdependency of various factors in a human-robot collaborative environment to facilitate more intuitive and effective interactions [2]. Through experimental validation, Zhang et al. [2] demonstrate that their method surpasses other prediction models, such as Long Short-Term Memory (LSTM) and Hidden Markov Model (HMM), in achieving a faster and more human-like response time in anticipating collaboration requests. By monitoring the same multi-channel inputs of human behavior, robot actions, and task state, the model could be trained to recognize patterns associated with a human intending to interrupt or take over the robot's current task. For example, sudden changes in human movement or interaction with objects relevant to the robot's task could be interpreted as signals of an impending takeover, prompting the robot to yield or pause its actions accordingly.

To foster a more intuitive and seamless collaborative assembly environment, Cramer et al. [111] formulate a POMDP framework that enables the robot to model the inherent uncertainty in predicting human intentions. Their approach leverages an intention graph, a representation of all feasible assembly sequences derived from the product's CAD data, effectively encapsulating the designer's intent [111]. By observing the human's interactions with components and tools via cameras, the robot utilizes the POMDP to dynamically update the probability distribution within the intention graph, highlighting the most likely assembly path the human is pursuing [111]. This allows the robot to anticipate the human's needs and take appropriate collaborative actions, such as fetching the next required part, without explicit communication and any pre-programmed instructions [111]. The authors demonstrate the effectiveness of this approach through a simulated assembly task where the robot, physically simulated by a human, successfully predicts and supports the human operator's actions [111]. With the source's primary focus on using a POMDP for the robot to infer the human's assembly intention and plan collaborative actions, the authors don't directly address task takeovers or mid-task interruptions. Therefore, while Cramer et al. [111] provide a foundation for understanding how a robot might predict human intent in collaborative assembly, they do not offer information on how this approach could specifically handle situations where the human interrupts the robot mid-task.

4 Study and Implementation

This chapter focuses on the design and implementation of a hypothesis-driven study. Therefore it builds upon and incorporates the concepts explored in the preceding chapters. Following Hoffman et al.'s [112] primer for conducting HRI research, the first sub-chapter 4.1 elaborates on the chronological approach of designing and running the experiment. This is followed by the second sub-chapter 4.2 discussing the key aspects of the implementation in greater detail, and the third chapter 4.3 elaborating on how the design and implementation were adapted after running the first pilot experiments.

4.1 Study Design

Owing to the huge variety of HRI research (discussed in chapter 2), there are numerous scholarly practices one can follow when exploring different HRI aspects. For investigating how a human perceives and interacts with a robot, conducting empirical research is often regarded as a crucial step in order to obtain new scientific insights [112]. Following the best practices of empirical research suggested by Hoffman et al. [112], one or more research questions must be clearly articulated before conducting any kind of experiment. For the purpose of the experiments conducted for this thesis, the research question *Q4* defined in section 1.2 will guide the empirical study and its design:

Q4.) To what extent do different HTRC channels impact team dynamics, such as perceived team fluency and trust, during task-takeovers?

4.1.1 Participants and Setting

To achieve a broad and diverse sample, word-of-mouth referrals were used to recruit participants from the general public. Additionally, a laboratory setting was chosen, as this allows for increased consistency across participants due to better experimental control and minimized external variables [112]. To further enhance the study's robustness, a within-participant design is utilized. This allows each participant to experience all conditions while reducing the impact of individual differences and increasing the statistical power of the experiment [112]. In order to mitigate order effects introduced by the within-participant design, counterbalancing the order of conditions ensures that any learning, novelty, fatigue, or familiarity effects are evenly distributed across participants [112]. Additionally, to mitigate order effects introduced by the task itself, participants completed a trial run to get accustomed to the task, the interface, and the different modes of interaction.

4.1.2 Research Constructs

A vital part of empirical research are the so-called constructs, which represent theoretical and abstract ideas directly derived from the research questions [112]. These constructs can have (a) a *causal predictor-outcome* relation, where a change in one construct directly translates to a change in another, or (b) a *correlational* relationship, where the analysis only suggests if two constructs are related at all without knowing the direction of the relationship between them [112]. In accordance with the research question discussed above, this thesis aims to test the relations between the predictor construct *Human to Robot Communication Channel* and outcome constructs associated with team dynamics such as *Team Fluency* or *Trust in the Robot*.

Following [55]'s message classification (see section 2.5), the message required to interrupt the robot and take over its tasks can be categorized as a simple command message. Therefore, it is sufficient to limit the channel of communication (section 2.6) to a non-verbal form. Additionally, the modality can be restricted to visual and tactile, as auditory HTRC modalities often have problems due to the interference of background noise, and neural modalities such as BCI based communication suffer from drifting sensor outputs and information redundancy, therefore, requiring advanced signal processing in order to extract reliable signals (section 3.2). Based on the distinction between explicit and implicit communication (section 2.6) and the discussion of the state-of-the-art technologies (section 3.2), the predictor construct is split into three levels: (a) The *Baseline* condition, based on a non-verbal tactile channel where participants can interrupt the robot through the push of a button, (b) the *Explicit* condition, featuring a non-verbal tactile channel where interruptions are triggered by physically interacting with the robot, and (c) the *Implicit* condition, facilitated via a non-verbal visual channel that infers intent based on the human's actions. The individual HTRC channels and their respective implementation are discussed in greater detail in section 4.2.4.

4.1.3 Hypothesis

After formulating research questions and deriving relevant constructs, the next stage of empirical research is the development of hypotheses, phrased as affirmative statements that propose expected relationships between the investigated constructs [112]. The following hypotheses will be tested:

H1: Explicit tactile HTRC will result in greater task fluency compared to the baseline condition.

H2: Implicit HTRC will result in greater task fluency compared to the explicit condition.

H3: Participants will exhibit higher trust in robots during explicit and baseline HTRC.

4.1.4 Task Design

Based on the research question, the hypothesis, the chosen context and participant structure, as well as the selected constructs, designing a feasible task that conforms to all of the above is not trivial. The first idea of a simple pick-and-place task where the robot is tasked to move cylindrical mint containers from one tray to another was scraped due to the lack of a useful task for the human that would occupy the human enough to not only have them stand by and interrupt the robot at random intervals. The second idea has the robot place colored items into specific bins while the human ensures that the items are correctly positioned and makes adjustments if necessary. However, this idea was also scrapped because it would create a dynamic where the human acts as the robot's supervisor rather than working alongside it as a peer (see table 2.3 on authorial relation [13]).

Keeping in mind the challenges of the first two ideas, the third and final idea evolves around assembling small Lego blocks in two distinct colors (blue and orange). While the human is tasked to collect 4 unique parts from a storage location, the robot delivers the fifth and final part required for assembling a given Lego design. Once human and robot start collecting, the human is allowed to change their mind and switch to the other color, but they need to communicate this switch to the robot. Based on the three levels of the predictor construct, this HTRC can be categorized as follows: (a) In the explicit condition, participants can interrupt the robot by touching it. (b) In the implicit condition, participants can simply reach for the object they want to interact with, and the robot uses a vision system to detect these human actions and interrupt its current task if necessary. (c) In the baseline condition, participants can interrupt the robot through the push of a button.

In order to increase consistency across sessions and participants there is a pseudo-randomized external cue to trigger participants to change their minds, this helps to ensure that each session has the same amount of interruptions. If the human chooses to interrupt the robot's task, they are required to collect the correct part themselves. Figure 4.1 depicts the ten chosen Lego designs, each with a unique combination consisting of one 4x2 brick, and four distinct other bricks.

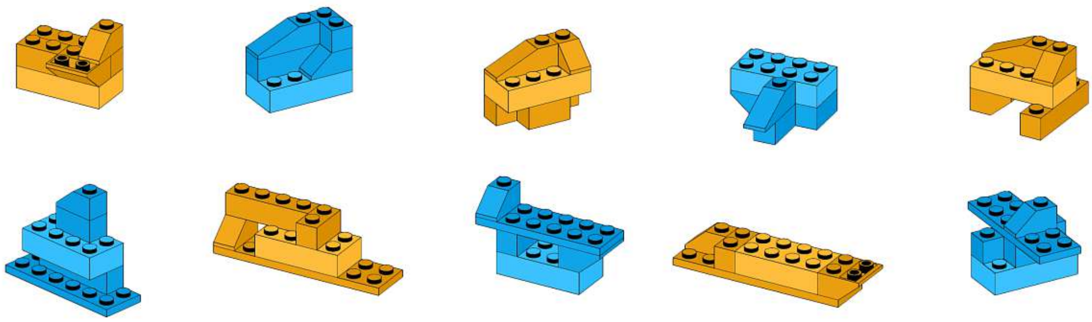


Figure 4.1: Set of Lego Designs

4.1.5 Physical Layout

With the task designed, the next step involves defining the physical layout of the research setup. The first design iteration of the setup, depicted in figure 4.2, consists of:

(a) Robot: One collaborative robot mounted on the first table at $(x, y) = (328, 317)$ mm measured from the lower left corner with its base rotated by an angle of 34°. It features an advanced motion control interface allowing low-level access to commanding the robot's motion and is equipped with force-limiting controllers, making it ideal for the tasks ahead. The robot's location allows it to interact efficiently with both the 3D-printed storage slides and the human operator.

(b) Robot Storage: The 4x2 bricks are delivered by the robot and stored separated by color in two 3D-printed slides, resulting in one pick-up and one drop-off location per slide (see Figure 4.3).

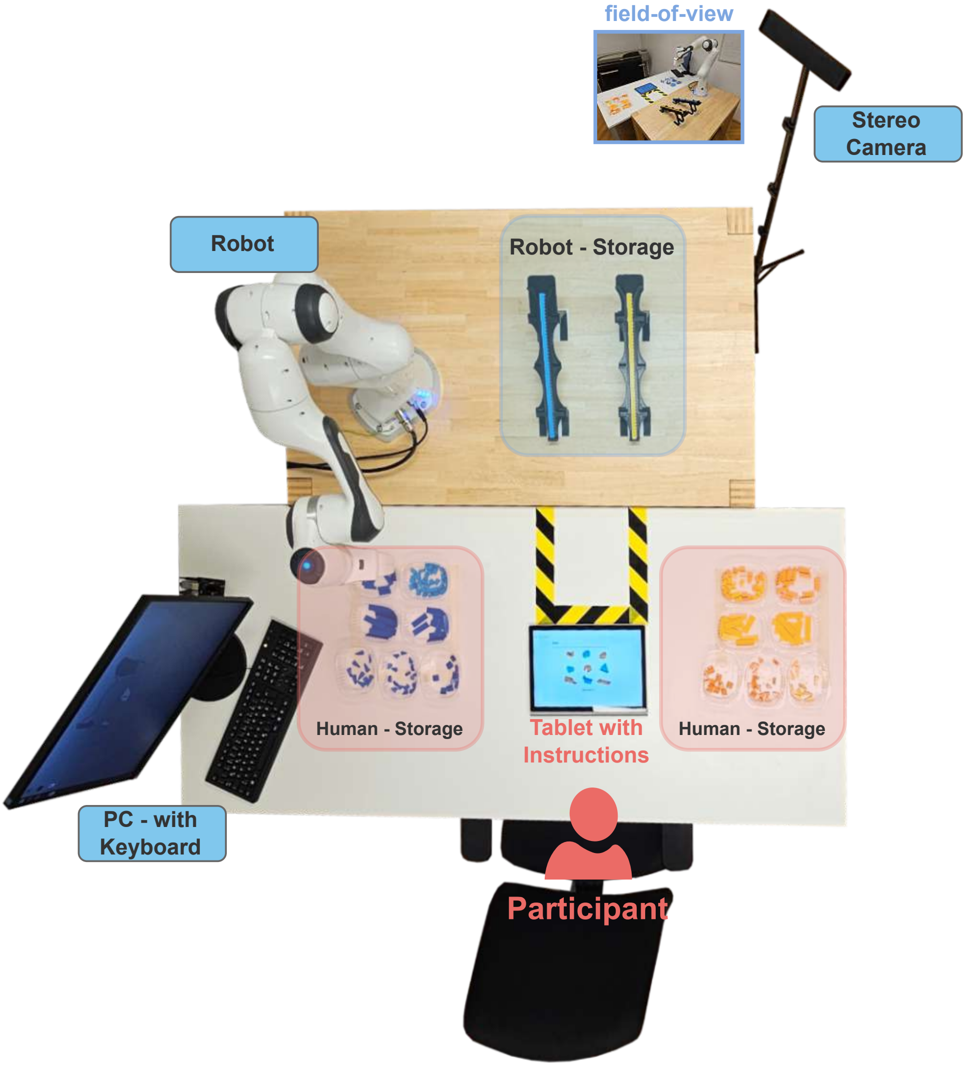


Figure 4.2: First Iteration of Physical Layout

Figure 4.3: 3D-printed Slide

(c) Stereo Camera: In order to equip the robot with spatial vision, a stereo camera is strategically mounted in the top right corner above the robot's workspace, providing a broad field of view. It captures real-time depth information and video, crucial for tracking the robot's surrounding environment.

(d) Human Storage: In contrast to the slides for the robot storage, the remaining seven types of bricks are stored in two dedicated areas on the second table in small color-separated containers where the human can easily access them. Allowing the human to quickly retrieve necessary components, thus streamlining the workflow.

(e) PC with Keyboard: To enable the baseline condition, the first design iteration includes a PC equipped with a keyboard that serves as the robot's control panel, enabling a direct and explicit way to interrupt the robot. It is strategically positioned on the far left of the table to be easily accessible by the human while also requiring a shift of attention away from the other tasks.

(f) Tablet: Last but not least, the setup includes a tablet PC to visualize instructions, guide participants through the experiment, and collect survey responses. It is positioned conveniently for the human operator to view at all times, ensuring that they can follow the progress of the tasks.

4.1.6 Methods of Evaluation - Variables and Measures

With the general task and its physical layout defined, the abstract constructs can be operationalized into observable and measurable variables. This involves carefully defining what each construct means in the selected context and choosing specific methods to measure the outcome constructs [112].

The outcome constructs trust and team fluency, can primarily be operationalized using subjective measures. For the purpose of this thesis, a self-reported questionnaire with 10-point Likert-scale items and one open-ended question is used to assess the participants' perceived fluency and trust after each session with the robot. The selected questions are based on Hoffmann's [113] subjective fluency metric scales and grouped by similarity to get composited indicators directly or indirectly related to fluency and/or trust. The composite score's reported Cronbach's $\alpha \geq 0.772$ should ensure internal consistency among the questionnaire items. For researching task interruptions, [113]'s composite-scales for "Robot-Relative Contribution", "Improvement" and their "Individual Measures” are suspected of providing little insight and are therefore dropped from the questionnaire and replaced by an open-ended question and a Likert-scale capturing how much participants liked a session, reducing the total number of repeated questions to twenty. Most questions are used without modification, only the first question of the “Human-Robot Fluency” composite scale is slightly adapted to be more clear. Additionally, some Likert scales are augmented with additional text to clarify what it means to strongly disagree or strongly agree. The chosen compound scales and their respective questions are depicted in table 4.1. In addition to these fluency metric questions, there are additional post-study questionnaire to capture the participants’ demographics, preferred mode as well as their overall impressions and thoughts about the experiment (see table 4.2).

Human-Robot Fluency
- The robot and I worked fluently together.
- The human-robot team’s fluency improved over time.
- The robot contributed to the fluency of the interaction.

Trust in Robot
- I trusted the robot to do the right thing at the right time.
- The robot was trustworthy.

Positive Teammate Traits
- The robot was intelligent.
- The robot was committed to the task.

Individual Liking & Open-ended
- How much did you like this specific mode of interrupting the robot?
- In 1-2 sentences, describe how you perceived this style of interrupting the robot

Working Alliance - Bond
- I feel uncomfortable with the robot.
- The robot and I understand each other.
- I believe the robot likes me.
- The robot and I respect each other.
- I am confident in the robot’s ability to help me.
- I feel that the robot appreciates me.
- The robot and I trust each other.

Working Alliance - Goal
- The robot perceives accurately what my goals are.
- The robot does not understand what I am trying to accomplish.
- The robot and I are working towards mutually agreed upon goals.
- I find what I am doing with the robot confusing.

Table 4.1: Subjective Fluency Metric Items adapted from [113]

Demographics
- What is your age?
- What is your gender?
- What is your occupation?
- How often have interacted with a robot, prior to this experiment?

Preference
- Which mode of interaction was your favorite?
- Which mode of interaction was your least favorite?

Open-Ended
- Do you have any additional comments, feedback, or thoughts about your experience with the robot during this experiment? Any insights on the interaction process, challenges faced, or suggestions for improvement would be appreciated

Table 4.2: Post Study Questionnaire Items

4.1.7 Study Protocol

Following the operationalization of constructs into measurable variables, [112] suggests getting explicit about every aspect of the study's procedure by outlining a step-by-step process through which participants engage with the experiment. This helps to ensure consistency across sessions, minimizes variability, and provides a clear framework for data collection, task administration, and addressing any issues that might arise [112]. In addition to defining what and how something should happen, this phase also includes estimating the timing for each of the experiment's steps. For the purpose of this study figure 4.4 depicts a short summary of all major steps as well as time estimates for each of them. The following paragraphs will explain each step in greater detail.

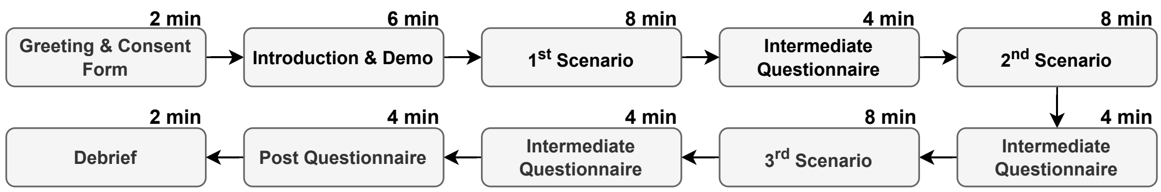


Figure 4.4: Study Procedure and Timing

4.2 Implementation

Besides the study design, the implementation of the system plays a vital role in the overall success of conducting experimental research. Following the four requirements to enable task interruptions, introduced in chapter 3, the subsequent sections elaborate on how these requirements can be implemented to facilitate the study design discussed in the previous section. First, section 4.2.1 details the general idea of the digital task representation, followed by section 4.2.2 discussing the how the developed software is structured and the robot controlled. Then, section 4.2.3 describes how the participants are guided by the study interface, and section 4.2.4 how the individual communication channels are implemented.

4.2.1 Digital Task Representation

After establishing the physical layout, the digital representation of the task needs to be defined. In this context, the discussion of the state-of-the-art in dynamic task allocation techniques and their readiness for mid-task interruptions (section 3.1) reveals that, while some of the discussed techniques are capable of re-allocating tasks, none are specifically designed to handle situations where a human worker purposefully interrupts the robot during its current task. Even though some techniques could be adapted to handle mid-task interruptions, none of the analyzed allocation methods can be directly applied. Instead, a finite-state-machine approach is chosen due to its conceptual similarity to conventional robot programming. Besides the conceptual similarities, the utilization of a finite-state machine allows more complex scenarios with more than one flow of execution. In addition to deciding on the general programming approach, a decision must also be made in regard to the utilized programming language. To streamline the development process, and ensure interoperability among all developed algorithms, Python is chosen as the primary programming language.

A graphical representation of the implemented finite-state-machine is depicted in figure 4.5. The individual robot states are visualized as boxes with rounded corners, and the transitions as arrows, with blue arrows representing automatic transitions and red arrows user-invoked transitions. Besides the depicted states there are several internal variables such as (a) $active\_storgae \in (1, 2)$, to determine whether the robot is supposed to grasp from the storage holding blue or from the storage holding orange bricks, or (b) $mode \in (Explicit, Implicit, Baseline, Demo)$ to enable/disable some of the robot's capabilities based on the active mode.

In total, there are eight states linked by eight transitions. For example, while the robot is moving toward the active storage's pick-up location (see figure 4.3) without having grasped an object, it is in the "*moving to active robot storage*" state. Upon reaching its goal destination successfully, the finite-state machine automatically transitions to the "*grasping action @ the active storage*" state. However, if an interruption occurs during this motion, it instead transitions to the *moving to home pose* state.

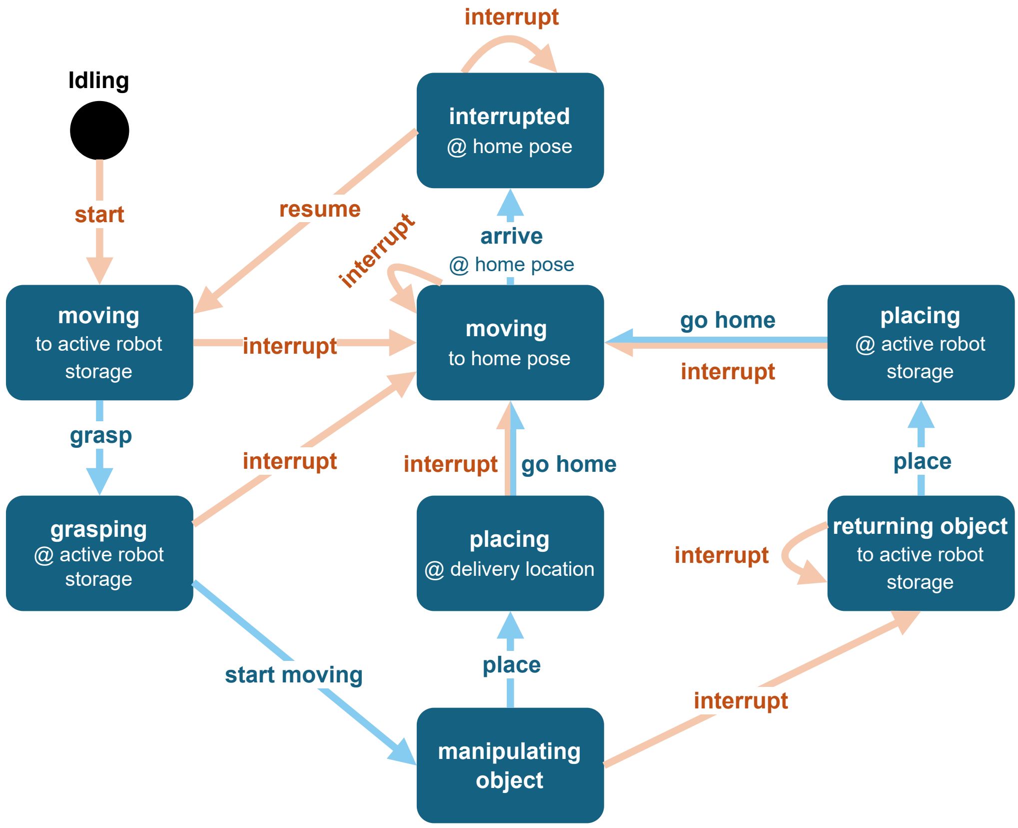


Figure 4.5: Digital Task Representation as Finite-State-Machine

4.2.2 Implemented Software Architecture

To control the robot's motion, the digital task representation must be equipped with interfaces to directly connect to the robot's built-in controller. In order to keep the code as little hardware dependent as possible, the open source Robot Operating System (ROS) [114] is used as middleware to facilitate communication between different software components. It provides standardized interfaces and tools, such as drivers, transformation libraries, inverse-kinematic solvers, and much more, simplifying robotic development.

Theoretically, any collaborative robot controllable via ROS can be used to explore the given research questions. Due to its high degree of sensitivity, precision, and ease of programming, the Franka Emika Panda is a popular choice for experimental setups. Additionally, the availability of open-source software such as libfranka offers a flexible platform for experimentation and development with a low-level interface for controlling the robot in real-time. On top of ROS and libfranka, MoveIt! [115] is utilized for planning and executing motions. Despite nearing its end-of-life (EOL), ROS remains a popular choice for controlling robotic systems. Firstly, in comparison to its predecessor ROS2, it has far more compatible tools and libraries as well as a larger and more active community. Secondly, most robot manufacturers provide stable ROS packages to work with their robots, whereas ROS2 support is often still experimental.

Figure 4.6 depicts a simplified version of the implemented software architecture. Its main components are (a) the robot and its controller with Franka Control Interface enabled, (b) a ROS node communicating with the robot to retrieve measurement data from the robot and send low-level motion commands to the robot, (c) a ROS node with the tasks digital representation and motion planning capabilities, (d) a ROS node to handle the vision-based HTRC channel, (e) a ROS node to handle the baseline HTRC channel, and (f) the anvil uplink to connect the local code to (g) the web app via (h) the cloud hosted anvil server. The individual HTRC nodes are discussed in greater detail in section 4.2.4.

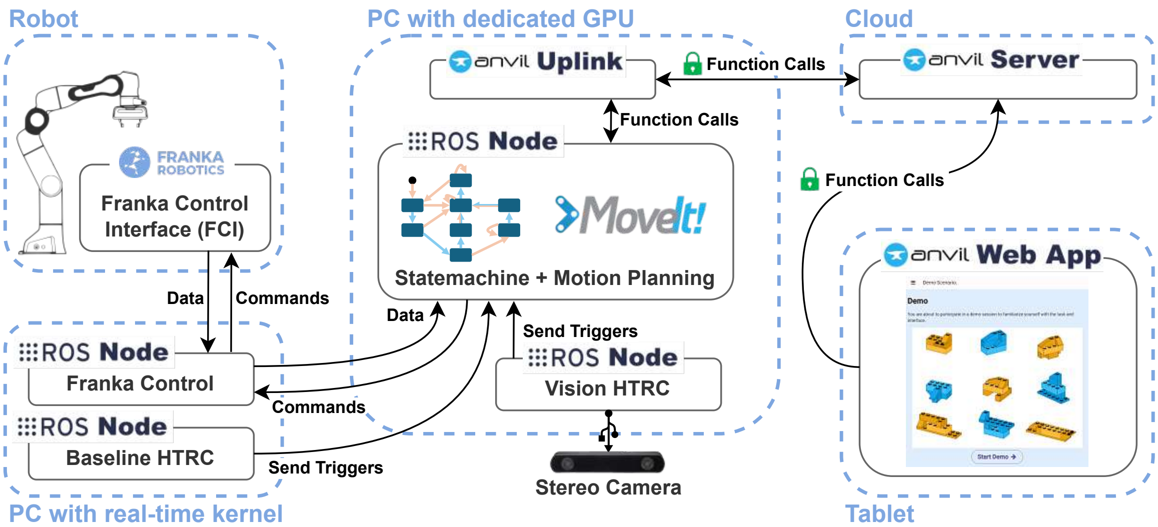
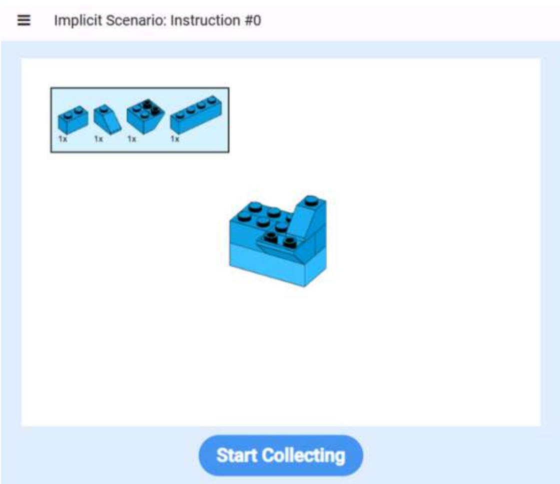


Figure 4.6: Software Architecture

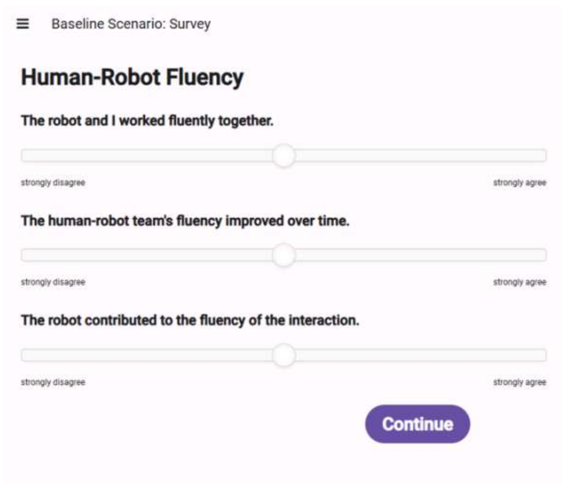
In principle, all ROS nodes can be run on the same machine, however, due to different system requirements in regard to performance and real-time priorities for the node communicating with libfranka and the vision node, the implemented architecture splits the discussed nodes on two PCs connected via an Ethernet cable. By default, ROS is configured as a distributed system allowing nodes to communicate over a network open to any device. Therefore, communication between individual ROS nodes is not encrypted. In contrast to that, communication between the anvil web app, server, and uplink must be encrypted as traffic is routed via the public Internet.

4.2.3 Study Interface

To guide the participants trough the experiment a custom anvil web app is developed. This web app is responsible for: (a) Showing the instructions of which Lego design to build. (see figure 4.7a) (b) Relaying commands to the robot, allowing it to resume its tasks and update the active storage accordingly. (c) Administering the survey by displaying the questions and enabling participants to respond directly within the web app, while also logging the timing of each response (see figure 4.7b). (d) Handling the task demonstration process, ensuring that participants become familiar with the tasks. (e) Demoing the different HTRC channels, by showing a short animation of how it works and allowing participants to test the mode before beginning with the experiment. (f) Setting up the experiment, which includes counterbalancing the assignment of scenarios, randomizing the order of instructions, and pseudo-randomly determining when an interruption should occur. (g) Modifying the base image for the selected Lego design according to the assigned color. This is necessary as every instruction has only one image and the colors can be modified as needed once the image is requested. (h) Logging the robot's actions and associating them with the correct participant, scenario, and instruction.



(a)



(b)

Figure 4.7: (a) Example Instruction Screen, (b) Example Survey Screen,

4.2.4 Implementation of Communication Channels

In accordance with the discussion above, there are multiple options to implement different HTRC channels to enable mid-task interruptions and task takeovers. Since the applicable communication channel depends on the active scenario, the following paragraphs discuss the implementation details for each level of the predictor construct.

Baseline

In order to include the current industry standard of how a human can interact with a robot, a simple haptic control element is used as a baseline. In contrast to a regular control interface, however, interrupting the task will not result in a cumbersome restart procedure as the interruption behavior is set to be the same as for the other scenarios. For the first design iteration, depicted in figure 4.2, the haptic control element is implemented as a separate ROS node monitoring for the push of the space key on the PC located on the left in figure 4.2.

Explicit

As pointed out in section 3.2.1, using the built-in force limiting functionality to communicate with a cobot is currently under explored in research. The basic idea is quite simple though, as it enables the human to interrupt a robot's task by intentionally causing a collision, effectively signaling the robot to pause, stop, or modify its current task without the need for additional interfaces. A major drawback of relying on the built-in mechanism of force limiting are the predefined thresholds required for detecting a collision and the limitation to only be able to detect collisions while the robot is moving. However, by directly monitoring the active forces on the robot's end-effector, the thresholds can be fine-tuned, and the robot can also be enabled to detect haptic interactions while it is idling. To enable this force monitoring, the finite-state machine is equipped with a ROS-Subscriber to the robot's measured external forces and an algorithm to detect peaks exceeding a certain threshold.

Implicit

Building on the discussion of techniques enabling implicit HTRC (section 3.2.2), vision-based systems, such as cameras and stereo-cameras, are the preferred technology when it comes to motion-based non-verbal HTRC. These systems are favored due to their relatively low price point, ease of use, simple installation, and readily available frameworks to extract relevant features from the stream of images. Additionally, they are capable of capturing complex motion cues without the need for physical contact or specialized sensors. Given these advantages, a stereo-vision-based system is implemented.

The vision module runs as a separate ROS node on the PC equipped with a dedicated GPU (see figure 4.6), once it detects the human intent to interrupt, it sends a trigger signal to the finite state machine, which then depending on its current state can initiate the expected interruption behavior of the robot. The implicit technique implemented for this thesis uses a ZED2 stereo camera[1] to monitor the robot's environment. The camera's depth data is processed at runtime by the ZED-SDK, and using a neural network, the 3D skeleton for every human within its field of view is computed. Owing to the physical setup and study design, the intent-prediction algorithm can be held rather simple. Based on the skeleton data, a custom algorithm extracts the wrist and elbow key points and extrapolates a hand key point. This hand keypoint is then used to monitor if the human is trying to reach for an object. In case any monitored hand key point is within a predefined zone, the vision node sends the interrupt trigger to the state machine.

4.3 Piloting

After designing the experimental procedure, implementing the required systems, and before inviting real participants, piloting is an essential phase to refine and validate the study design [112]. By conducting a pilot study, researchers gain insights that can help identify unforeseen issues, such as unclear instructions or operational errors, that might not surface during the planning phase, ultimately ensuring a smoother execution during the actual study [112]. During the piloting for this study, several design aspects, but most notably (a) the physical layout and (b) the style of the baseline interruption were slightly adapted.

(a) Physical Layout: During the piloting sessions, participants reported minor hesitance when asked to pick blue Lego bricks due to their closer proximity to the robot (see figure 4.2). The initial layout discussed in section 4.1.5 was optimized for longer robot trajectories as well as robot motions closer to the human workspace to enable

[1]www.stereolabs.com/products/zed-2

quicker interruptions by the human. However, this design choice also led to a non-symmetric layout as the robot's mounting position as well as the chosen home pose were both located on the left end of the table and therefore closer to the human's workspace for picking blue Lego bricks (see figure 4.2). In order to maximize the layout's symmetry ideally the robot should be remounted, however, since remounting the robot would also influence other experiments conducted with it and require substantial modifications to the entire setup a good compromise, still increasing the symmetry, is found by adapting the robots home-pose as depicted in figure 4.8.

In addition to the layout change to increase symmetry, the positioning of the stereo camera is also adapted to mitigate false positives, where the vision system interrupts the robot even though the human did not intend to interrupt it. Further testing revealed the root cause of these false positives to be linked to the robot partially occluding one of the human's hands causing the neural network responsible for extracting the human skeleton to every now and then fuse this occluded human arm with the robot and therefore produce false estimations of where and what the human is doing. The modified position, as well as the new field of view, are depicted in figure 4.8.

(b) Style of Baseline Interruption: Initially, the baseline condition was implemented as a simple press of the spacebar on a keyboard. However, during piloting, it became obvious that following this approach would lead to distorted results as the participants were simply waiting for the robot to start moving and would immediately press the button, with very little effort. Therefore, to make the baseline a bit more cumbersome the participants are now required to press a virtual button using a mouse. To further increase the required attention the mouse is programmed to randomly move away from the button to force a realignment of cursor and button for each interruption.

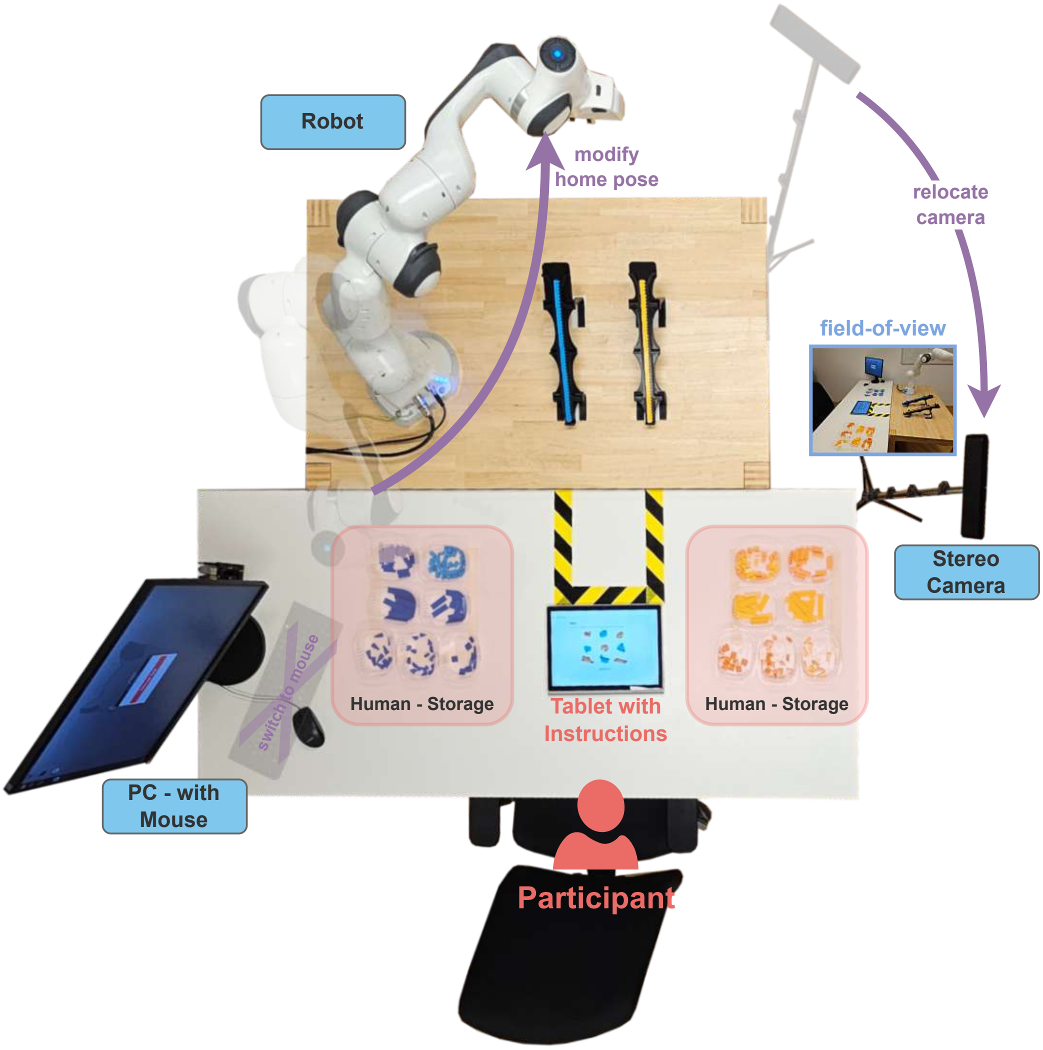


Figure 4.8: Second Iteration of Physical Layout

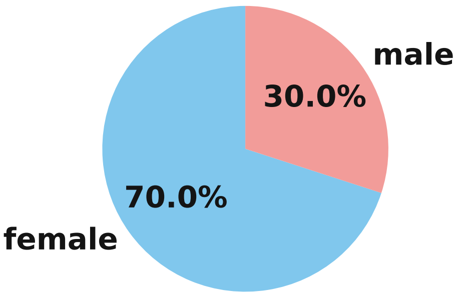
5 Results

The following chapter presents the findings of the experimental research, analyzing participants' responses to the questionnaires introduced in section 4.1.6. In total 23 participants were recruited to participate. Three of these were used as pilot sessions, therefore their responses are excluded, and the reported results are based on the 20 remaining participants. The first section 5.1 reports the results of the post-study questionnaire, followed by the second section 5.2 reporting the results of the repeatedly measured subjective fluency metric questionnaire, and the final section 5.3 summarizing the responses to the open-ended questions.

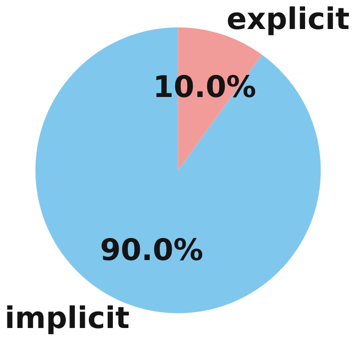
5.1 Demographics and Descriptive Statistics

The majority of the questions of the post-study questionnaire, such as age, gender, prior robot experience, and participants' most and least preferred channel of communication, can be analyzed via descriptive statistics. To calculate the statistics the open-source software JASP was utilized.

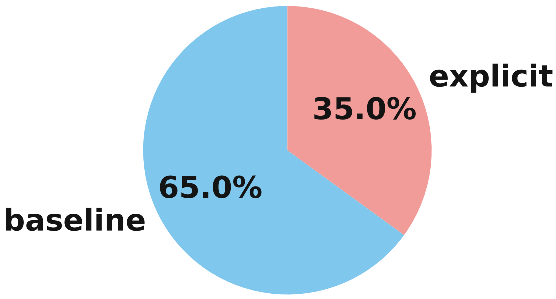
The recruited participants were, on average, $M = 26.3$ years old with a standard deviation of $SD = 2.452$. 70% identified as female and 30% as male, the remaining options "non-binary", and "prefer not to say" were not selected by any participant (see figure 5.1a). On a scale from one to ten, the self-reported prior experience was on average $M = 2.950$ with a standard deviation of $SD = 2.724$. In regard to preference, 90% reported the implicit, 10% the explicit, and none the baseline as their favored channel of communication (see figure 5.1b). The most disliked channel was the baseline, with 65% reporting it as their least favorite, followed by explicit, with 35% picking it as least preferred (see figure 5.1c). The implicit channel was never chosen as the least favorite.



(a)



(b)



(c)

Figure 5.1: (a) Gender Distribution, (b) Favorite Mode, and (c) Least Favorite Mode

5.2 Repeated Measures - Questionnaire

Based on the chosen participant structure (within-participant) (section 4.1.1) and the three levels of the predictor construct (explicit, implicit, and baseline) (section 4.1.2) the results of the subjective fluency metric questionnaire can be analyzed with the help of a Repeated Measurements Analysis of Variance (RM-ANOVA). However, the individual questions were not analyzed directly. Instead, compound scores, calculated as the mean of individual ratings, were utilized as input for the analysis. For each compound score, the assumption of sphericity was tested, and pairwise post hoc analysis using Holm correction was used to compare individual communication channels where the F-scores of the RM-ANOVA showed significant deviations. Additionally, using a Helmert coding contrast, the implicit channel's mean was compared with the means of explicit and baseline.

Human-Robot Fluency

The descriptive statistics for the perceived fluency compound scale are shown in table 5.1 and figure 5.2a. Based on Mauchly's test [116], the assumption of sphericity was violated ($W(2) = 0.607,\ X^2 = 8.974,\ p = 0.011,\ \epsilon = 0.718$). Therefore, the results were corrected using Greenhouse-Geisser's [117] sphericity correction. The calculated results indicate that perceived fluency differed significantly between HTRC channels with $F(1.436, 27.287) = 5.939,\ p = 0.013,\ \eta_p^2 = 0.238$. The compound value's distribution is plotted in figure 5.2b. A pairwise post hoc analysis using Holm correction showed that while there was no significant difference between explicit-vs-baseline channels ($p_{Holm} = 0.479,\ d = 0.206$), perceived fluency was significantly different for implicit-vs-baseline ($p_{Holm} = 0.007,\ d = 0.944$) and implicit-vs-explicit ($p_{Holm} = 0.029,\ d = 0.738$) channels. The calculated Helmert contrast indicated significance with $p = 0.002$. Based on these results, the hypothesis *H1* must be rejected, whereas *H2* can be confirmed.

Communication Channel	Mean	SD
implicit	8.033	1.213
explicit	6.900	1.734
baseline	6.583	1.611

Table 5.1: Descriptive Statistics for Human Robot Fluency

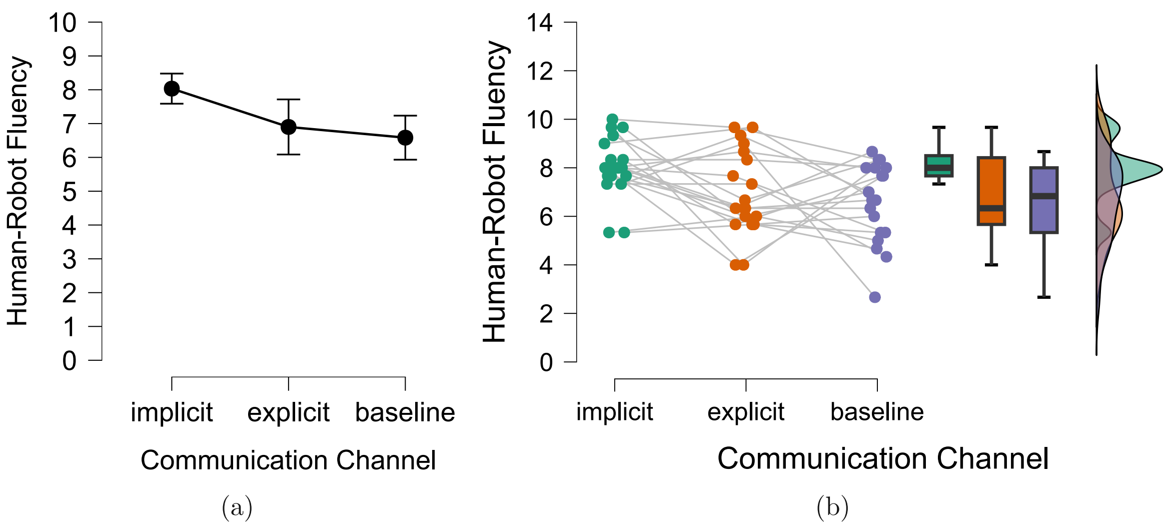


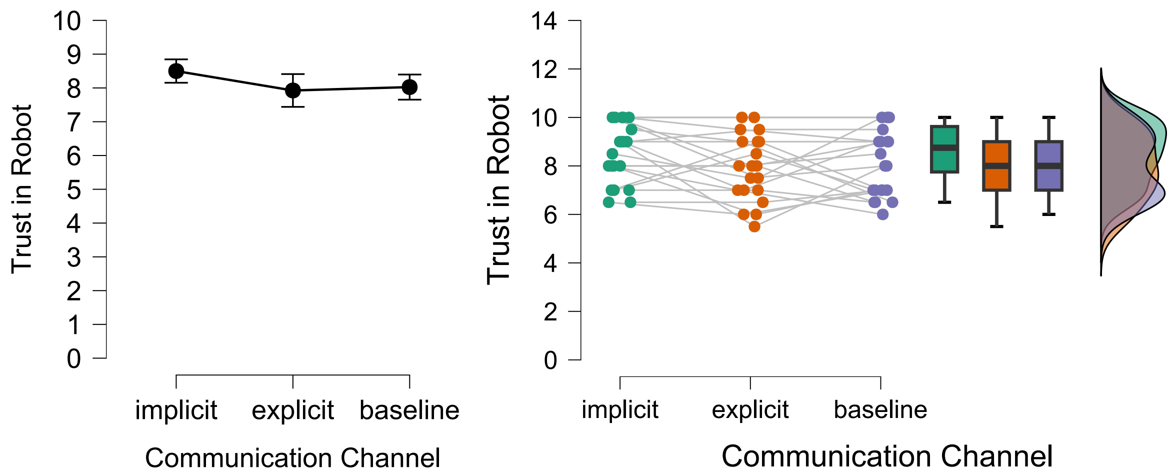
Figure 5.2: Human-Robot Fluency: (a) Descriptive Plot with 95% Confidence Intervals, (b) Compound Score Raincloud Plot

Trust in Robot

RM-ANOVA results suggested that the human's trust in the robot did not vary notably across HTRC channels, with $F(2,38) = 2.512$, $p = 0.095$, $\eta_p^2 = 0.117$. Therefore, hypothesis *H3* must be rejected in favor of its alternative. However, the Helmert contrast indicated a significant difference between implicit and the other channels with $p = 0.033$. The descriptives and raincloud distribution of trust levels are illustrated in table 5.2 and figure 5.3.

Communication Channel	Mean	SD
implicit	8.500	1.235
explicit	7.925	1.370
baseline	8.025	1.235

Table 5.2: Descriptive Statistics for Trust in Robot



(a) (b)

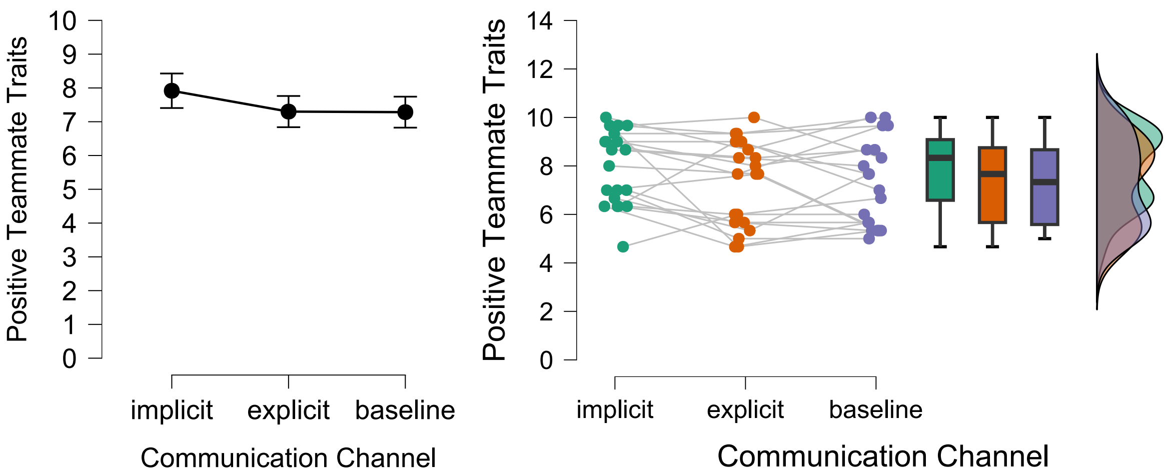
Figure 5.3: Trust in Robot: (a) Descriptive Plot with 95% Confidence Intervals, (b) Compound Score Raincloud Plot

Positive Teammate Traits

Findings indicated no significant variation for the compound score of positive teammate traits, with $F(2, 38) = 2.500$, $p = 0.095$, $\eta_p^2 = 0.116$. Figure 5.4a and table 5.3 visualize the descriptive statistics, and figure 5.4b the compound score's raincloud distribution. Similar to the trust score, a Helmert contrast reports the positive teammate traits compound value for the implicit condition to be significantly different from the rest with $p = 0.031$.

Communication Channel	Mean	SD
implicit	7.917	1.573
explicit	7.300	1.747
baseline	7.283	1.771

Table 5.3: Descriptive Statistics for Positive Teammate Traits



(a) (b)

Figure 5.4: Positive Teammate Traits: (a) Descriptive Plot with 95% Confidence Intervals, (b) Compound Score Raincloud Plot

Working Alliance - Bond

The distribution for implicit, explicit, and baseline scores (see Table 5.4 and Figure 5.5a), demonstrated a statistically significant difference with $F(2, 38) = 11.788$, $p \leq 0.001$, $\eta_p^2 = 0.383$. Figure 5.5b displays the raincloud distribution pattern. Post hoc analysis with Holm correction indicated considerable differences between implicit-vs-baseline ($p_{Holm} < 0.001, d = 0.478$) as well as explicit-vs-baseline ($p_{Holm} = 0.012, d = 0.289$) communication channels, implicit-vs-explicit ($p_{Holm} = 0.064$, $d = 0.189$) indicated no statistically difference. The Helmert contrast between implicit and the other two channels was also significant at $p \leq 0.001$.

Communication Channel	Mean	SD
implicit	5.614	1.935
explicit	5.264	1.828
baseline	4.729	1.787

Table 5.4: Descriptive Statistics for Working Alliance - Bond

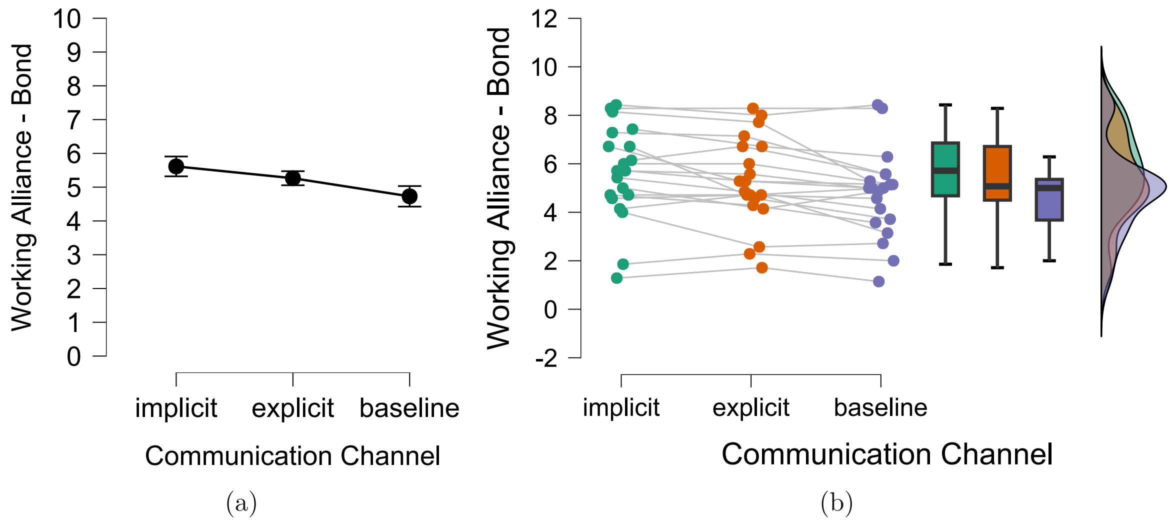


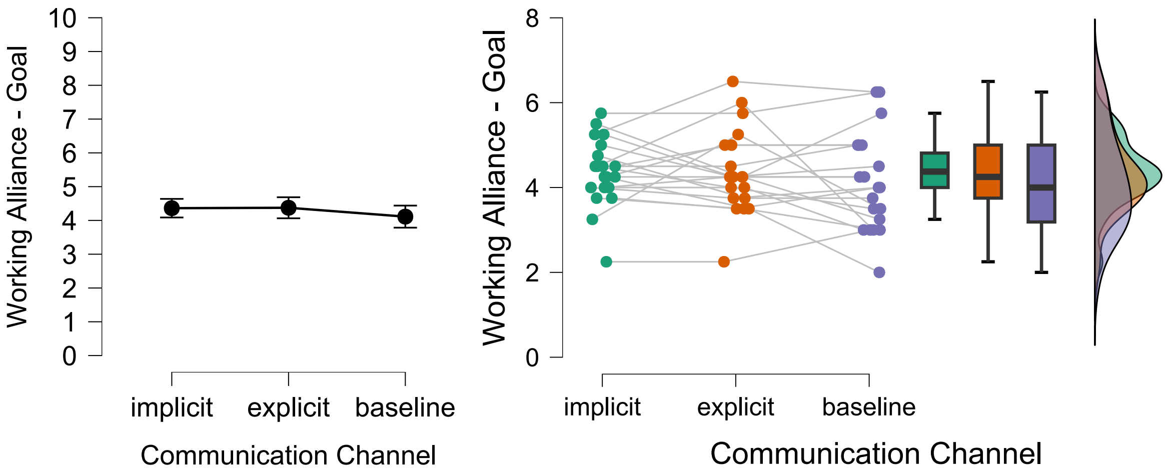
Figure 5.5: Working Alliance - Bond: (a) Descriptive Plot with 95% Confidence Intervals, (b) Compound Score Raincloud Plot

Working Alliance - Goal

In regard to goal alignment, the analysis helped identify no significant differences with $F(2, 38) = 1.031,\ p = 0.051,\ \eta_p^2 = 0.366$. The descriptives are presented in table 5.5 and figure 5.6a, and the raincloud distribution in figure 5.6b. Unlike the previous compound values, the goal alignment scores showed no notable difference of means when compared via a Helmert contrast table.

Communication Channel	Mean	SD
implicit	4.362	0.805
explicit	4.375	0.988
baseline	4.112	1.157

Table 5.5: Descriptive Statistics for Working Alliance - Goal



(a) (b)

Figure 5.6: Working Alliance - Goal: (a) Descriptive Plot with 95% Confidence Intervals, (b) Compound Score Raincloud Plot

Individual Liking

The individual liking scores, demonstrated a statistically significant difference between conditions ($F(2,38) = 21.442,\ p < 0.001,\ \eta_p^2 = 0.530$). Table 5.6 and figure 5.7 depict the descriptive statistics, and figure 5.7 the distribution of acceptance across channels. Holm-adjusted post hoc analysis found significant differences between implicit-vs-baseline ($p_{Holm} < 0.001, d = 2.125$), implicit-vs-explicit ($p_{Holm} < 0.001, d = 1.431$), and explicit-vs-baseline ($p_{Holm} = 0.043, d = -0.693$). The calculated Helmert contrast between implicit and the other two channels was significant at $p \leq 0.001$.

Communication Channel	Mean	SD
implicit	8.850	1.309
explicit	5.650	2.796
baseline	4.100	2.337

Table 5.6: Descriptive Statistics for Individual Liking

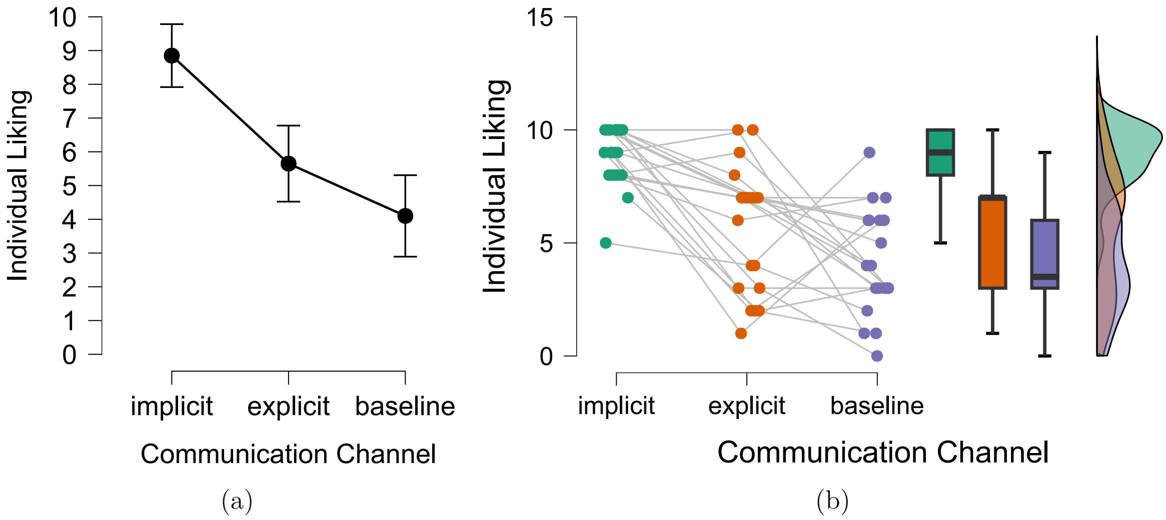
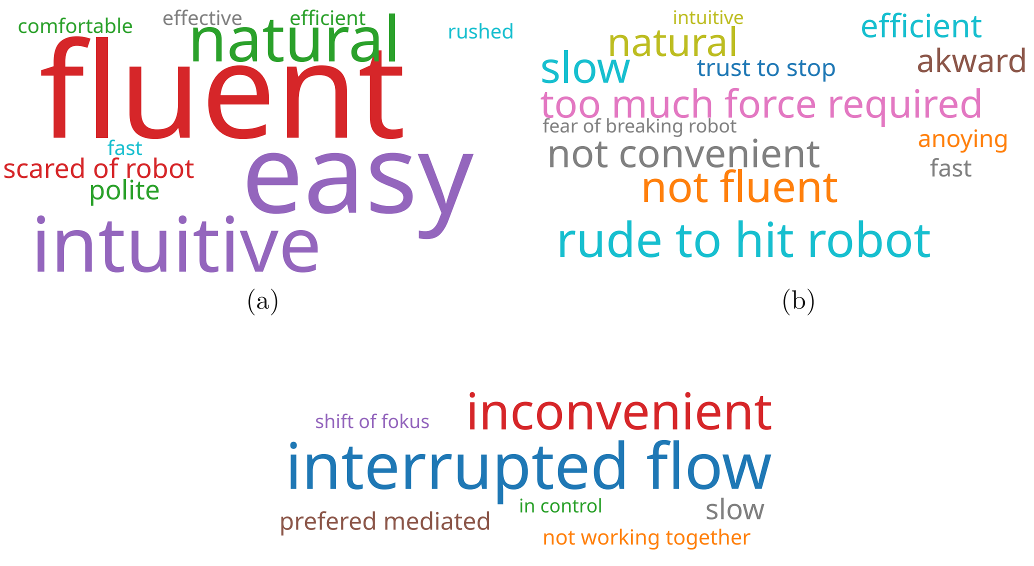


Figure 5.7: Individual Liking: (a) Descriptive Plot with 95% Confidence Intervals, (b) Compound Score Raincloud Plot

5.3 Open-Ended

To analyze the responses to the open-ended question, participants' comments were thoroughly read to gain a general sense of common themes. The most common themes were coded for each channel of communication and are shown in Figure 5.8 (a-c).



(a) (b) (c)

Figure 5.8: coded common themes for (a) Implicit, (b) Explicit, and (c) Baseline channels

Overall, participants reported the implicit channel of communication to be intuitive, comfortable, and as having a good flow. Comments included "it was a fluent interaction", "this mode requires least effort and is the smoothest", or "the flow of the work was good". Some participants, however, were a bit scared of the robot - "I found it a little bit scary" - , and reported being hesitant when prompted to interrupt as they were afraid of causing a collision - "it seemed like we could collide". Despite that, the majority of responses suggest that implicit communication allowed the most fluent and most natural collaboration.

In contrast to that, while some participants still experienced the explicit scenario as intuitive, most felt awkward, unpolite, or afraid to damage the robot when asked to interrupt it by physically touching it. Comments ranged from "it felt a little bit rude and impolite", "Rough, not very elegant", or "a bit brutal but manageable'" to "it's comfortable knowing the robot stops when you want it to", "Quick and effective", or " it felt more natural'". Additionally, some participants noted that they would have liked to receive more training on how to properly physically interrupt the robot, and if the robot was more sensitive in detecting interruptions.

The baseline scenario was primarily perceived as an inconvenient extra task that interrupted the workflow and was uncomfortable due to the requirement to shift attention toward the mouse and screen. The majority of participants suggested rearranging the study's physical layout to bring the mouse closer to their workspace. Even though some participants noted familiarity with this modality, it was never the preferred channel of communication. Comments included "it interrupted me in my workflow", "this way was uncomfortable", or "didn't feel like we are working together".

5.3.1 Additional Observations

In addition to the quantitative and qualitative data captured with the questionnaire, the experimenter took notes of additional observations while administering the experiment. During the explicit condition, the most common observation was people being hesitant to grasp the correct brick after changing their minds. Some even waited as long as it took the robot to return to its safe pose before collecting the brick from the robot-storage. Another interesting observation, seen across all conditions, was people enjoying it when the robot had to return the already grasped brick, this manifested in participants waiting to interrupt until the robot had grasped a Lego. During the implicit condition, this waiting and seeing sometimes turned into participants actively trying to fool the robot by moving extra fast and just seconds before it would deliver the brick to them.

6 Discussion

As a sub-field of Human-Robot Interaction (HRI), Human-Robot Collaboration (HRC) largely focuses on the division of labor between humans and robots. Current industry standards and research typically define task boundaries based on spatial and temporal separation of humans and robots or based on task interdependencies. However, when humans and robots collaborate, strict task boundaries can reduce team adaptability and flexibility, especially when a human needs to interrupt a robot during the execution of a task. The concept of task takeovers turns out to have not been previously researched in the domain of HRC. However, lending from other collaborative systems, like autonomous driving, it can be defined as the dynamic transfer of control between agents. The missing key components of enabling humans to take over tasks currently executed by a robot are: (1) a task representation capable of handling interruptions and (2) a way to communicate the intent to interrupt. This thesis, therefore, intended to bridge this gap by exploring how interruptions can be handled and how different Human to Robot Communication (HTRC) channels influence the team dynamics like perceived team fluency and trust during task takeovers. Specifically, it addressed the following research questions:

Q1.) Which communication theories can be used to model a takeover request?

In order to understand how to effectively communicate a human's takeover request to a robot, it turns out to be essential to consider the underlying theoretical communication framework and the channel used to communicate. Theories on human-robot communication often draw on human-to-human communication theories such as Watzlawick's five axioms [51], Laswell's linear framework [57], Barnlund's transactional model [61] or Kincaid's convergence model [60] (see section 2.4). The existing research was found to include various studies on HRI specific frameworks focused on adapting transactional models to better depict the reciprocal aspect of communication (see section 2.4.4). However, some researchers like Frijns et al. [21] also argue that with current technologies, communication between robots and humans can not be symmetrical due to inherent differences in capabilities. This is especially true for non-humanoid robots which lack human-like skills such as displaying and perceiving emotions.

Furthermore, unlike ongoing dialogues that benefit from using concepts of mutual understanding, feedback loops, and dynamic agent roles, takeover scenarios can be viewed as a discrete communication of intent and a simple response of acknowledgment, e.g., the robot interrupting its task. Therefore, this thesis suggests discretizing the required communication for interrupting a robot and aligning it with the linear approach suggested by Kunold et al. [50], as this allows for simplified modeling and analysis of communication acts by focusing on clear, actionable exchanges. Applying this framework to the context of human-robot task takeovers, the human can be seen as a communicator sending an instrumental message to the robot with the intentional effect of coordinating task-related actions. At the center of this discrete act of communication, the channel, comprised of form and modality, can be viewed as crucial in determining how the takeover request gets communicated.

Q2.) Which channels of communication can be utilized to communicate a task-takeover request?

As outlined in section 3.2, various channels can be used to convey a human's takeover request to a robot. These can generally be categorized as either (1) explicit, where the human actively invokes something, or (2) implicit, where the robot tries to infer human intent based on the user's actions.

Explicit communication was found to be accomplishable via multiple combinations of form and modality. However, not all of these channels and their respective technologies turned out to be equally suited for deployment. For example, while microphones could offer a natural *verbal-auditive* channel, their effectiveness is still highly dependent on factors such as trained vocabulary, the number of speakers, and the level of background noise. Other channels like a *nonverbal-neural*, based on BCI technology, were found to require extensive training and adaptation to each individual user, limiting their immediate applicability. Similarly, a *nonverbal-visual* channel based on cameras or a *nonverbal-kinesthetic* channel utilizing motion sensors, both designed to detect gestures, turned out to require the user to know and learn which gestures could be used while also being limited by the available training data. In contrast to these more complex options, a *nonverbal-tactile* channel facilitated by simple haptic control elements, although often stationary, was found to be a well-established way of sending command messages to a robot, e.g., via its teach pendant. Additionally, the use of internal force/torque monitoring turned out to be a viable, less explored but promising *nonverbal-tactile* channel. While modern cobots are known to employ force and torque limiting primarily to adhere to safety standards, its potential for detecting explicit human input during task execution appeared to be largely unexplored in research.

Following Watzlawick's axiom that "one cannot not communicate" [51], the notion of communication-free collaboration must be reconsidered. Even in implicit interaction scenarios, human actions serve as communicative signals that the robot must correctly interpret to ensure successful collaboration. However, unlike explicit HTRC, implicit communication turned out to be limited to certain channels and their respective technologies. The examination of previous research revealed the most common implicit channels as either (a) *nonverbal-visual* based on 2D or 3D camera systems or (b) *nonverbal-kinesthetic* enabled by motion sensors. Most reviewed implicit techniques were found to require some sort of machine learning algorithms to analyze human motion patterns to either recognize and label or predict future human motion. In general, implicit channels appeared to enable a more natural interaction, as humans can communicate without interrupting their workflow. However, they require robust algorithms to accurately interpret/predict human actions to avoid unintended takeovers.

While the discussed asymmetries in humans and robots allow for structured and linear communication, they may not always align with human expectations for natural interaction. This is particularly relevant when using simplified interfaces such as GUIs or simple buttons, which lack the fluidity of human-like exchanges. Nonetheless, they can serve as an ideal baseline condition for comparison, as they provide a clear, controlled, and well-established method for communicating with a robot.

Based on the discussion above, the three different channels (a) *baseline*, mediated via a GUI and mouse, (b) *explicit*, facilitated by physically touching the robot, and (c) *implicit*, based on human actions, where chosen to investigate how different Human to Robot Communication (HTRC) channels affect task takeovers.

Q3.) What are current task allocation techniques, and how do they integrate task-takeovers and communication?

The analysis of prior literature revealed numerous techniques, none of which have previously been used for the proposed concept of task takeovers. The four major categories of dynamic task allocation methods identified were:

(a) AND/OR Graphs (AOGs): Modern AOGs provide a directed graph representation of tasks with alternative execution paths capable of online re-planning. Despite these capabilities, the reviewed techniques must be criticized for assuming optimal human behavior and limiting humans to a passive role. Moreover, communication in the analyzed papers was often limited to GUIs or facilitated via gestures. (section 3.1.1)

(b) Behavior Trees (BTs): The concept of BTs originated from game design, transferred to HRC it offers a modular technique to handle task allocation. While the analyzed method excels in handling low-level ad-hoc decisions, it lacks the capability to consider hierarchical orders and human-robot communication. (section 3.1.2)

(c) Hierarchical Task Networks (HTNs): To increase understandability for humans, HTNs decompose complex tasks into smaller subtasks. In order to optimize task completion time and minimize spatial interference, the reviewed HTN methods prioritize robot actions parallel to human actions. However, similar to AOGs and BTs, the examined methods did not account for unexpected human behavior or task takeovers. In the HTN-based systems, communication was primarily implicit and relied on visual-nonverbal channels by recognizing or predicting human actions. (section 3.1.3)

(d) Partially Observable Markov Decision Processes (POMDPs): The reviewed POMDPs provide a probabilistic framework for decision-making under uncertainty where the (human-robot) system's state cannot be observed directly. In the analyzed literature, POMDPs have been utilized to minimize task completion time and to give robots the capability to reason about when to communicate. However, POMDPs require numerous assumptions about human behavior and extensive modeling expertise. Additionally, they have yet to be tested for their applicability to handle task interruptions. (section 3.1.4)

A key limitation of all analyzed techniques was their assumption that once a task is assigned, it must also be completed, making it difficult to integrate takeovers within existing frameworks. Even though some task allocation techniques could, in theory, be adapted to accommodate takeovers, none of the reviewed methods offered a native mechanism for handling purposeful human interruptions. Consequently, as an alternative strategy to handle interruptions, a finite-state-machine approach was chosen to speed up development. In addition to being conceptually similar to standard robot programming, where multiple commands are sequentially linked, a finite-state machine allows for the creation of more complex scenarios by constructing a network of states and transitions.

Q4.) To what extent do different HTRC channels impact team dynamics, such as perceived team fluency and trust, during task-takeovers?

In order to answer research question Q4 an experimental setup was developed, and a user study was conducted with a total of 20 participants. The quantitative results of the experiment suggest considerable differences in how HTRC channels affect task takeovers (see section 5.2). The implicit channel scored highest in all subjective compound metrics, with the explicit channel typically coming in second and the baseline third. The conducted Repeated Measurements Analysis of Variance (RM-ANOVA) showed statistical significance for the "Human-Robot Fluency", "Working Alliance - Bond" and "Individual Liking" scores. Even though it might seem counterintuitive, the remaining scores "Trust in Robot", "Positive Teammate Traits", and "Working Alliance - Goal" showed no notable effect associated with HTRC channels. This indicates that the level of trust during the takeover scenario was not influenced by how the human communicated the intention to take over. Similarly, it can be speculated that the robot's traits, such as perceived intelligence and commitment to the task, were not affected by the communication channel.

Overall, the quantitative results align with the qualitative results based on the answers to the open-ended questions. All findings suggested that the implicit HTRC channel is not only the most liked channel but also perceived as having the highest team fluency. One possible explanation for this is the implicit channel's intuitiveness, as it requires no training or familiarization, while also being the least disruptive channel to the human's workflow, enabling almost a human-co-worker-like dynamic that supports task fluency. The explicit HTRC channel's results, on the other hand, were a bit surprising, as it was generally perceived as less fluent and encountered usability challenges that may have affected the scores. The primary usability issues seem to stem from participants feeling awkward and rude about physically interrupting the robot. Additionally, the force thresholds may have been too high, and participants might have lacked sufficient experience to judge how much force was required. In contrast to that, the baseline HTRC channel's poor performance and its overall bad ratings proved to be within expectations, given the inherent limitations of its design. The fact that almost all participants highlighted the inconvenience of having to use a mouse to click a virtual button, emphasizes how interactions mediated via a GUI disrupt the workflow and require a shift of focus, thus reducing perceived task fluency.

Limitations

Like most experimental research, the generalizability and validity of this study's results have several limitations. First of all, the statistical power and transferability are constrained by the small sample size of 20 participants. Moreover, the decision to conduct the experiments in a controlled laboratory setting with a simple Lego® assembly task can not fully reflect the complexity of real-world scenarios. However, the controlled laboratory setting and the abstract task of assembling Legos allowed for better isolation of key variables, minimized external noise, and increased experimental control.

Besides these limitations associated with the study design, it is also necessary to address a number of technical restrictions. Firstly, the explicit mode suffered from the robot's internal force/torque limits being fixed at a relatively high value. Unfortunately, lowering these thresholds results in random false positives as the robot moves and is exposed to the dynamic forces imposed by its physical embodiment. For the explicit condition, the robot's physical location must be named as an additional limiting factor, as it caused participants to delay their interruptions until the robot was closer and within a more convenient proximity to interrupt. Last but not least, the implicit channel's basic logic to infer human intent could benefit from enhanced reasoning capabilities.

Future Research

The discussion above highlights the need for future research on how to handle and communicate task takeovers in collaborative human-robot scenarios. One valuable area of future research should address the limitations of the study's design by expanding the participant pool and exploring more complex and realistic scenarios. This will increase generalizability and enhance the statistical power and reliability of future findings.

In order to tackle the responsiveness issue caused by the robot's internal force/torque limits, future work could look into using ML-enabled algorithms to detect anomalies instead. This could increase the detection accuracy and make the explicit mode work faster as the robot would no longer be required to recover from the collision. Similarly, the implicit channel could be refined through advanced intent recognition and prediction capabilities such as constantly monitoring hand velocity and trajectory to perform predictions of future hand movements and, for example, only trigger an interruption if the robot and the human are about to grasp the same object.

New dimensions for future work could investigate the long-term effects of different HTRC channels and examine the impact of varying task complexity. Another possible area of future research to increase objectivity could be the incorporation of objective metrics, such as eye tracking, to quantify human attention and cognition during task takeovers. Moreover, future research could focus on integrating multimodal communication approaches to offer richer interactions. In addition to these new dimensions, future work could incorporate a multi-factorial design and combine HTRC with Robot to Human Communication (RTHC) to explore the reciprocal effects of communication and provide a more holistic view.

7 Conclusion

This thesis introduced the research domain of collaborative robotics to the concept of task takeovers, where a human momentarily interrupts a robot to assume control of its task. Within a linear framework of human-robot communication, it then explored the effects of selected Human to Robot Communication (HTRC) channels on team dynamics such as perceived intelligence, fluency, trust, liking, goal alignment, and human-robot bond during task takeovers by conducting an experimental user study. Within the study three HTRC channels were examined: (a) *baseline*, mediated by an Graphical User Interface (GUI) on a PC, (b) *explicit*, facilitated by physical touching the robot, and (c) *implicit*, based on inferring intent from human motion.

The analysis of the results suggests that the implicit channel allowed for smoother and more natural task takeover as it was perceived as significantly more fluent and received higher individual liking ratings than the other two channels. While the explicit condition was able to achieve significantly higher individual liking scores than the baseline channel, no notable difference was observable in regard to perceived fluency. Contrary to expectations, the trust in the robot was not significantly affected by switching HTRC channel. This suggests that the human's trust in the robot's capabilities to handle a task takeover might not be linked to the channel used to communicate the intent to interrupt. Similarly, perceived robot traits such as task commitment and the robot's intelligence, appeared to be unrelated to the chosen HTRC channel.

This thesis' findings provide valuable insights for developing and designing future robotic systems equipped with intuitive and fluent human-robot interactions capable of handling frequent interruptions and task takeovers.

List of Figures

List of Tables

Bibliography

[1] Y. Cheng, L. Sun and M. Tomizuka, "Human-Aware Robot Task Planning Based on a Hierarchical Task Model," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 1136–1143, Apr. 2021. DOI: 10.1109/LRA.2021.3056370.

[2] R. Zhang, J. Li, P. Zheng, Y. Lu, J. Bao and X. Sun, "A fusion-based spiking neural network approach for predicting collaboration request in human-robot collaboration," *Robotics and Computer-Integrated Manufacturing*, vol. 78, p. 102 383, Dec. 2022. DOI: 10.1016/j.rcim.2022.102383.

[3] K. Ramachandruni, C. Kent and S. Chernova, "UHTP: A User-Aware Hierarchical Task Planning Framework for Communication-Free, Mutually-Adaptive Human-Robot Collaboration," en, *ACM Transactions on Human-Robot Interaction*, vol. 13, no. 3, pp. 1–27, Sep. 2024. DOI: 10.1145/3623387.

[4] Fitts, Paul M, "Human engineering for an effective air-navigation and traffic-control system," *National Research Council*, 1951.

[5] I. Georgievski and M. Aiello, "HTN planning: Overview, comparison, and beyond," *Artificial Intelligence*, vol. 222, pp. 124–156, May 2015, ISSN: 00043702. DOI: 10.1016/j.artint.2015.02.002.

[6] T. Kiyokawa, N. Shirakura, Z. Wang *et al.*, "Difficulty and complexity definitions for assembly task allocation and assignment in Human–Robot collaborations: A review," *Robotics and Computer-Integrated Manufacturing*, vol. 84, p. 102 598, Dec. 2023. DOI: 10.1016/j.rcim.2023.102598.

[7] A. Tausch, A. Kluge and L. Adolph, "Psychological Effects of the Allocation Process in Human–Robot Interaction – A Model for Research on ad hoc Task Allocation," *Frontiers in Psychology*, vol. 11, 2020. DOI: 10.3389/fpsyg.2020.564672.

[8] R. Salehzadeh, J. Gong and N. Jalili, "Purposeful Communication in Human–Robot Collaboration: A Review of Modern Approaches in Manufacturing," *IEEE access : practical innovations, open solutions*, vol. 10, pp. 129 344–129 361, 2022. DOI: 10.1109/ACCESS.2022.3227049.

[9] T. Hellström and S. Bensch, "Understandable robots - What, Why, and How," *Paladyn, Journal of Behavioral Robotics*, vol. 9, no. 1, pp. 110–123, Jul. 2018, Publisher: De Gruyter Open Access. DOI: 10.1515/pjbr-2018-0009.

[10] G. A. Boy, Ed., *The handbook of human-machine interaction: A human-centered design approach*. Boca Raton; London New York: CRC Press, Taylor & Francis Group, 2011, ISBN: 978-0-7546-7580-8.

[11] M. Nardo, D. Forino and T. Murino, "The evolution of Man–Machine interaction: The role of human in Industry 4.0 paradigm," *Production & Manufacturing Research*, vol. 8, no. 1, pp. 20–34, Jan. 2020. DOI: 10.1080/21693277.2020.1737592.

[12] H. Ayanoğlu and E. Duarte, Eds., *Emotional Design in Human-Robot Interaction: Theory, Methods and Applications* (Human–Computer Interaction Series). Cham: Springer International Publishing, 2019. DOI: 10.1007/978-3-319-96722-6.

[13] J. Burke, R. Murphy, E. Rogers, V. Lumelsky and J. Scholtz, "Final Report for the DARPA/NSF Interdisciplinary Study on Human–Robot Interaction," *IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews)*, vol. 34, no. 2, pp. 103–112, May 2004. DOI: 10.1109/TSMCC.2004.826287.

[14] B. H. Kantowitz and R. D. Sorkin, *Human Factors: Understanding People-System Relationships*. Wiley, Mar. 1983, ISBN: 978-0-471-09594-1.

[15] S. K. Card, T. P. Moran and A. Newell, *The Psychology of Human-computer Interaction*. L. Erlbaum Associates, 1983, ISBN: 978-0-89859-243-6.

[16] S. Schleidgen, O. Friedrich, S. Gerlek, G. Assadi and J. Seifert, "The concept of "interaction" in debates on human–machine interaction," *Humanities and Social Sciences Communications*, vol. 10, no. 1, p. 551, Sep. 2023. DOI: 10.1057/s41599-023-02060-8.

[17] K. Capek, *R.U.R.: Rossum's Universal Robots*, Englisch, trans. by D. Wyllie. Wildside Press, Sep. 2010, ISBN: 978-1-55742-255-2.

[18] J. G. C. Devol, "Programmed article transfer," US2988237A, Jun. 1961.

[19] A. Bonarini, “Communication in Human-Robot Interaction,” en, *Current Robotics Reports*, vol. 1, no. 4, pp. 279–285, Dec. 2020. DOI: 10.1007/s43154-020-00026-1.

[20] L.-E. Janlert and E. Stolterman, “The Meaning of Interactivity—Some Proposals for Definitions and Measures,” *Human–Computer Interaction*, vol. 32, no. 3, pp. 103–138, May 2017, Publisher: Taylor & Francis. DOI: 10.1080/07370024.2016.1226139.

[21] H. A. Frijns, O. Schürer and S. T. Koeszegi, “Communication Models in Human–Robot Interaction: An Asymmetric MODel of ALterity in Human–Robot Interaction (AMODAL-HRI),” *International Journal of Social Robotics*, vol. 15, no. 3, pp. 473–500, Mar. 2023. DOI: 10.1007/s12369-021-00785-7.

[22] F. Hegel, S. Gieselmann, A. Peters, P. Holthaus and B. Wrede, “Towards a typology of meaningful signals and cues in social robotics,” in *2011 RO-MAN*, Atlanta, GA, USA: IEEE, Jul. 2011, pp. 72–78. DOI: 10.1109/ROMAN.2011.6005246.

[23] S. Bensch, A. Jevtić and T. Hellström, “On Interaction Quality in Human-Robot Interaction:” in *Proceedings of the 9th International Conference on Agents and Artificial Intelligence*, Porto, Portugal: SCITEPRESS - Science and Technology Publications, 2017, pp. 182–189. DOI: 10.5220/0006191601820189.

[24] N. C. Krämer, A. von der Pütten and S. Eimler, “Human-Agent and Human-Robot Interaction Theory: Similarities to and Differences from Human-Human Interaction,” in *Human-Computer Interaction: The Agency Perspective*, M. Zacarias and J. V. de Oliveira, Eds., Berlin, Heidelberg: Springer, 2012, pp. 215–240. DOI: 10.1007/978-3-642-25691-2_9. (visited on 24/03/2024).

[25] D. Premack and G. Woodruff, “Does the chimpanzee have a theory of mind?” *Behavioral and Brain Sciences*, vol. 1, no. 4, pp. 515–526, Dec. 1978. DOI: 10.1017/S0140525X00076512.

[26] C. Vesper, E. Abramova, J. Bütepage *et al.*, “Joint Action: Mental Representations, Shared Information and General Mechanisms for Coordinating with Others,” *Frontiers in Psychology*, vol. 07, Jan. 2017. DOI: 10.3389/fpsyg.2016.02039.

[27] R. Dale, R. Fusaroli, N. D. Duran and D. C. Richardson, “The Self-Organization of Human Interaction,” in *Psychology of Learning and Motivation*, vol. 59, Elsevier, 2013, pp. 43–95, ISBN: 978-0-12-407187-2. DOI: 10.1016/B978-0-12-407187-2.00002-2.

[28] F. Sherwani, M. M. Asad and B. Ibrahim, "Collaborative Robots and Industrial Revolution 4.0 (IR 4.0)," in *2020 International Conference on Emerging Trends in Smart Technologies (ICETST)*, Karachi, Pakistan, Mar. 2020, pp. 1–5. DOI: 10.1109/ICETST49965.2020.9080724.

[29] M. Ahmad, M. H. Khan, A. Bouabdallah, N. Zemoura and M. Imran, "Exploring Human-Robot Interaction and Collaboration for Real-World Applications," en, *IEEE-SEM*, vol. 12, no. 2, Feb. 2024, ISSN: 2320-9151. [Online]. Available: https://www.ieeesem.com/researchpaper/Exploring_Human_Robot_Interaction_and_Collaboration_for_Real_World_Applications.pdf (visited on 05/04/0204).

[30] A. Grau, M. Indri, L. Lo Bello and T. Sauter, "Robots in Industry: The Past, Present, and Future of a Growing Collaboration With Humans," *IEEE Industrial Electronics Magazine*, vol. 15, no. 1, pp. 50–61, Mar. 2021. DOI: 10.1109/MIE.2020.3008136.

[31] European Parliament: Directorate-General for Parliamentary Research Services and Gambao, Ernesto, *Analysis exploring risks and opportunities linked to the use of collaborative industrial robots in Europe.* European Parliament, 2023. DOI: 10.2861/021129.

[32] C. Petzoldt, M. Harms and M. Freitag, "Review of task allocation for human-robot collaboration in assembly," *International Journal of Computer Integrated Manufacturing*, vol. 36, no. 11, pp. 1675–1715, Nov. 2023. DOI: 10.1080/0951192X.2023.2204467.

[33] J. E. Colgate and M. A. Peshkin, "Cobots," US5952796A, Sep. 1999.

[34] W. Bauer, M. Bender, M. Braun, P. Rally and O. Scholtz, "Studie Leichtbauroboter," Fraunhofer - Institut für Arbeitswirtschaft und Organisation IAO, Stuttgart, Tech. Rep., 2016. [Online]. Available: https://www.engineering-produktion.iao.fraunhofer.de/content/dam/iao/tim/Bilder/Projekte/LBR/Studie-Leichtbauroboter-Fraunhofer-IAO-2016.pdf (visited on 28/04/2024).

[35] R. Müller, M. Vette and O. Mailahn, "Process-oriented Task Assignment for Assembly Processes with Human-robot Interaction," *Procedia CIRP*, vol. 44, pp. 210–215, 2016. DOI: 10.1016/j.procir.2016.02.080.

[36] S. El Zaatari, M. Marei, W. Li and Z. Usman, "Cobot programming for collaborative industrial tasks: An overview," en, *Robotics and Autonomous Systems*, vol. 116, pp. 162–180, Jun. 2019. DOI: 10.1016/j.robot.2019.03.003.

[37] M. Bernabei and F. Costantino, "Adaptive automation: Status of research and future challenges," *Robotics and Computer-Integrated Manufacturing*, vol. 88, p. 102 724, Aug. 2024. DOI: 10.1016/j.rcim.2024.102724.

[38] B. Hayes and B. Scassellati, "Challenges in Shared-Environment Human-Robot Collaboration," en, in *ACM/IEEE International Conference on Human-Robot Interaction*, Jan. 2013. [Online]. Available: https://scazlab.yale.edu/sites/default/files/files/hayes_challenges_HRI_collab_2013(1).pdf (visited on 02/05/2024).

[39] C. Schmidbauer, S. Schlund, T. B. Ionescu and B. Hader, "Adaptive Task Sharing in Human-Robot Interaction in Assembly," in *2020 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, Dec. 2020, pp. 546–550. DOI: 10.1109/IEEM45057.2020.9309971.

[40] F. Ranz, V. Hummel and W. Sihn, "Capability-based Task Allocation in Human-robot Collaboration," *Procedia Manufacturing*, 7th Conference on Learning Factories, CLF 2017, vol. 9, pp. 182–189, Jan. 2017, ISSN: 2351-9789. DOI: 10.1016/j.promfg.2017.04.011.

[41] T. Bänziger, A. Kunz and K. Wegener, "Optimizing human–robot task allocation using a simulation tool based on standardized work descriptions," *Journal of Intelligent Manufacturing*, vol. 31, no. 7, pp. 1635–1648, Oct. 2020. DOI: 10.1007/s10845-018-1411-1.

[42] M. C. Gombolay, R. A. Gutierrez, S. G. Clarke, G. F. Sturla and J. A. Shah, "Decision-making authority, team efficiency and human worker satisfaction in mixed human–robot teams," *Autonomous Robots*, vol. 39, no. 3, pp. 293–312, Oct. 2015. DOI: 10.1007/s10514-015-9457-9.

[43] K. Darvish, B. Bruno, E. Simetti, F. Mastrogiovanni and G. Casalino, "Interleaved Online Task Planning, Simulation, Task Allocation and Motion Control for Flexible Human-Robot Cooperation," in *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, ISSN: 1944-9437, Nanjing, China, Aug. 2018, pp. 58–65. DOI: 10.1109/ROMAN.2018.8525644.

[44] W. Y. Kwon and I. H. Suh, "A temporal Bayesian network with application to design of a proactive robotic assistant," in *2012 IEEE International Conference on Robotics and Automation*, ISSN: 1050-4729, Saint Paul, MN, USA, May 2012, pp. 3685–3690. DOI: 10.1109/ICRA.2012.6224673.

[45] R. Schulz, P. Kratzer and M. Toussaint, "Preferred Interaction Styles for Human-Robot Collaboration Vary Over Tasks With Different Action Types," *Frontiers in Neurorobotics*, vol. 12, Jul. 2018, Publisher: Frontiers. DOI: 10.3389/fnbot.2018.00036.

[46] R. Parasuraman and C. D. Wickens, "Humans: Still Vital After All These Years of Automation," *Human Factors*, vol. 50, no. 3, pp. 511–520, Jun. 2008, Publisher: SAGE Publications Inc. DOI: 10.1518/001872008X312198.

[47] P. Wintersberger, C. Schartmüller, S. Sadeghian, A.-K. Frison and A. Riener, "Evaluation of Imminent Take-Over Requests With Real Automation on a Test Track," *Human Factors*, vol. 65, no. 8, pp. 1776–1792, Dec. 2023, Publisher: SAGE Publications Inc. DOI: 10.1177/00187208211051435.

[48] C. Breazeal, "Social Interactions in HRI: The Robot View," *IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews)*, vol. 34, no. 2, pp. 181–186, May 2004. DOI: 10.1109/TSMCC.2004.826268.

[49] C. Breazeal and B. Scassellati, "How to build robots that make friends and influence people," in *Proceedings 1999 IEEE/RSJ International Conference on Intelligent Robots and Systems. Human and Environment Friendly Robots with High Intelligence and Emotional Quotients (Cat. No.99CH36289)*, vol. 2, Kyongju, South Korea: IEEE, 1999, pp. 858–863. DOI: 10.1109/IROS.1999.812787.

[50] L. Kunold and L. Onnasch, "A Framework to Study and Design Communication with Social Robots," *Robotics*, vol. 11, no. 6, p. 129, Nov. 2022. DOI: 10.3390/robotics11060129.

[51] P. Watzlawick, J. B. Bavelas, J. H. Beavin and D. D. Jackson, *Menschliche Kommunikation: Formen, Störungen, Paradoxien.* Huber, 1972, ISBN: 978-3-456-30389-5.

[52] J. Seibt, "Classifying Forms and Modes of Co-Working in the Ontology of Asymmetric Social Interactions (OASIS)," in *Envisioning Robots in Society – Power, Politics, and Public Space*, IOS Press, 2018, pp. 133–146. DOI: 10.3233/978-1-61499-931-7-133.

[53] E. Cha, A. D. Dragan and S. S. Srinivasa, "Perceived robot capability," in *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, Kobe, Japan, Aug. 2015, pp. 541–548. DOI: 10.1109/ROMAN.2015.7333656.

[54] E. Sandry, *Robots and Communication.* London: Palgrave Macmillan UK, 2015. DOI: 10.1057/9781137468376.

[55] P. Gustavsson, M. Holm, A. Syberfeldt and L. Wang, "Human-robot collaboration – towards new metrics for selection of communication technologies," *Procedia CIRP*, 51st CIRP Conference on Manufacturing Systems, vol. 72, pp. 123–128, Jan. 2018. DOI: 10.1016/j.procir.2018.03.156.

[56] C. E. Shannon, "A Mathematical Theory of Communication," *Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, Jul. 1948. DOI: 10.1002/j.1538-7305.1948.tb01338.x.

[57] H. D. Lasswell, "The Structure and Function of Communication in Society," in *The Communication of Ideas: a Series of Addresses*, L. Bryson, Ed., New York: Institute for Religious and Social Studies, 1948.

[58] D. K. Berlo, *The process of communication: an introduction to theory and practice*, eng. New York Chicago San Francisco Atlanta Dallas Montreal Toronto London Sydney: Holt, Rinehart and Winston, Inc, 1960, ISBN: 978-0-03-007490-5.

[59] R. E. Day, "The "conduit metaphor" and the nature and politics of information studies," *Journal of the American Society for Information Science*, vol. 51, no. 9, pp. 805–811, 2000. DOI: 10.1002/(SICI)1097-4571(2000)51:9<805::AID-ASI30>3.0.CO;2-C.

[60] D. Kincaid, "The convergence model of communication," *Honolulu : East-West Communication Institute*, vol. Communication Institute. Papers, Sep. 1979. [Online]. Available: http://hdl.handle.net/10125/48094 (visited on 14/04/2024).

[61] D. C. Barnlund, "A TRANSACTIONAL MODEL OF COMMUNICATION," in *Language Behavior*, J. Akin, A. Goldberg, G. Myers and J. Stewart, Eds., DE GRUYTER, Dec. 1970, pp. 43–61, ISBN: 978-90-279-1244-2. DOI: 10.1515/9783110878752.43.

[62] C. Lawson, R. Gill, A. Feekery, M. Witsel, M. Lewis and P. Cenere, *Communication Skills for Business Professionals*, en. Cambridge University Press, Jun. 2019, ISBN: 978-1-108-59441-7.

[63] P. Savage-Knepshield and J. Chen, Eds., *Advances in Human Factors in Robots and Unmanned Systems: Proceedings of the AHFE 2016 International Conference on Human Factors in Robots and Unmanned Systems, July 27-31, 2016, Walt Disney World®, Florida, USA* (Advances in Intelligent Systems and Computing), en. Springer International Publishing, 2017, vol. 499. DOI: 10.1007/978-3-319-41959-6.

[64] S. J. Lackey, D. J. Barber and S. G. Martinez, "Recommended Considerations for Human-Robot Interaction Communication Requirements," in *Human-Computer Interaction. Advanced Interaction Modalities and Techniques*, D. Hutchison, T. Kanade, J. Kittler *et al.*, Eds., vol. 8511, Springer International Publishing, 2014, pp. 663–674. DOI: 10.1007/978-3-319-07230-2_63.

[65] J. Banks and M. d. Graaf, "Toward an Agent-Agnostic Transmission Model: Synthesizing Anthropocentric and Technocentric Paradigms in Communication," *Human-Machine Communication*, vol. 1, no. 1, Feb. 2020. DOI: 10.30658/hmc.1.2.

[66] E. J. de Visser, M. M. M. Peeters, M. F. Jung *et al.*, "Towards a Theory of Longitudinal Trust Calibration in Human–Robot Teams," *International Journal of Social Robotics*, vol. 12, no. 2, pp. 459–478, May 2020. DOI: 10.1007/s12369-019-00596-x.

[67] B. Siciliano and O. Khatib, Eds., *Springer Handbook of Robotics*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008. DOI: 10.1007/978-3-540-30301-5.

[68] B. R. Burleson, "The Nature of Interpersonal Communication: A Message-Centered Approach," en, in *The Handbook of Communication Science*, 2455 Teller Road, Thousand Oaks California 91320 United States: SAGE Publications, Inc., 2010, pp. 145–164. DOI: 10.4135/9781412982818.n9.

[69] E. Sandry, "Re-evaluating the Form and Communication of Social Robots," *International Journal of Social Robotics*, vol. 7, no. 3, pp. 335–346, Jun. 2015. DOI: 10.1007/s12369-014-0278-3.

[70] W. Gerke, *Technische Assistenzsysteme: vom Industrieroboter zum Roboterassistenten*. Berlin, München, Boston: DE GRUYTER, Dec. 2014, ISBN: 978-3-11-034370-0. DOI: 10.1515/9783110343717.

[71] H. Su, W. Qi, J. Chen, C. Yang, J. Sandoval and M. A. Laribi, "Recent advancements in multimodal Human–Robot interaction," *Frontiers in Neurorobotics*, vol. 17, 2023. DOI: 10.3389/fnbot.2023.1084000.

[72] S. MacKenzie, *Human-computer Interaction*. Elsevier, 2013. DOI: 10.1016/C2012-0-02819-0. (visited on 25/03/2024).

[73] M. Schmauder and B. Spanner-Ulmer, *Ergonomie: Grundlagen zur Interaktion von Mensch, Technik und Organisation*, de, 2nd ed. München: Carl Hanser Verlag GmbH & Co. KG, Mar. 2022. DOI: 10.3139/9783446473584.

[74] K. Landau, "Lexikon Arbeitsgestaltung : Best practice im Arbeitsprozeß," ger, in *Lexikon Arbeitsgestaltung : best practice im Arbeitsprozeß*, 1. Aufl., Stuttgart [Stuttgart]: Gentner Ergonomia-Verl., 2007, ISBN: 3-87247-655-6.

[75] A. Carrera-Rivera, W. Ochoa, F. Larrinaga and G. Lasa, "How-to conduct a systematic literature review: A quick guide for computer science research," *MethodsX*, vol. 9, p. 101 895, Nov. 2022. DOI: 10.1016/j.mex.2022.101895.

[76] N. Biggs, E. K. Lloyd and R. J. Wilson, *Graph Theory, 1736-1936*, en. Clarendon Press, 1986, ISBN: 978-0-19-853916-2.

[77] K. Darvish, F. Wanderlingh, B. Bruno, E. Simetti, F. Mastrogiovanni and G. Casalino, "Flexible human–robot cooperation models for assisted shop-floor tasks," *Mechatronics*, vol. 51, pp. 97–114, May 2018. DOI: 10.1016/j.mechatronics.2018.03.006.

[78] H. Karami, A. Carfì and F. Mastrogiovanni, "Branched AND/OR Graphs: Toward Flexible and Adaptable Human-Robot Collaboration," in *2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)*, ISSN: 1944-9437, Vancouver, BC, Canada, Aug. 2021, pp. 527–533. DOI: 10.1109/RO-MAN50785.2021.9515512.

[79] R. A. Knepper, D. Ahuja, G. Lalonde and D. Rus, "Distributed Assembly with AND/OR Graphs," en, *In Workshop on AI Robotics at the Int. Conf. on Intelligent Robots and Systems (IROS).*, 2014. [Online]. Available: https://people.csail.mit.edu/gdk/iros-airob14/papers/andor.pdf (visited on 17/05/2024).

[80] L. Johannsmeier and S. Haddadin, "A Hierarchical Human-Robot Interaction-Planning Framework for Task Allocation in Collaborative Industrial Assembly Processes," *IEEE Robotics and Automation Letters*, vol. 2, no. 1, pp. 41–48, Jan. 2017. DOI: 10.1109/LRA.2016.2535907.

[81] P. E. Hart, N. J. Nilsson and B. Raphael, "A Formal Basis for the Heuristic Determination of Minimum Cost Paths," *IEEE Transactions on Systems Science and Cybernetics*, vol. 4, no. 2, pp. 100–107, Jul. 1968, Conference Name: IEEE Transactions on Systems Science and Cybernetics. DOI: 10.1109/TSSC.1968.300136.

[82] A. Martelli and U. Montanari, "Additive AND/OR graphs," in *Proceedings of the 3rd international joint conference on Artificial intelligence*, ser. IJCAI'73, San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., Aug. 1973, pp. 1–11.

[83] K. Darvish, E. Simetti, F. Mastrogiovanni and G. Casalino, "A Hierarchical Architecture for Human–Robot Cooperation Processes," *IEEE Transactions on Robotics*, vol. 37, no. 2, pp. 567–586, Nov. 2020, Conference Name: IEEE Transactions on Robotics. DOI: 10.1109/TRO.2020.3033715.

[84] F. Fusaro, E. Lamon, E. D. Momi and A. Ajoudani, "A Human-Aware Method to Plan Complex Cooperative and Autonomous Tasks using Behavior Trees," in *2020 IEEE-RAS 20th International Conference on Humanoid Robots (Humanoids)*, ISSN: 2164-0580, Munich, Germany, Jul. 2021, pp. 522–529. DOI: 10.1109/HUMANOIDS47582.2021.9555683.

[85] M. Colledanchise and P. Ögren, *Behavior Trees in Robotics and AI*, 0th ed. CRC Press, Jul. 2018. DOI: 10.1201/9780429489105.

[86] B. Hayes and B. Scassellati, "Autonomously constructing hierarchical task networks for planning and human-robot collaboration," en, in *2016 IEEE International Conference on Robotics and Automation (ICRA)*, Stockholm, Sweden: IEEE, May 2016, pp. 5469–5476, ISBN: 978-1-4673-8026-3. DOI: 10.1109/ICRA.2016.7487760. [Online]. Available: http://ieeexplore.ieee.org/document/7487760/ (visited on 17/05/2024).

[87] A. Roncone, O. Mangin and B. Scassellati, "Transparent role assignment and task allocation in human robot collaboration," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, Singapore: IEEE, May 2017, pp. 1014–1021. DOI: 10.1109/ICRA.2017.7989122.

[88] G. Buisan, A. Favier, A. Mayima and R. Alami, "HATP/EHDA: A Robot Task Planner Anticipating and Eliciting Human Decisions and Actions," in *2022 International Conference on Robotics and Automation (ICRA)*, Philadelphia, PA, USA, May 2022, pp. 2818–2824. DOI: 10.1109/ICRA46639.2022.9812227.

[89] B. Hayes, "First Links in the Markov Chain," *American Scientist*, vol. 101, no. 2, p. 92, 2013. DOI: 10.1511/2013.101.92.

[90] R. Bellman, "A Markovian Decision Process," *Indiana University Mathematics Journal*, vol. 6, no. 4, pp. 679–684, 1957, ISSN: 0022-2518.

[91] R. A. Howard, *Dynamic programming and Markov processes*. Technology Press of Massachusetts Institute of Technology, 1960, ISBN: 978-0-262-08009-5.

[92] A. Favier, S. Shekhar and R. Alami, "Models and Algorithms for Human-Aware Task Planning with Integrated Theory of Mind," in *2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, Place: , Republic of tex.eventtitle: 2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), Busan, Korea: IEEE, Aug. 2023, pp. 1279–1286. DOI: 10.1109/RO-MAN57019.2023.10309437.

[93] P. Poupart, "Partially Observable Markov Decision Processes," in *Encyclopedia of Machine Learning*, C. Sammut and G. I. Webb, Eds., Boston, MA: Springer US, 2010, pp. 754–760. DOI: 10.1007/978-0-387-30164-8_629.

[94] V. V. Unhelkar, S. Li and J. A. Shah, "Decision-Making for Bidirectional Communication in Sequential Human-Robot Collaborative Tasks," in *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '20, New York, NY, USA: Association for Computing Machinery, 2020, pp. 329–341. DOI: 10.1145/3319502.3374779.

[95] K. Mahadevan, M. Sousa, A. Tang and T. Grossman, ""Grip-that-there": An Investigation of Explicit and Implicit Task Allocation Techniques for Human-Robot Collaboration," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, ser. CHI '21, New York, NY, USA: Association for Computing Machinery, 2021, pp. 1–14. DOI: 10.1145/3411764.3445355.

[96] B. W. Weaver and P. R. DeLucia, "A Systematic Review and Meta-Analysis of Takeover Performance During Conditionally Automated Driving," en, *Human Factors*, vol. 64, no. 7, pp. 1227–1260, Nov. 2022, Publisher: SAGE Publications Inc. DOI: 10.1177/0018720820976476.

[97] J. Arents, V. Abolins, J. Judvaitis, O. Vismanis, A. Oraby and K. Ozols, "Human–Robot Collaboration Trends and Safety Aspects: A Systematic Review," *Journal of Sensor and Actuator Networks*, vol. 10, no. 3, p. 48, Sep. 2021. DOI: 10.3390/jsan10030048.

[98] D. Yu and L. Deng, *Automatic Speech Recognition: A Deep Learning Approach* (Signals and Communication Technology). London: Springer London, 2015. DOI: 10.1007/978-1-4471-5779-3.

[99] S. Alharbi, M. Alrazgan, A. Alrashed *et al.*, "Automatic Speech Recognition: Systematic Literature Review," en, *IEEE Access*, vol. 9, pp. 131 858–131 876, 2021. DOI: 10.1109/ACCESS.2021.3112535.

[100] A. Angleraud, Q. Houbre, M. Netzev and R. Pieters, "Cognitive Semantics For Dynamic Planning In Human-Robot Teams," in *2019 IEEE 15th International Conference on Automation Science and Engineering (CASE)*, Vancouver, BC, Canada, Aug. 2019, pp. 942–947. DOI: 10.1109/COASE.2019.8842842.

[101] A. Gottardi, M. Terreran, C. Frommel *et al.*, "Dynamic Human-Aware Task Planner for Human-Robot Collaboration in Industrial Scenario," in *2023 European Conference on Mobile Robots (ECMR)*, ISSN: 2767-8733, Coimbra, Portugal, Sep. 2023, pp. 1–8. DOI: 10.1109/ECMR59166.2023.10256268.

[102] H. Liu and L. Wang, "Gesture recognition for human-robot collaboration: A review," *International Journal of Industrial Ergonomics*, vol. 68, pp. 355–367, Nov. 2018. DOI: 10.1016/j.ergon.2017.02.004.

[103] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei and Y. Sheikh, *OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields*, May 2019. DOI: 10.48550/arXiv.1812.08008.

[104] A. Hameed, A. Ordys, J. Możaryn and A. Sibilska-Mroziewicz, "Control System Design and Methods for Collaborative Robots: Review," *Applied Sciences*, vol. 13, no. 1, p. 675, Jan. 2023. DOI: 10.3390/app13010675.

[105] U. E. Ogenyi, J. Liu, C. Yang, Z. Ju and H. Liu, "Physical Human–Robot Collaboration: Robotic Systems, Learning Methods, Collaborative Strategies, Sensors, and Actuators," *IEEE Transactions on Cybernetics*, vol. 51, no. 4, pp. 1888–1901, Apr. 2021. DOI: 10.1109/TCYB.2019.2947532.

[106] C. Yang, H. Wu, Z. Li, W. He, N. Wang and C.-Y. Su, "Mind Control of a Robotic Arm With Visual Fusion Technology," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 9, pp. 3822–3830, Sep. 2018. DOI: 10.1109/TII.2017.2785415.

[107] T. B. Tuli and M. Manns, “Explainable human activity recognition based on probabilistic spatial partitions for symbiotic workplaces,” *International Journal of Computer Integrated Manufacturing*, vol. 36, no. 12, pp. 1783–1800, Dec. 2023. DOI: 10.1080/0951192X.2023.2177742. (visited on 10/05/2024).

[108] Y. Cheng and M. Tomizuka, “Long-Term Trajectory Prediction of the Human Hand and Duration Estimation of the Human Action,” *IEEE Robotics and Automation Letters*, vol. 7, no. 1, pp. 247–254, Jan. 2022. DOI: 10.1109/LRA.2021.3124524.

[109] M.-L. Lee, W. Liu, S. Behdad, X. Liang and M. Zheng, “Robot-Assisted Disassembly Sequence Planning With Real-Time Human Motion Prediction,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 53, no. 1, pp. 438–450, Jan. 2023. DOI: 10.1109/TSMC.2022.3185889.

[110] H. R. Lee, E. Cheon, C. Lim and K. Fischer, “Configuring Humans: What Roles Humans Play in HRI Research,” in *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, Sapporo, Japan, Mar. 2022, pp. 478–492. DOI: 10.1109/HRI53351.2022.9889496.

[111] M. Cramer, K. Kellens and E. Demeester, “Probabilistic Decision Model for Adaptive Task Planning in Human-Robot Collaborative Assembly Based on Designer and Operator Intents,” *IEEE Robotics and Automation Letters*, vol. 6, no. 4, pp. 7325–7332, Oct. 2021. DOI: 10.1109/LRA.2021.3095513.

[112] G. Hoffman and X. Zhao, “A Primer for Conducting Experiments in Human–Robot Interaction,” en, *ACM Transactions on Human-Robot Interaction*, vol. 10, no. 1, pp. 1–31, Mar. 2021, ISSN: 2573-9522. DOI: 10.1145/3412374. [Online]. Available: https://dl.acm.org/doi/10.1145/3412374 (visited on 09/09/2024).

[113] G. Hoffman, “Evaluating Fluency in Human–Robot Collaboration,” *IEEE Transactions on Human-Machine Systems*, vol. 49, no. 3, pp. 209–218, Jun. 2019. DOI: 10.1109/THMS.2019.2904558.

[114] M. Quigley, B. Gerkey, K. Conley *et al.*, “ROS: An open-source Robot Operating System,” en, in *Proc. of the IEEE Intl. Conf. on Robotics and Automation (ICRA) Workshop on Open Source Robotics*, Kobe, Japan, May 2009.

[115] D. T. Coleman, I. A. Sucan, S. Chitta and N. Correll, “Reducing the Barrier to Entry of Complex Robotic Software: A MoveIt! Case Study,” *Journal of Software Engineeering in Robotics*, 2014. DOI: 10.6092/JOSER_2014_05_01_P3.

[116] J. W. Mauchly, "Significance Test for Sphericity of a Normal n-Variate Distribution," en, *The Annals of Mathematical Statistics*, vol. 11, no. 2, pp. 204–209, Jun. 1940, ISSN: 0003-4851. DOI: 10.1214/aoms/1177731915.

[117] S. W. Greenhouse and S. Geisser, "On Methods in the Analysis of Profile Data," en, *Psychometrika*, vol. 24, no. 2, pp. 95–112, Jun. 1959, ISSN: 0033-3123, 1860-0980. DOI: 10.1007/BF02289823.