Lehrstuhl für Bildverarbeitung und Mustererkennung
Fakultät für Informatik
Technische Universität München

Legibility of Robot Behavior

Investigating Legibility of Robot Navigation in Human-Robot Path Crossing Scenarios

Christina Lichtenthäler

Vollständiger Abdruck der von der Fakultät für Informatik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzende: Univ.-Prof. Gudrun J. Klinker, Ph.D.

Prüfer der Dissertation:

1. Univ.-Prof. Dr. Daniel Cremers
2. Prof. Dr. Alexandra Kirsch
3. Dr.-Ing. Marc A. Hanheide, Reader University of Lincoln, UK

Die Dissertation wurde am 17.06.2014 bei der Technischen Universität München eingereicht und durch die Fakultät für Informatik am 09.11.2014 angenommen.

Acknowledgements

I would like to thank my supervisor, Professor Alexandra Kirsch, for her support over the last years. I also want to thank my colleague and fellow in misery Tamara Lorenz for her help, advise, support, for every encouraging discussion during my PhD. I would like to thank, in particular, Annika Peters and Sascha Griffith for our fruitful cooperation, a very nice time in Bielefeld, and for tips and hints to survive a PhD. A very special thanks is due to three persons of the Physioduo team, Martina Hangen, Thomas Ehlert, and Claudia Wagner for keeping my body and mind healthy, for welcome distraction and for kilograms of chocolate.

Last but not least I will say "Thank you" to those people who supported me along the way. First of all, there is my husband Jürgen, who helped me through all the mental breakdowns and tried to keep me sane, which I have to admit hasn't been easy – sometimes. And at the end I would like to thank my beloved sons for loving me despite the fact that I was ever reading papers when we were at the open-air swimming pool.

Contents

List of Figures

List of Tables

Abstract

A key requirement for a smooth and safe human-robot collaboration is for the robot to make its intentions and goals clear to its human collaborator. Such behavior is referred to as legible robot behavior. The thesis at hand investigates legibility of robot navigation behavior in human-robot path crossing situations. We conducted five different experiments in order to explore how legible state-of-the-art navigation algorithms are perceived and how the robot should behave in a path crossing situation to be as legible as possible. Furthermore, we investigated how the spatial relationship influences human expectations in a path crossing scenario and which other important Human-Robot Interaction properties like perceived safety, comfort, and likability are influenced by the legibility of robot navigation behavior.

Zusammenfassung

Eine wesentliche Voraussetzung für eine reibungslose und sichere Mensch-Roboter-Kooperation ist, dass der Roboter seine Absichten und Ziele dem interagierenden Menschen erkennbar macht. Ein solches Verhalten bezeichnen wir als lesbar. Die vorliegende Arbeit untersucht die Lesbarkeit von Roboter Bewegungen in Situationen, in denen der Roboter den Weg des Menschen kreuzt. Wir führten fünf verschiedene Experimente durch, um zu ermitteln, wie lesbar das Roboterverhalten bei der Verwendung von state-of-the-art Navigationsalgorithmen ist und wie ein Roboter sich in einer Kreuzungs-Situation verhalten sollte, um seine Ziele und Intentionen so deutlich wie möglich zu zeigen. Darüber hinaus untersuchten wir, wie die räumliche Beziehung von Mensch und Roboter zueinander, die Erwartungen bezüglich des Roboterverhaltens beeinflusst. Wir analysierten ebenfalls inwieweit andere wichtige Mensch-Roboter Interaktionseigenschaften, wie zum Beispiel Sicherheitsempfinden, Komfort, und Sympathie von der Lesbarkeit der Roboternavigation beeinflusst werden.

1. Introduction

Nothing in life has to be feared, it only has to be understood.

Marie Curie

1.1. Motivation

Beside the main message of the quote ahead to foster scientific research, we can also derive that understanding is a crucial factor in scientific research and everyday life. New technologies will only be accepted and used when its functionality, its intentions, its behavior is understood. For example, when looking around in the subway then you see a lot of old people, young people, normal people, playing around with their IPad, Smartphone, Kindle, etc. They all using the new technology of taking a small computer wherever they go. Would these people use a 80's computer with a text-based command line system? No, no-one would *understand* such a computer, but they all understand how to use an IPad. Furthermore, one of the design rules for future things Don Norman published in his book [87] is that a "smart" machine has to be predictable and the output has to be understandable. For smooth and comfortable interaction we have to make future technology understandable. For example, we use the flashing indicator to show other drivers that we are planning to turn right or left in order to enable other drivers to react according to our intention to turn, which makes the interaction smooth, comfortable and also safe. We can conclude that understanding and predictability are important key factors for new technologies.

The technology the thesis at hand discusses is robotics and in particular the research field of Human-Robot Interaction (HRI). According to Dautenhahn [25] HRI is a challenging research field at the intersection of psychology, cognitive science, social science, artificial intelligence, computer science, robotics, engineering, and human-computer interaction. HRI investigates the interaction and communication of humans with robots. The outcome of scientific research, industrial or sociopolitical studies [10, 47, 56] shows us that the use of robots in our daily life is more than just science fiction and that people would like to have a robot assistant [24, 27]. Robots are already there, cleaning our houses[1] or carrying laboratory samples and medications through hospitals[2]. Robots will definitely enter our daily life and the aim of HRI research is to understand, design, and evaluate robotic systems to use with humans [43].

Going back to the previous conclusion that understanding and predictability are important factors for future technology we now consider understanding and predictability in the field of Human-Robot Interaction (HRI). When interacting with an embodied agent, like a robot, it becomes crucial to understand and predict the robot's behavior for a smooth, comfortable, and safe interaction. Imagine a robot that assists you with the household chore, such as preparing a meal. The robot fulfills its duty,

[1] `http://www.irobot.com/`

[2] `http://www.swisslog.com/de/Products/HCS/Automated-Material-Transport/RoboCourier-Autonomous-Mobile-Robot`

but manipulates objects with sudden, unpredictable movements. It rushes through the kitchen with rapid changes of direction or ignores obvious errors like a pot not being placed properly on the stove. It opens the tap before getting the pot to fill the water in. One would be irritated by the unpredictable and not understandable robot behavior. It is obvious that a smooth, comfortable, and in particular safe interaction is hardly possible. In the literature, we have found several publications regarding HRI and predictability confirming our assumption that predictability is an important HRI property. An example is the result of a study Dautenhahn et al. [27] conducted in 2005. They concluded that a robot has to be predictable from their results of a questionnaire regarding the attitudes towards the idea of a future robot companion for the home. Furthermore, Koay et al. [62] found that predictability is one main concern for their participants when the robot is approaching them for interaction. Bortot et al. [15] also found that predictability of robot behavior is one key aspect in industrial co-worker scenarios. We can ,therefore, conclude, that predictability is an important property for human-robot interaction scenarios that have to be further investigated.

Dautenhahn et al. [23] and Alami et al. [1,2,86] coined the term "legibility" in 2005/2006. According to them, legibility means more than only predictability, it means to make the robot‘s actions and behavior *understandable* and *predictable* to a human. Since then, much research has been carried out in order to examine legibility of robot behavior [31, 62, 106] as well as to develop methods to make robot behavior more legible [32, 64, 67, 100]. The thesis at hand is inspired and based on the research conducted so far regarding legibility of robot behavior and aims to extend previous findings and methods.

In particular, the robot behavior we examine in detail with the thesis at hand is robot navigation. As mentioned before, robots will enter our daily life and wherever they will be used, in factories as a co-worker, in nursing homes or hospitals as a care assistant, as a guide in supermarkets, or as a household-robot one crucial behavior, which they have all in general, is navigation. The ability to move autonomously differentiates robots from other new smart technologies and is ,therefore, one behavior we have specifically to focus on. Furthermore, according to Syrdal et al. [105] a moving robot is perceived as more likable and the spatial relationship between human and robot plays an important role concerning the relationship between them. In particular, regarding legibility, navigation is very interesting to investigate. Remember our example of the path crossing cars and transfer it to Human-Robot Interaction. Without knowing the direction and intentions of the approaching robot, a smooth interaction is hardly possible. We finally conclude that navigation is a key behavior for robots and legibility is a key property, which has to be taken into account seriously.

To this end, the following thesis investigates legibility of robot navigation behavior. We will investigate how a robot should move in the presence of humans to be as legible as possible.

1.2. The Scenario - Human-Robot Path Crossings

The HRI aspect we want to investigate further with the thesis at hand is legibility of robot navigation behavior and in particular of robot navigation in path crossing scenarios. We confine ourselves to the particular scenario of human-robot path crossings because it is one of the most crucial interactions for a robot with a human while navigating through its workspace. Remember our car example, when the robot crosses our path it is essential to know about its intentions to perform this kind of interaction smooth, comfortable, and especially safe. The behavior where a robot approaches a sitting or standing human is also very interesting for robot navigation and already investigated so far [4, 23, 62, 94]. Regarding human-robot path crossing scenarios much research deals with head-on encounters, which are often occurring in hallway settings. [58, 78, 115]. They all commonly agree in their navigation approaches. The robot moves to one side in order to avoid the human. Side encounters are studied, for example, in a human-human experiment by Basili et al. [9] and their results are implemented in a navigation method and evaluated regarding comfort and confusion by Kruse et al. [64, 65]. Their approach is to go straight ahead and decrease the velocity to allow the human to pass. However, their investigations only consider $90°$ crossings. Another navigation approach, which aims to be legible, is proposed by Sisbot et al. [97], but they did not consider human-robot path crossing situations. Moreover, the property legibility has currently not been investigated further in human-robot path crossings so far. Therefore, the aim of the thesis at hand is to investigate the scenario of human-robot path crossings regarding legibility.

In the following, we want to formally define our scenario of investigation:

Definition 1. *A human-robot path crossing situation is defined as a situation where 1) the robot crosses the human's path and 2) both are located before the crossing point.*
(Example depicted in Figure Fig. 1)

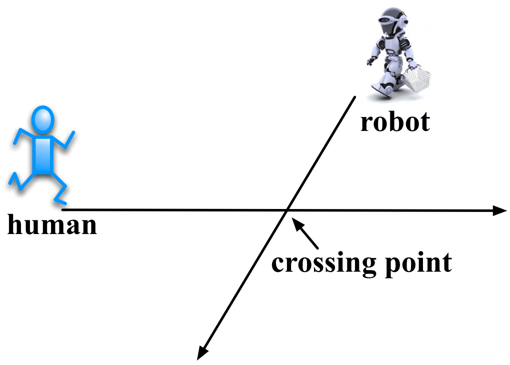


Fig. 1: The diagram depicts a crossing situation. The robot crosses the human's path and both are located before the crossing point.

Due to our definition (Def. 1) a path crossing situation extends over a period where human and robot are approaching each other. In order to reduce the time interval we start to consider a path crossing situation not until the crossing point is less than 3.6m away. We choose the distance of 3.6m due to the proxemics theory [48], which is commonly used and investigated in the HRI community [84, 107, 112]. Proxemics is the study of how humans use and manipulate distances between each other with regard to social behavior [112]. Edward T. Hall coined the term proxemic in 1963 and emphasized the impact of proxemic behavior (the use of space) on interpersonal communication. Hall identified four proxemic zones for human-human social interaction (see Figure Fig. 2 on the following page), an intimate space (0.15m to 0.46m), a personal space (0.46m to 1.2m), a social space (1.2m to 3.6m), and a public space (3.6m to 7.6m). Most interactions, like talking to each other, take place within the personal space. The same applies for human-robot interactions. Hüttenrauch et al. found

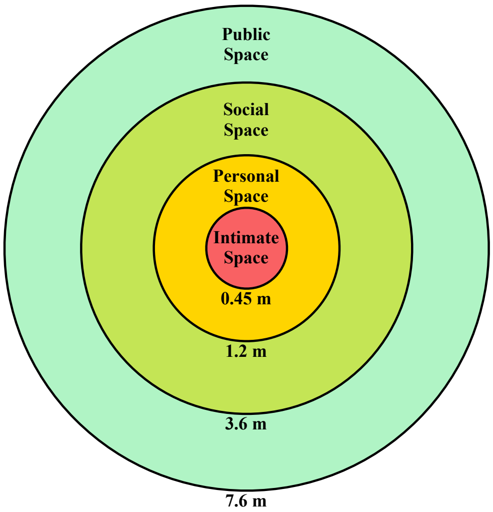


Fig. 2: The diagram depicts the proxemic zones around a human according to Hall's proxemics theory [48].

that most participants preferred interacting with the robot at a distance of 0.46m to 1.22m, which is within the personal space. With regard to our path crossing scenario Yoda et al. [115] found in a human-human frontal approaching experiment that it is preferred to avoid each other at a distance of approximately 2.1m, which is within the social space. [89]

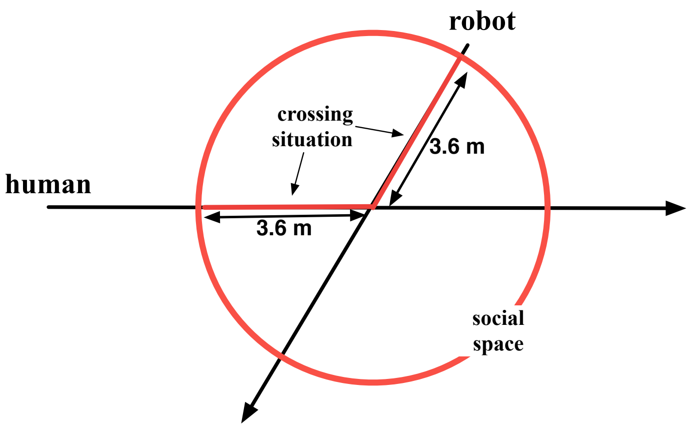


Fig. 3: The diagram depicts which part of a crossing situation we investigate. The robot crosses the humans path and both are located at a maximum distance of 3.6m before the crossing point.

Based on these findings we can assume that most of the interaction takes place within the personal and social space, meaning that the distance between human and robot is less than 3.6m. Therefore, we confined the space of investigation based on the social space distance of 3.6 m. Different from the classical proxemics theory we defined the crossing point as the center of the social space circle (see Figure Fig. 3). To this end, we investigate all situations where both robot and human are within the social space around the crossing point. By choosing the crossing point as the center of our space of investigation, we consider not only the distance d between human and robot but also the two distances d_h human to crossing point and d_r robot to the crossing point (see Figure Fig. 4 on the facing page). Therefore, we investigate not only the distance of the human to the robot as in common proxemics, but also the spatial relationship of human and robot regarding the crossing point, which provides us way more information about the crossing situation. Note, that in a crossing situation for one human-robot distance d exists an infinite number of combinations of robot to crossing point distance d_r and human to crossing point distance d_h. One objective of the work at hand is to investigate the proxemics of human-robot path crossings with regard to the crossing point, which is a very recent approach in robot navigation.

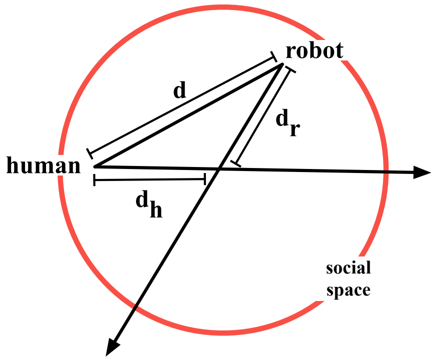


Fig. 4: The diagram depicts the distances we investigate in a crossing situation regarding the crossing point. With d we denote the distance human to robot, with d_h we denote the distance human to crossing point, and with d_r we denote the distance of the robot to the crossing point.

1.3. Objectives

The objective of the thesis at hand is to investigate legibility of robot navigation behavior in human-robot path crossing situations. To this end, the first goal is to define and formalize the term legibility and to develop methods to measure legibility in human-robot path crossing scenarios. Furthermore, we investigate which other HRI properties are correlated with legibility in order to find how legibility fits in the HRI context. For example, one question we want to answer is if safety and comfort are associated with legibility. Our main objective is to evaluate legibility of robot navigation behavior as well as to investigate which kind of robot navigation behavior is perceived as highly legible. Finally, we aim to conclude all our findings to propose how we can realize a legible robot navigation for path crossing scenarios.

1.4. Methods

In order to accomplish our objectives we performed literature research to get an overview of current research regarding legibility and to derive a general definition for legibility. Based on theoretical foundations and previous research we elaborated experimental methods to evaluate robot behavior regarding legibility. After that, we implemented the experimental methods to evaluate legibility of robot navigation behavior in path crossing situations. Furthermore, we used the evaluation methods to figure out which kind of behavior is perceived as legible. We analyzed the gathered data with standardized statistical methods from experimental psychology and drew conclusions.

1.5. Contributions

The thesis at hand contributes several results to field of HRI research regarding our investigations of legibility of robot behavior.

- a common definition of legibility
- suggestions of experimental methods to evaluate legibility of arbitrary robot behavior using self-assessment, interview , task performance metrics, psychophysiological measurements, and behavioral measures
- experimental methods to evaluate legibility in human-robot path crossing situations
- experimental method to capture human expectations of robot navigation behavior
- experimental results showing which kind of behavior is highly legible

1.6. Thesis Outline

In the following we provide a short outline of further chapters of this thesis.

Chapter 2 (Legibility of Robot Behavior) presents the results of a literature research regarding investigations, definitions, evaluation methods, implementations, and correlated HRI properties of legibility.

Chapter 3 (Legibility Metrics) suggests several methodologies and measures to evaluate the legibility of robot behavior.

Chapter 4 (Measuring the Legibility of Robot Navigatio Behavior) presents the experimental setups and results of the experiments we conducted to investigate the legibility of robot navigation behavior. We measured legibility of different navigation methods and investigated which kind of navigation behavior is most legible in a human-robot path crossing scenario.

Chapter 5 (Determine Human Expectations of Robot Navigation Behavior) describes our method to capture human expectations and presents the results of the conducted study to capture human expectations in human-robot path crossing scenarios. Furthermore, we show how to use the spatial relationship as predictor for legible navigation behavior.

Chapter 6 (Discussion) discusses the results of the thesis and draws final conclusions regarding a legible robot navigation. We also point out open questions and considerable factors for legible robot navigation.

2. Legibility of Robot Behavior - A Literature Review

The following chapter presents a literature research regarding legible robot behavior. The purpose of the review is to answer the questions: How do other researchers define legibility? How can legibility be measured? How can one implement legible robot behavior? Which factors are correlated with legibility? We searched all relevant publication databases and systematically reviewed all publications addressing legibility.

2.1. Introduction

In the chapter ahead, we introduced the term legibility and motivated that legibility is an important property for smooth, comfortable and safe human-robot interaction. In the chapter at hand we want to see the big picture of research regarding legible robot behavior. How is legibility seen in the HRI community, what are the conclusions regarding legibility so far, or which methods are used to evaluate legibility. These questions we want further to investigate by conducting a literature research. Furthermore, due to the novelty of the topic of legible robot behavior and also the novelty of the emerging field of Human-Robot Interaction (HRI) no generally accepted definition or guidelines for legibility have been proposed so far. Therefore, one objective is to find out what other researchers in the HRI community have published regarding legible robot behavior in order to extract a general definition as well as further interesting facts about legibility as methods to measure legibility, or how can we generate legible behavior and what needs to be considered. Another interesting aspect is to examine which other acceptance measures [51] or HRI concepts [7] are correlated with legibility. To this end, we conducted a literature review where we systematically investigated articles regarding the following research questions:

- How is legibility defined? Which terms are used synonymously?
- How is legibility measured? What research methods have been used to investigate legibility?
- How can legible robot behavior be realized?
- Which HRI properties[1] are influenced by legibility?

Based on the review at hand we want to conclude our findings to a common definition of legible robot behavior, which is in line with the reviewed articles.

[1] As HRI property we determine all variables indicating (1) how a human-robot interaction was perceived by the human interactor or (2) how we can assess an interaction.

2.2. Review Protocol

In order to find all publications related to what we define as "legibility of robot behavior" we searched the ACM Digital Library[2] and IEEExplore Digital Library[3]. Additionally, we used the literature search engine "Web of Knowledge" [4] to take into account a wide range of journals. From our previous work and former literature research we know that other authors describe, what we call legibility, with terms like readability [106], anticipation [40], or simply predictability [15] of robot behavior and motion. Based on this knowledge we determined the following search terms to find publications concerning legible robot behavior:

- legibility/legible AND motion/behavior AND robot
- readability/readable AND motion/behavior AND robot
- anticipatory/anticipate AND motion/behavior AND robot
- predictability/predictable AND motion/behavior AND robot

After a first selection process where we removed all non related publications[5] we found 32 publications dealing with legibility of robot behavior [2, 8, 11, 15, 23, 27, 28, 31, 32, 34, 40, 45, 60, 61, 64, 67, 73, 73, 74, 74, 75, 75, 79, 80, 86, 96–98, 100, 101, 106]. As mentioned before, some authors use other notations for what we denote as legibility. Most of the authors (21) are using the term legibility [1, 2, 11, 20, 21, 28, 31, 32, 45, 60, 61, 64, 67, 73–75, 79, 80, 96–98, 100, 101] a few (5) are using predictability [8, 15, 23, 27, 34], one is using the term anticipatory [40] and one is using the term readability [106].

As shown in Figure Fig. 5 on the next page a high ratio of the publications were published in the Proceedings of the Human-Robot Interaction Conference (HRI) and the International Symposium on Robot and Human Interactive Communication (RO-MAN). The first articles dealing with legibility were published in 2005. The histogram in Figure Fig. 6 on the facing page shows the distribution of publications over the time.

2.3. Results of the Literature Research

In the following, we first give a brief overview of the reviewed articles by summarizing the content of each publication. Afterwards, we report our findings regarding the aforementioned questions.

[2]ACM: `http://dl.acm.org/`

[3]IEEE: `http://ieeexplore.ieee.org`

[4]Web of Knowlege: `http://apps.webofknowledge.com`

[5]e.g. publications where the search terms are not related to each, like motion does not refer to robot motion and predictable refers to a statistical test. Sometimes the term legibility refers to a figure, or the topic was human intention recognition

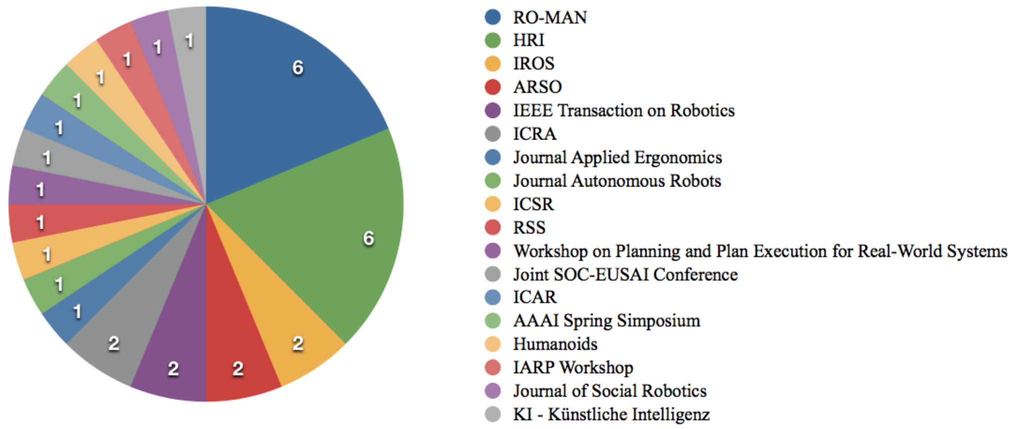


Fig. 5: Pie Chart showing the distribution of legibility articles in the HRI publication venues. A high ratio (12 of 32) appeared in proceedings of the ACM International Conference on Human-Robot Interaction (HRI) and IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN).

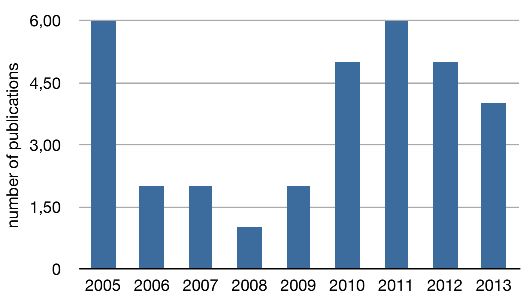


Fig. 6: Histogram showing the distribution of legibility articles over the last eight years (2005-2013).

2.3.1. Summary

We start our summary with a list of the articles where legibility is only briefly mentioned as an important factor or new challenge for HRI [15,20,21,23,27,86]. For example, Colic et al. [21] wrote: *"One major key point is that the robot must act in a way judged as legible and acceptable by humans."* Furthermore, one finding of a study Dautenhahn et al. [27] conducted to explore people's perception towards the future use of robot companions was that the behavior of a robot has to be predictable (= legible). As a main result, they state, *"a future robot companion would need to be predictable, controllable, considerable and polite"* [27].

A vast amount of work was published by researchers from the Robotics and Artificial Intelligence Group at LAAS/CNRS in Toulouse, France, headed by Rachid Alami [1, 2, 20, 21, 28, 79, 80, 96–98, 100, 101]. One goal of the group is the development of a human aware motion and manipulation framework, which synthesizes safe, comfortable, socially acceptable, and legible motions. Therefore, the main topic of the publications from this group is the development, implementation and evaluation of such a system, which consists of a Human-Aware Navigation Planner (HANP) [97, 100], generating legible paths, and a Human-Aware Manipulation Planner (HAMP) [79,80,96,98,99,101], performing legible hand-over tasks. They have integrated cost functions modeling human comfort, safety, and visibility in their navigation and motion planning system in order to generate comfortable, safe and visible motions. The cost functions are partly based on the findings of the human-robot experiments conducted by Dautenhahn et al. [23, 27]. An evaluation of HAMP using psychophysiological measures was conducted by Dehais et al. [28]. Furthermore, some work was done towards a Human-Aware Task Planner (HATP) [1, 2] which should produce legible high level robot task plans.

Kruse et al. [67] extended the Human Aware Navigation Planner [100] by relaxing some constraints and changing cost functions in order to exploit the fact that humans might act cooperatively and make way for the robot, which should increase the legibility of HANP. Another approach to increase the legibility of HANP was also proposed by Kruse et al. [64]. They assume that human-like behavior leads to a higher legibility. Therefore, they added an additional cost function to HANP in order to produce navigation behavior similar to the behavior humans show in a human-human path crossing experiment [9]. Also Guzzi et al. [45] followed the "human-likeness increases legibility" assumption and introduced a navigation algorithm based on simple obstacle avoidance heuristics modeling pedestrian behavior. Beetz et al. [11] published another implementation of this "human-likeness increases legibility" assumption. They use a dynamic movement primitive approach to mimic human arm trajectories. Furthermore, their proposed method for mobile manipulation should generate more stereotypical motions and behavior, which, according to their assumption, leads to a more legible behavior.

A high level joint-task planning framework was proposed by Kirsch et al. [60]. Their assumption is that robot plans must take into account human preferences and abilities in order to generate legible plans. Therefore, they present an integrated planning and learning framework. In [61] Kirsch et al. described several planning and action-related HRI problems like collaborative planning, navigation and joint manipulation and pointed out the challenges. One challenge they point out is to achieve legible robot behavior. Furthermore, they claim that *"legibility is a prerequisite to establish human* *comfort"* and that *"Legibility is connected to the perceived safety and comfort"*.

Despite the promising results of the aforementioned methods and approaches legibility has not been objectively evaluated. Only the Human-Aware Manipulation Planner [100] HAMP was tested regarding legibility in an experiment conducted by Dehais et al. [28]. The aforementioned work is not only focussing on legibility, other factors like comfort, social acceptability, and safety are also taken into account (e.g., [45, 67, 96, 100]). To conclude, in the work presented so far the authors either model social constraints, human preferences and abilities and/or mimic human motions in order to generate legible as well as comfortable, safe and socially acceptable robot behavior.

The work from Dragan et al. and Gielniak et al. [31, 32, 40] is focusing explicitly on generating legible motions. Both present a method to generate legible motions as well as an evaluation method to test the proposed method regarding legibility. Gielniak et al. [40] presents an algorithm to increase the legibility of a given gesture. In their work, a legible motion is determined as an anticipatory motion. Dragan et al. [31, 32] present a mathematical definition and model for goal directed legible arm motions. In their work the authors distinguish between legibility, which is defined as the ability to anticipate the goal and predictability, which is defined as the ability to predict the trajectory. This distinction is very different to the comprehension about legibility of all the other authors presented in this review. Furthermore, additionally to the proposed model to generate legible as well as predictable goal directed robot arm motions Dragan et al. [31,32] presented the design and results of experiments where legibility ,as well as predictability, were measured. Their experimental results support their hypothesis that legibility and predictability are contradictory properties of motion.

In the following, we elaborate on the experiments measuring legibility in human-robot interaction and then subsequently present our own published work on legibility. Eyssel et al. [34] investigated the effect of predictability on anthropomorphism and acceptance of a robot in a video based experiment. They found no significant correlation between predictability and anthropomorphism, but a significant correlation between predictability and acceptance. Takayama et al. [106] used animation techniques in order to let the robot show forethought. They assume that showing forethought would increase the readability (= legibility) of robot behavior and also influence the perception of the robot. They presented the design and the results of a video-based experiment measuring legibility and other factors like intelligence and safety. Results could not support their hypothesis that forethought increases legibility, but they could show *"forethought makes people more sure of their interpretations of robot behavior, and make the robot seem more appealing and approachable."* One real-live experiment was conducted by Basili et al. [8]. The participants had to predict the goal of an approaching robot. The experimental design is similar to the experiments presented by Dragan et al. [31, 32], however in [8] the robot mimics human motions. Basili et al. compared the human-robot results with the results of a human-human experiment and found out that although the robot behaves the same way as humans the ability to predict the goal of the robot was three-fold below the human-human results. The experiment conducted by Bartot et al. [15] supports the assumption stated by Beetz et al. [11], that stereotypical motions, which are more predictable than variable motions, are leading to increased human well-being. Nevertheless, legibility was not measured in the experiment. In our own work regarding legibility [73–75] we first introduced a framework to measure legibility in human-robot interaction [74] and presented the design and results of two different experiments where we measured legibility and other factors like comfort, reliability and perceived safety of state-of-the-art navigation methods.

One interesting thing we want to point out is that a vast amount of early work in the field of legible robot behavior was done within the COGNIRON project `http://www.cogniron.org` [1, 2, 20, 21, 23, 27, 86, 97–100]. The earlier mentioned human aware motion and manipulation planner and the studies conducted by Dautenhahn [23, 27] were an outcome of the COGNIRON project. Based on our review we can claim that the COGNIRON project was an important starting point for legibility as an HRI key factor and the development of methods providing legible robot behavior.

2.3.2. How is legibility defined?

A first definition of legibility was given by Nehaniv et al. [86]. Legibility is defined as "*making the robot's actions and behaviour **understandable** and **predictable** to a human*." Clodic et al. [20] stated that a robot "*needs to be able **to explain its task** by exhibiting a legible behavior*." Similar to the aforementioned definitions is the definition given in [2, 100], where the authors wrote *"a motion is legible when a human partner can easily **understand** the robot's **intentions** by observing its motions"* which is in turn very similar to the definition from Kirsch et al. [61] *"We use the term legible to describe a behavior that is intuitively **understood** by humans."* Sisbot and Alami extended this definition in [96, 101] and stated: *"The human partner should **understand** clearly the **intention** of the robot **without further communication.**"* A more specific definition regarding legibility of robot motions is given in [8]: *"The human is able to **attribute goals** and **predict** motion **trajectories**"* and in [11] *"humans can recognize the intentions of the motion."* Another definition regarding legible robot motion is given by Guzzi et al. [45]. They state that a legible motion means *"a person observing robot motion can intuitively **understand the spatial target** the robot is heading to"*. Furthermore, another definition in the same line is given in [96] *" With a legible motion, the robot must **make clear its intention**"* or in [64, 67] *"legibility means that an ordinary, uninstructed person can **understand** and **anticipate** the robot's **actions**."* Our own definition of legibility conforms to the definitions stated so far. In [73, 74] we define that *"robot behavior is legible, if a human can **infer** the next **actions**, **goals** and **intentions** of the robot with high accuracy and confidence,"* which is extended in [75] with *"and the robot behavior **fulfills the expectations** of human interaction partner."*

Another aspect of legibility was pointed out by Takayama et al. [106]. In addition to the aforementioned factors of legible robot behavior, understandability and predictability Takayama et al. [106] adds the factor **effectiveness**. They state that robot behavior is legible — in their work they call it readable - when *"people can figure out what the robot is doing, reasonably **predict** what the robot **will do next**, and ultimately interact with the robot in an **effective way**."*

The definition given by Gielniak et al. [40] refers to communicative gesture motions, which are passing a particular information. They state that a legible, they call it anticipatory, motion is an **intent expressive** motion, meaning that a human can **understand** the communicated **message**.

According to the aforementioned definitions of legibility we can conclude the following factors defining legible robot behavior:

- understandable intentions ([2, 11, 20, 61, 64, 67, 73, 74, 86, 96, 100, 101, 106])
- understandable message ([40, 86])
- predictable actions ([64, 67, 73–75, 86, 106])
- predictable motion trajectories ([8, 86])
- predictable goals ([8, 45, 73, 74, 86])
- fulfills expectations ([75])
- effective interaction ([45, 106])

Only Dragan et al. [31, 32] deviate from these common definitions and present a new and very interesting finding regarding legible grasping motions. In their work they differentiate between legibility, which is defined as the ability to anticipate the goal, and predictability, which is defined as the ability to predict the trajectory. We will further investigate and discuss these differences later on (See section 6.2. on page 93).

2.3.2.1. Common Definition of Legibility

From the definitions that most researchers agree on, we want to extract a compact definition of legibility in the following. For this common definition we first summarize the above mentioned factors. The term *understandable intentions* includes the factors *understandable message, predictable action, predictable motion trajectory* and *predictable goal*. Because, if the human observer or interactor understands the robots' intentions — meaning the human understands what the robot is going to do, knows its plans and purpose, and anticipates the outcome — then the ability to predict actions, trajectories, and goals is included. Furthermore, when the human observer understands the intention of a gesture, then he/she understands the message, the robot wants to communicate. Therefore, we claim that *understandable intention* is an umbrella term for all these factors. The factor effective interaction can more be seen as a correlated factor of legibility. Effectivity is more a consequence of legible behavior and not a factor defining legibility. Therefore, we exclude *effective interaction* from our common definition and shift this factor to the *correlated factors*. To conclude, after summarizing and excluding factors defining legibility we end up with the two factors *understandable intentions* and *fulfills expectations*. Out of this we derive the following definition of legible robot behavior:

Definition 2 (Legible Robot Behavior)**.** *Robot behavior is legible if: (Factor 1) a human observer or interactor is able to understand its intentions, and (Factor 2) the behavior met the expectations of the human observer or interactor.*

At this point, we want to state, that only Dragan et al. [31, 32] deviates from these general definition. In their work they differentiate between legibility, which is defined as the ability to anticipate the goal, and predictability, which is defined as the ability to predict the trajectory. We will further investigate and discuss these differences later on (See section 6.2. on page 93).

2.3.3. How is Legibility Measured?

Methods to measure legibility are described in 12 publications. In the following, we will give an overview of the presented methods and briefly report the results. Within this review we only concentrate on the methods to measure legibility and disregard the measurement of other factors.

In a study presented by Dautenhahn et al. [27], the participants were asked about their desires for a future robot companion. Among other factors, the participants were directly asked to rate how predictable a future robot should be on a 5-point Likert scale. No specific robot behavior was evaluated. Most participants (54%) wanted a high predictable or predictable (36%) robot behavior, only 11% were neutral about the potential predictability. These results show us how important the factor predictability and ,therefore,legibility be for the development of robot behavior.

In the experiments presented in [15, 34] predictability was used as an independent variable in order to investigate correlations. Eyssel et al. [34] conducted an experiment in order to examine the correlation between predictability and anthropomorphism, as well as acceptance. They showed the participants a video in order to introduce the robot. A short description. followed the video. By chance, the participants got one of two different descriptions manipulating the perceived predictability of the robot. However, high predictable or low predictable robot behavior was imagined by the participants and not implemented. Hence, again no specific robot behavior was evaluated. After seeing the video, the participants were asked to complete a questionnaire with questions measuring anthropomorphism and acceptance. They did not find a significant correlation between predictability and anthropomorphism, but they could show a correlation between predictability and acceptance. A real life experiment using an industrial robot arm was conducted by Bortot et al. [15] to examine the influence of the arm trajectory on user well-being and performance. They hypothesized that a variable trajectory (= unpredictable) leads to lower results in performance and well-being. They tested this hypothesis in an experiment where the participants had to solve two different tasks while sitting at a workbench. The robot arm moved in five different conditions, whereby one condition was the unpredictable condition (=variable), see Fig Fig. 7 on the next page. Human performance and well-being were measured by evaluating the task performance and by using the NASA-TLX[6] [50], STAI-S [102] questionnaires as well as subjective evaluations. Results supported their hypothesis. Therefore, we can conclude that a stereotypical motion leads to higher performance and human-well being.

The aforementioned Human Aware Motion Planner (HAMP) was evaluated by Dehais et al. in [28]. They tested three different motion trajectories (HAMP vs. HAMP without grasp detection vs. No HAMP) in a human-robot hand over task. After each hand-over the participants were asked to rate legibility, as well as safety and comfort on a 9-point visual analog scale. Additionally to the questionnaires they used also psychophysiological measures (skin conductance, EMG, occulometry) in order to get objective measures for comfort and safety. As hypothesized by the authors the HAMP trajectory revealed the best results from the ratings as well as from the psychophysiological measures and from the results we can assume a correlation between legibility, safety and comfort.

In [31, 32, 40] the authors let the participants predict the intention of a robot motion and used the

[6]NASA-TLX: http://humansystems.arc.nasa.gov/groups/TLX/index.html

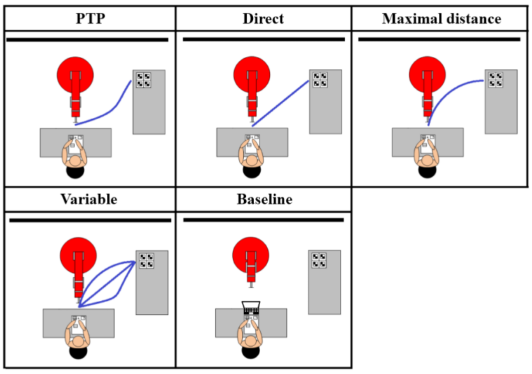


Fig. 7: Design of the experiment conducted by Bortot et al. [15] in order to investigate the influence of the robot arm trajectory human on performance and well-being. (Picture taken from [15])

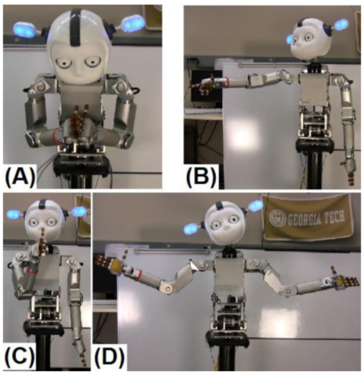


Fig. 8: Example of gestures used in the experiment conducted by Gielniak et al. [40] in order to validate their proposed method to increase legibility of a given communicative gesture. The pictures are showing the gestures (A) Bow (B) Point (C) Shhh... (D) I Don't Know. (Picture taken from [40]).

prediction time to measure legibility of developed methods following the assumption that a shorter prediction time indicates a higher legibility. The purpose of the experiment Gielniak et al. [40] conducted was to verify that their proposed algorithm increases the legibility of a given gesture, see Figure Fig. 8. They showed the participants videos of six different communicative gestures (beckon, stop, I don't know, wave, point, bow) performed by a robot, randomly either the original gesture or the modified gesture. The participants were instructed to press a button immediately when they feel certain to predict the meaning of the gesture. The time the participants need to anticipate the gestures meaning was measured. They could show that their method increases legibility of a communicative gesture. Dragan et al. [31, 32] differentiate between legibility and predictability and proposed two models generating either legible or predictable robot arm motions. Therefore, the objective of their experiments was to evaluate the proposed models [31] and to evaluate model parameters [32]. In both publications, the method is equal. To measure the legibility they used a video based method similar to the aforementioned experiments. Dragan et al. showed their participants videos of different robot arm trajectories (see Figure Fig. 9 on the next page) towards one of two goals. They asked the participants to stop the video as soon as they felt confident to predict the goal and measured the prediction time. This methodology to measure legibility is also used by Gielniak et al. [40]. Dragan et al. combined the time measure and correctness of the goal prediction to one score measure: guessing wrong gets a score of 0, and guessing right gets a higher score if it happens earlier. Predictability was measured by asking the participants to draw the expected trajectory and showing them the trajectory videos and

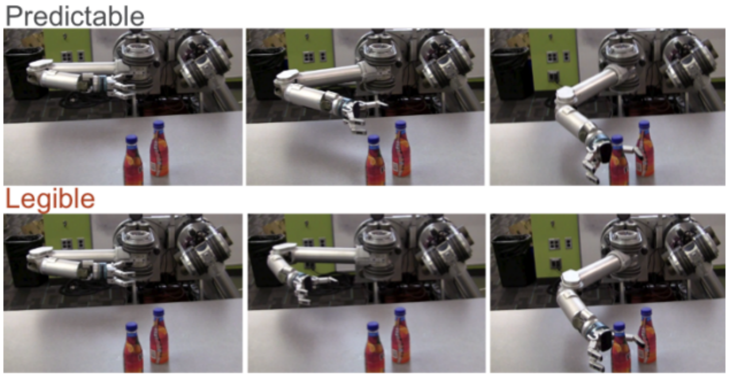


Fig. 9: Experimental setup presented by Dargan et al. [31]. They showed the participants videos of a robot arm performing different trajectories towards one of the two goals and asked the participants to stop the video by pressing a button as soon as they felt confident to predict the goal. (Picture taken from [31]).

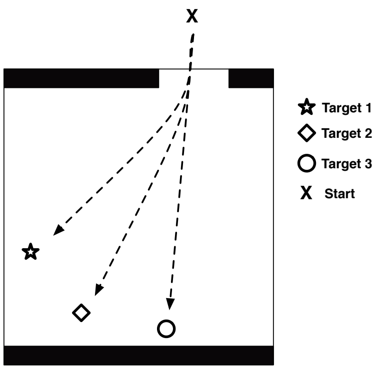


Fig. 10: Experimental Setup from an experiment conducted by Basili et al. [8]. The participant were placed on one target and the human or robot is either approaching the participant or another target.(Picture based on [8]).

asking them to rate how much the behavior met their expectations. Results showed that, according to their definitions, predictable trajectories and legible trajectories are different.

In the following, we present experiments where participants were asked to predict the robot's intention. In this case, a high rate of correct predictions indicates a high legibility. A real world experiment was conducted by Basili et al. [8]. They address the question if a motion is more predictable when a human or a robot is performing it? Therefore, they tested three different conditions, (1) a human wearing sunglasses and a scarf to avoid gazing effects, (2) a human with gaze effects, and (3) a robot. All three were performing similar motions. Figure Fig. 10 on the following page depicts the experimental setup. The participants were placed by chance on one of the three target positions (see Figure Fig. 10) and the human/robot is either heading towards them or to another target. The participants were asked to predict as soon as possible whether or not the human/robot was to approach the participant. Results show that the legibility was higher for the two human conditions. The "gaze" condition scored with 98.9 % correct answers, "no gaze" with 93.55% and the robot condition with 81.25 % correct answers. Takayama et al. [106] conducted an experiment to address the question: does the use of animation principles to show forethought make the robot behavior more legible and how does it influence people's perception of robots? The authors also tested whether showing goal-oriented responses

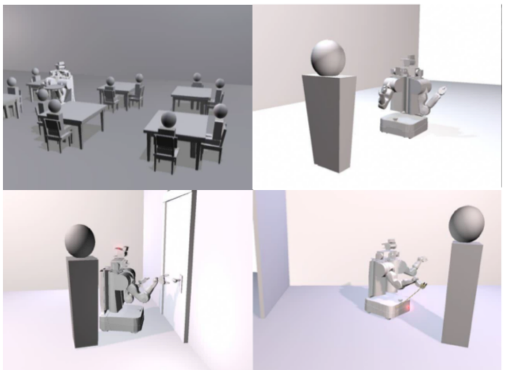


Fig. 11: Experimental Setup from an experiment conducted by Takayama et al. [106] in order to investigate if showing forethought increases legibility. (Picture taken from [106]).

Fig. 12: Experimental setup from our own experiment [73] where we measured the legibility of different navigation algorithms. The robot moves to one of the three tables and the human is crossing the robot's path. (Picture taken from [73]).

influenced people's perception, but this is irrelevant for our legibility review. They prepared simulated videos of a robot performing four different tasks (opening a door, delivering a drink, ushering a person into a room, requesting help from a person to plug into an outlet) either showing forethought or not (see Figure Fig. 11). The videos were divided into two parts, the first part shows the behavior of the robot before the robot actually opens the door, delivers the drink, asks for help, or directs the person. The second part shows the rest of the task. In their online study, the participants saw the first part of the video and were asked to predict the robot's intentions and how confident they felt about their answer. Results show that forethought was not significantly correlated to legibility, measured with the right prediction. However, the participants were significantly more sure of their predictions. The method of our own experiments [73, 75] regarding legible robot navigation is derived from the aforementioned method from Takayama et al. [106] to measure legibility. In both experiments we wanted to investigate how legible different state-of the-art navigation methods are for a human observer. In the first experiment [73] we recorded short simulator videos with a robot moving to one of three goals and a human crossing the robot's path (see Figure 2.3.3.) using four different navigation methods. As in Takayama et al. [106] we divided the video into two parts. The first part ends before the human crosses the robot's path. We showed the participants the first part and asked them to predict the goal and to rate their confidence about their prediction. Afterwards we showed them the second part and asked them if the robot's actual behavior was as expected and how surprising it was. With these three questions we measured the legibility of the four different navigation methods. We found no significant association between the number of correct goal predictions, but the ratings for expectation and surprise differed significantly, which is in turn very similar to the results in [106]. In our second experiment [75] we used a similar method to investigate the correlation between legibility and

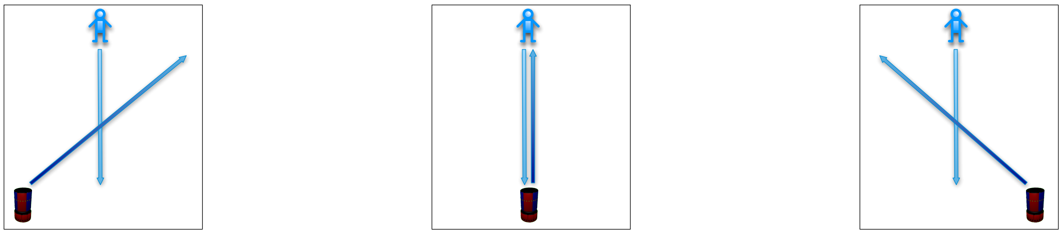


Fig. 13: Experimental setup fom [73]. The robot moves to one of the three tables and the human is crossing the robot's path. (Picture taken from [73]).

perceived safety. Different from [73] we used only two different navigation methods and prepared real world videos using a head mounted camera to provide a first-person perspective. As depicted in Figure Fig. 13 the robot approached the human frontally and from each side. Similar to [75, 106] we split up the videos at the point before the robot crosses the human's path. After the participants had seen the first part we asked them to predict the direction and velocity. After we showed them also the second part and asked them to rate whether the robot's behavior met their expectation. Additionally, we used the Godspeed V questionnaire [7] to measure perceived safety. Results show a significant correlation between perceived safety and legibility measured with correct predictions as well with the expectation ratings.

In [74] we present different approaches of how legibility can be measured. Additionally to the formerly described use of questionnaires, we suggest to use psychophysiological measures. We recommend to measure prediction times, idle times, gaze behavior and skin conductance to assess how the observed robot behavior was perceived by humans. Prediction times were already used by [31,32,40] as well as gaze behavior and skin conductance by [28].

To conclude, within our review we found the following methods to measure legibility:

- Either show the participants robot motions/behavior or let them interact directly with the robot. Afterwards ask the participants to rate how legible the behavior was perceived [28].
- Either show the participants robot motions/behavior or let them interact directly with the robot and use psychophysiological measures like skin conductance, gaze behavior (pupil size, fixations, saccades), to measure the arousal and anxiety. Thus, one gets an indirect measure for legibility. [28, 74]
- Either show the participants robot motions/behavior or let them interact directly with the robot and ask to rate how surprising the behavior was perceived or if the behavior met the participants expectations. [31, 73, 75]
- Show the participants robot motions/behavior and ask them to indicate immediately when they feel certain to predict the robot's intentions/goals and measure the time as indicator for legibility. Additionally, correctness and timing can be combined to one measure. [8,31,32,40]
- Show the participants a video of robot motions/behavior and stop the video before the intention is clear and ask the participants to predict the robot's intentions. [73, 75, 106]

- In order to capture humans' expectations about robot motions one can ask participants to draw the expected trajectory [31, 32]

Furthermore, legibility was used as independent variable in order to investigate how legibility influences other factors [15, 34]. Another indirect measuring method is to measure correlated factors of legibility. We present our findings regarding correlations of legibility in section 2.3.5. on page 25.

2.3.4. Methods to Achieve Legible Robot Behavior

The question we want to answer in the following is how we can generate legible robot behavior. First of all we will elaborate the proposed assumptions and approaches on how to create legible robot behavior. Later on we describe briefly the proposed methods.

2.3.4.1. Assumptions and Approaches

One obvious assumption regarding legible robot behavior is that **human-like** behavior would be perceived as legible [11, 45, 64] because human behavior is well-known for humans. Therefore, the development of methods imitating human motions is very common in the HRI community.

Furthermore, Beetz et al. [11] claimed that a **stereotypical motion** is predictable and thus legible. Results from Bortot et al. [15] are supporting this assumption. In [86] the authors claim that the use of **complementary motions** made by the robot could achieve legibility (e.g., during a hand-over arm motion looking to the object). Therefore, using their proposed gesture classification can improve the legibility of the robot. Furthermore, [96] integrated complementary gestures in order to make the motion more intend expressive. Moreover, [106] claims that the **use of animation principles** makes the robot behavior more legible. They implemented additional gestures in order to let the robot show forethought and the results of their conducted study supported their assumption. The "complementary gesture" assumption is also supported by results from [8]. They could show that gaze behavior increases the ability to predict where someone is heading to.

Another assumption is to **take into account social constraints, personal preferences and capabilities** [2, 61]. Following this Kirsch et al. [60, 61] proposed an approach to achieve legible task execution behavior. They suggest to learn personal preferences and capabilities in order to integrate this knowledge into a high level task planner.

Visibility is not only a prerequisite for legibility, because a human is not able to anticipate anything from a hidden motion, it is also a crucial factor for generating legible motions. Sisbot et al. [98] claims *" a first step is to make the robot motion legible, is to make the handling position as visible as possible."*. Dehais et al. [28] shares this view. This assumption is implemented in the Human Aware Motion Planner [80, 96, 99, 101] as well as in the Human Aware Navigation Planner [97, 100]. The visibility assumption is based on results from Dautenhahn et al. [23].

To conclude, in order to generate legible robot behavior, the following assumptions were proposed in the reviewed articles:

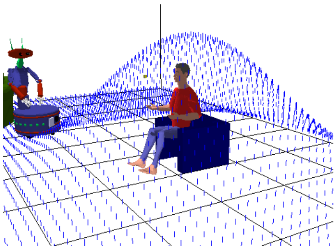
- model human-like behavior [11, 45, 64]
- generate stereotypical motions [11, 15]
- add complementary motions (gestures) in order to clarify intentions (e.g. gaze, pointing, use animation principles) [8, 86, 96, 106]
- take into account social constraints, human abilities, and preferences [2, 60, 61]
- robot motion must be as visible as possible [28, 98].

2.3.4.2. Methods

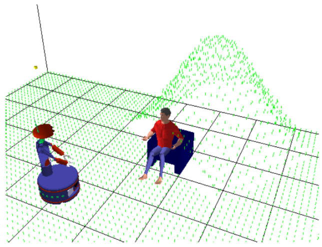
In the following, we briefly present the proposed methods to generate legible robot behavior. We will only describe the idea of the methods, for detailed descriptions we refer to the respective publications. In the literature at hand 18 articles are proposing methods to create legible robot behavior like legible robot high-level plans [1,2,60], legible navigation [45,64,67,97,100], legible arm motions [11,31,32], legible hand-over motions [79, 80, 96, 98, 99, 101], or present techniques to maximize the legibility of a gesture [40].

Legible Robot Task Planning The Human Aware Task Planning method (HATP) proposed by Alami et al. [1, 2] and Alili et al. [3] considers personal preferences and social constraints. The authors assume that taking into account personal preferences and social constraints leads to legible plans. The HATP planning process is an extension of the SHOP2 planner [85], which permits to specify cost functions. In order to select the preferred behavior they used cost functions denoting the difficulty and pleasure an agent has in action realization, undesirable states, as well as undesirable action sequences, and representing social or cultural constraints. Kirsch et al. [60] suggest to use learning methods in order to acquire knowledge about personal preferences in order to integrate this information into the planning process. In [60] they recommend to use the Robot Learning Language (RoLL) [59] for learning and updating personal preference models.

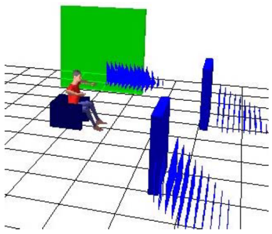
Legible Robot Navigation In order to consider social constraints like safety, comfort and **visibility** for navigation path planning Sisbot et al. [97, 100] model these constraints as cost functions (see Figure Fig. 14 on the next page) and use a classical A* path planning algorithm to compute legible paths. In [67] Kruse et al. propose a modification of the aforementioned HANP in order to achieve cooperative behavior that is more suitable for moving humans. They relaxed the safety cost function and shifted the safety constraint to the plan execution level to exploit the fact that the human might make room for the robot. Furthermore, they add a predictive cost function for moving humans to consider the future path of the human (see Figure 2.3.4.2. on the facing page). Navigation methods,



(a) Safety cost function — as farther the robot is from the human the safer he interaction is — high costs for low distances



(b) Visibility cost function - based on the human field of view, high costs for invisible regions



(c) Visibility cost function representing hidden zones — high costs for regions covered by obstacles

Fig. 14: Visualisation of the cost functions used by Sisbot et al. [97, 100] to model social constraints like safety and visibility in order to generate legible navigation paths [97, 100] or hand-over motions. (Picture taken from [97]).

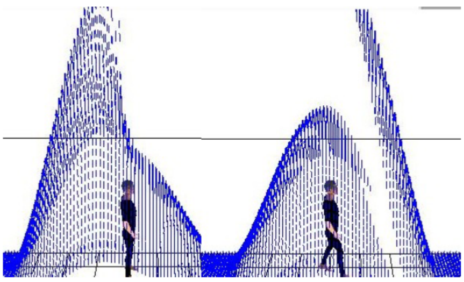


Fig. 15: Visualisation of the cost functions used by Kruse et al. [67] to model moving humans for robot path planning. As you can see in the right picture (moving) the costs in front of the human are higher to avoid moving in the humans way.

imitating human behavior, were implemented by Kruse et al. [64] and Guzzi et al. [45]. Based on results of a human-human path crossing experiment, Kruse et al. [64] added a context-dependent cost model to their extension of HANP. Consequently, they imitate human behavior of moving straight forward and decreasing the velocity. Another human-inspired approach was presented by Guzzi et al. [45]. Contrary to Kruse et al. [64], who are considering one to one path crossing situation, the navigation algorithm implemented by Guzzi et al. [45] imitates human pedestrian behavior in crowded multi-agent scenarios. The proactive method attempts to avoid potential collisions by using two simple heuristic rules to determine first the optimal direction towards the goal taking into account obstacles and other moving agents and second to determine a safe and smooth velocity adjustment.

Legible Arm Motions Beetz et al. [11] use Dynamic Movement Primitives to learn human reaching trajectories from a motion capturing dataset. These motion primitives were used to generate human like trajectories for a robot arm. A different approach was presented by Dragan et al. [31, 32]. They model the observers expectations as a cost function C in order to find the optimal trajectory. Due to their definition of a legible trajectory and the assumption that a goal is more likely when the movement is towards it and away from the other goal, the cost function C computes the probability for a goal given a trajectory. Thus, they compute the optimal trajectory in terms of the ability to predict the goal as soon as possible. Predictability is modeled as an efficient trajectory with a cost function C determined by the path length.

Legible Hand-Over Motions In order to generate legible hand-over motions Sisbot et al. developed the Human-Aware-Manipulation-Planner (HAMP) based on the Human Aware Navigation Planner HANP using the same cost function approach [96, 98, 99, 101]. The planner consists of three stages. First, the planner determines the optimal spatial coordinates for the object exchange point that is safe, visible, and comfortable to reach for a human. To achieve this, they search for a point with the lowest costs considering safety (see Figure Fig. 14 on the previous page (a)) and visibility (see Figure Fig. 14 on the preceding page (b)) cost functions, which had already been used in the HANP. Furthermore, they use an arm comfort cost function, which determines costs representing how comfortable it is for a human arm to reach a point, based on the kinematic model. The second step is to find the optimal path for the object from the initial position to the object exchange point. Equally to the method used for HANP, the motion planner calculates the optimal path for the object transfer using the A* algorithm and the safety and visibility cost functions (see Figure Fig. 14 on the previous page). Third, the robot motion has to be calculated considering the object path. To this end, they use a generalized inverse kinematics algorithm [99] and a Soft Motion Trajectory Planner, which transforms a path into a trajectory with bounded jerk and acceleration, thus generating a smooth robot motion that is aimed to be more natural and human friendly. [101]. Furthermore, Sisbot et al. [98, 101] added additional camera and head motions in order to increase legibility by expressing intention using gaze behavior.The robot moves its head and cameras to look at the object during the handover task. Mainprice et al. [79, 80] enhanced the existing HAMP by using a different path planning algorithm.

Enhance Legibility Takayama et al. [106] increased the legibility of a given behavior, like opening the door, by adding an additional gesture showing forethought to the original behavior. Their approach is to use animation principles from the Disney Studio [109]. However, the motions are animated using an animation software suite and not implemented (see Figure Fig. 11 on page 19). Another approach to increase legibility is the algorithm presented by Gielniak et al. [40] to enhance the legibility of a given gesture. They analyzed a given gesture, e.g. an *"I don't know"* gesture, in order to find the body configuration, which represents a symbol. A symbol is defined as the most expressive part of the gesture (see Figure Fig. 8 on page 17). Afterwards, they created a motion from the original gesture motion in which the *symbol* occurs as soon as possible.

2.3.5. HRI Properties Correlated to Legibility

In the last section of our literature review we present the results of our research regarding the question, which other HRI factors are correlated with legibility. In seven publications correlations are assumed and in seven scientifically evaluated.

The HRI properties, safety and comfort, seem very likely to be correlated with legibility because an unforeseeable and sudden movement of the robot could cause a collision, let the human feel unsafe and also very uncomfortable. Sisbot et al. [96] wrote in the context of the Human Aware Motion Planner that *"even if the robot position and objects position will be "good," an unclear motion that does not reflect the robot's intention (= not legible) can **surprise** human and cause **discomfort**."* Furthermore, in [101] Sisbot et al. stated *"a legible interaction adds to safety."* Also Kirsch et al. [61] claimed *"legibility is a prerequisite to establish human **comfort** and **perceived safety**, because if the intention of the robot is understandable, its actions can be expected and are not felt as a threat."* The statements are in line with Kruse et al. [67]. Dehais et al. [28] conducted an experiment to investigate the influence of legibility on safety and comfort. Although, the authors did not calculate a correlation coefficient, their presented results let us presume a correlation between legibility, safety and physical comfort (see Table 2.1).

Subjective variables	Motion-1	Motion-2	Motion -3
Legibility	7.33 ($\pm$ 1.18)	4.00 ($\pm$ 0.72)	3.58($\pm$ 0.54)
Safety	7.00 ($\pm$ 0.39)	2.25 ($\pm$ 1.05)	4.66 ($\pm$ 0.57)
Comfort	6.33 ($\pm$ 0.43)	2.83 ($\pm$ 0.92)	1.83 ($\pm$ 1.03)

Table 2.1: human-robot regarding legibility, safety and comfort using a 9-point visual analog scale (1 for very low, 9 for very high. The values were obtained by Dehais et al. [28].

Similar to the investigations of Dehais et al. [28] we also evaluated the legibility, safety and comfort of different algorithms [73]. The results let us also presume a correlation between legibility, safety, and comfort[7]. In our second experiment, where we further investigated the influence of legibility on perceived safety [75] we found a significant correlation between legibility and safety.

Another factor influenced by legibility is the **efficiency** of the interaction. Guzzi et al. [45] explained

[7] A subsequent data analysis on the dataset revealed a significant correlation between legibility (measured by the level of surprise) and safety ($r = -0.323^{**}$),as well as legibility and comfort ($r = -0.395^{**}$). ($^{**}: p < 0.01$)

this correlation with a vivid example from robot navigation: *"If navigation algorithms generate unpredictable trajectories, humans have to change their local plans frequently to move around the robots, ultimately resulting in less efficient navigation for both groups."* Also Kruse et al. [64, 67] assumed a correlation between legibility and the efficiency of the interaction. Bortot et al. [15] measured the human performance in a cooperative task and their results confirmed the assumption.

Takayama et al. [106] investigated the factors appealing, intelligence, competence, safety, approachable and confident, but their results do not show any clear correlation. The effect of legibility on anthropomorphism was studied by Eyssel et al. [34]. However, they found no significant correlation between the factors. In [74] we assume that the legibility of the robot behavior influences the perceived value of a robot.

One insight of an experiment conducted by Dragan et al. [31] is that the factors legibility and predictability of goal directed arm motions are inversely correlated.

We conclude from our literature review that the following factors are correlated with legibility:

- safety (assumed in [61, 101], experimental investigations in [28, 75])
- comfort (assumed in [61, 96], experimental investigations in [28, 73])
- surprise (assumed in [79, 96], experimental investigations in [73, 75])
- efficiency (assumed in [45, 64, 67], experimental investigations in [15])
- perceived value (assumed in [74, 75])

2.4. Conclusion

With the review at hand we summarized the findings, assumptions and methods other researchers made and developed regarding legible robot behavior. First of all, we concluded a definition for legible robot behavior. Robot motion is legible if a human observer/interactor is able to **(1) understand the robot's intentions**, **(2) the robot behavior meets expectations**, and **(3) both are interacting in an effective way**. Intention understanding means that a human can figure out what the robot is actually doing. In particular, regarding our navigation scenario, it means that the human interactor can predict the robots behavior and/or the goal of the movement. In order to achieve legible robot behavior we fund the assumptions in the reviewed articles, that a motion is legible when it is human-like, stereotypical, efficient, as visible as possible, considers social constraints and human preferences. Furthermore, we fund the suggestion that and one could add complementary motions to clarify the robots' intentions. We presented several ways to measure legibility (see Section 2.3.3. on page 16) e.g. by using self-reports or by measuring the time one needs to predict the intention of a robot motion. Furthermore, we elaborated methods to implement legible motions (see Section 2.3.4. on page 21) and summarized the factors that are correlated with legibility like safety, comfort, surprise, efficiency, and the perceived value of a robot.

One very interesting thing we found is that almost all reviewed articles are holding a similar view on legibility and the results are in line with each other, except the work from Dragan et al. [31, 32]. In their experiment, they had a very interesting result regarding the factors legibility and predictability namely that *predictability and legibility are fundamentally different and often contradictory properties of motion"* [31]. As opposed to the other authors they defined and measured legibility as the ability to infer the goal as soon as possible and predictability as the ability to infer the trajectory of a robot arm moving towards one of two goals. In the other articles of the review at hand, these two features are not differentiated. Furthermore, no other HRI properties like safety, efficiency or comfort were investigated within their experiments. Considering the results from Bortot et al. [15] we can conclude that the straight and according to Dragan et al. [31] predictable motion is more efficient than the legible motion. It would be very interesting to investigate the safety and comfort of these two aspects of motion. We will further discuss the differences of Dragan et al. [31] to our work in the discussion chapter (Section 6.2. on page 93). Our overview of measurement instruments to determine the legibility of robot behavior (see Section 2.3.3. on page 16)) adds to the field of HRI metrics [7, 14, 51] and extends the common HRI measurements by the factor legibility. Nevertheless, much work is to do in order to develop further techniques to measure legibility. Furthermore, the fact that several authors present concrete methods to achieve legible robot behavior and our findings regarding correlated factors support our former assumption that legibility is one important factor of robot acceptance [51] and should line up with the HRI key concepts [7].

Limitations This review is not intended to be exhaustively investigation legible robot behavior. To this end, the field of human-robot interaction is new and emerging research area and furthermore legibility is also a very new factor for human-robot interaction. We found only 32 articles dealing with legibility. It may be, that other researchers use different terms for legible robot behavior.

3. Legibility Metrics

We developed legibility metrics based on theoretical foundations and further literature regarding the evaluation of robot behavior in common and in particular of legibility. The chapter at hand elaborates measurements, methods and experimental designs to evaluate legibility of robot behavior.

3.1. Introduction

In the previous chapter, we proposed a definition for legible robot behavior. In the following, we address the question of how can we measure legibility in order to enable researchers to evaluate their algorithms and approaches regarding legibility. Without the possibility to objectively assess robot behavior regarding legibility, we cannot judge if an approach increases legibility or not. Furthermore, Steinfeld et al. [103] stated that HRI needs more metrics to investigate the interactions and Bethel et al. [14] pointed out that the development of methods to evaluate HRI scenarios are lacking. In 2008 Burghart and Steinfeld organized a Workshop at the HRI Conference in Amsterdam [17] in order to foster the discussion and the development of metrics for human-robot interaction. With the chapter at hand, our objective is to propose measurements as well as methods to assess legibility of robot behavior based on our experiences and further literature.

As examples for related work regarding metrics for HRI we refer to Bartneck et al. [7], Heerink et al. [51], and Weiss et al. [113]. Bartneck et al. [7] presented the commonly known Godspeed questionnaire to measure the five HRI key concepts anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety that is widely used in the HRI community (e.g., in [36, 42, 110]). Heerink et al. [51] presented a questionnaire based toolkit to measure a robot's social acceptance. (used e.g. by [35,42,55]) Another questionnaire based metric to evaluate social acceptance of a robot is presented by Weiss et al. [113]. Both questionnaires are based on the UTAUT (Unified Theory of Acceptance and Use of Technology) model. All three metrics are self-assessment methods using questionnaires as in most of the HRI studies conducted so far [13]. However, with the chapter at hand we also propose evaluation techniques for other experimental methodologies like psychophysiological measurements or behavioral measures in order to move a first step towards common legibility metrics.

In the following, we first propose dependent measures to evaluate legibility of robot behavior in HRI scenarios. Afterwards, we elaborate evaluation methods for each of the five primary methods of evaluation in HRI [13]: (1) self-assessment, (2) interview, (3) task performance, (4) psychophysiological measurements, and (5) behavioral measures. Finally, we propose one generic experimental setup based on our own used setup in order to give future researchers one example setup for further investigations regarding legible robot behavior.

3.2. Legibility Measures

Bartneck et al. [7] used different dependent measures to quantify each the HRI properties anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety. For example, the perceived safety was measured by asking the participants to rate their emotional state regarding anxiety, calmness, and surprise. Also Heerink et al. [51] used a list of acceptance factors like trust or perceived usefulness in order to determine the social acceptance of a robot. Just as the previously mentioned authors we also want to measure legibility by using different measurements, because it is very difficult to measure such a complex concept like legibility, which consist of different aspects. To this end, we will derive in the following measurable factors of legibility from our definition, the insights of the literature review, our own research findings, and further literature. First of all, we divide our formerly concluded definition of legibility into its parts. The factors defining legibility that we have found in our review are understandable intention, understandable message, predictable action, predictable motion, predictable goal, and expectation-fulfillment (see Sec. 2.3.2. on page 14). We can conclude that **predictability** is one measurable legibility factor and we can measure it, for example, by asking participants directly to predict the robots' goal (goal-predictability), to predict the future behavior (behavior-predictability) like the motion trajectory (trajectory-predictability), or to predict the meaning of a gesture (message-predictability). Furthermore, we can ask if the behavior meets the participant's expectation. To this end, the second measure is the **expectation-fulfillment** factor. Moreover, if the robot behavior meets the participant's expectation, then the participant is not surprised while otherwise he/she is surprised. The **surprise** factor is an inverse measure, which means that high values for surprise indicating a low legibility. Surprise is obviously correlated with the expectation-fulfillment factor. We proved this correlation in one of our legibility experiments [72]. Another factor we can measure is **efficiency** and **effectivity**. Currently, this correlation is only vaguely verified by Bortot et al. [15] and needs further investigation, but we strongly assume that legibility increases the efficiency, and also the effectivity of a cooperative task. For example, an assistive robot that should help a physically disabled person to prepare a meal. The robot fulfills its duty but, the robot manipulates objects with sudden, unpredictable movements. It rushes through the kitchen with rapid changes of direction, ignores obvious errors like a pot not being placed properly on the stove, or performs complex procedures in a strange order like heating up the stove before placing the pot. Even if this robot will eventually somehow help to prepare a meal, the ineligible behavior would lead to a less efficient and also effective behavior. Another investigation of the efficiency effect is investigated by Lu et al. [78]. In their experiment, they examined human-robot path crossings in a hallway setting and found that the interaction is way more efficient when the robot starts to avoid the person at an earlier time, which is more legible because the participant can predict the robot movements at an earlier time-point.

To conclude, we suggest the following legibility measurements:

- predictability
- expectation-fulfillment
- surprise

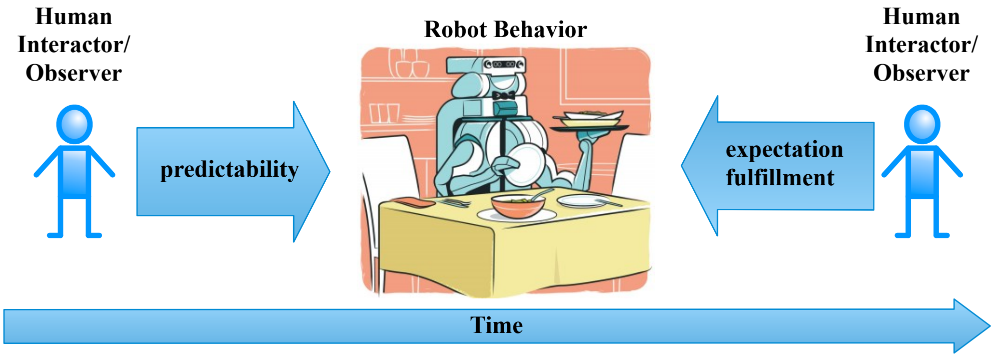


Fig. 16: This picture depicts the moment at which of the two legibility measures predictabilits and expectation fulfillment has to be measured.

- efficiency & effectivity

We want to point out that this list of measurements is a first step towards legibility metrics. Much further research has to be done to verify the measurements as well as to find more measures. Furthermore, we can measure everything what is connected to surprise and irritation using psychophysiological measures. We will briefly discuss this measurement methods in the following.

3.3. Timing for Legibility Measurements

Another fact we want to mention is the timing for the two major legibility factors predictability and expectation-fulfillment, which is important for the experimental design. According to our definition we have two different moments for measuring. As depicted in Figure Fig. 16 one has to measure predictability before or during the robot behavior. Because, one can only predict what happens in the future before the behavior happened. The expectation-fulfillment factor has to be measured afterwards, because one can only determine if the behavior met his/her expectation when the interaction has occurred. The same is applicable for all the other factors, surprise, which is correlated with expectation-fulfillment can be measured afterwards, this also applies to efficiency & effectivity and hesitation.

3.4. Legibility Evaluation Methods

In the following, we elaborate different evaluation methodologies how to measure the different legibility measurements in an HRI experiment. We suggest methods for each of the five primary methods of evaluation for HRI studies [13]: (1) self-assessment, (2) interview, (3) task performance, (4) psychophysiological measurements, and (5) behavioral measures. For a detailed description of the five primary methods and for an overview of how to plan and conduct HRI experiments we refer to Bethel et al. [13].

3.4.1. Self-Assessment

The easiest and according to Bethel et al. [13] the most common evaluation method in the HRI community is to use self-reports like questionnaires. By using a questionnaire we ask the participants directly to judge, to rate, or to draw. In the following, we give some examples of how to measure the different legibility properties by using a questionnaire. However, one very simple possibility to measure legibility, which Dehais et al. [28] implemented in an experiment, is to ask the participants to rate the legibility directly.

Predictability One can ask the participants to predict the future behavior of the robot, before the behavior happened (see Figure Fig. 16 on the previous page), to measure the predictability of the robot's behavior. Here, one can use different scales by providing a selection of possible outcomes: a yes/no scale and ask the participant to judge which behavior the robot will perform or a Likert / Visual-Analog Rating scale by asking the participant to rate how likely each of the listed behaviors is. Figure Fig. 17 depicts an example questionnaire to measure predictability. For more examples we refer to [73, 75, 106].

How will the robot behave in the immediate future ?

Yes/No Scale

☐ drive on
☐ drive to the left side
☐ drive to the right side
☐ stop
☐ drive backwards

Likert Scale

Please rate how likely you think each behavior is

drive on	very likely	☐☐☐☐☐☐☐	very unlikely
drive to the left side	very likely	☐☐☐☐☐☐☐	very unlikely
drive to the right side	very likely	☐☐☐☐☐☐☐	very unlikely
stop	very likely	☐☐☐☐☐☐☐	very unlikely
drive backwards	very likely	☐☐☐☐☐☐☐	very unlikely

Fig. 17: Brief example of a questionnaire to measure predictability of robot behavior using a yes/no scale (top) and a Likert scale (bottom).

Another method to measure predictability of a motion trajectory is to ask the participants to draw the expected trajectory. Here one can show the participants the first part of the robot movement or only a picture of the start configuration and ask them to imagine the trajectory they expect the robot will perform. Then, one ask to draw what they expected. For an example of this method, we refer to [31]. The issues of this method are that one can only draw two-dimensional lines and it does not take place any real human-robot interaction. Later on we present an approach to compensate these disadvantages (see Section 3.4.5. on page 37).

Expectation-Fulfillment and Surprise As depicted in Figure Fig. 16 on the preceding page one can measure the expectation-fulfillment factor after the robot behavior has happened. Therefore, after the interaction has finished one ask the participant if or how much the actual behavior met his/her expectations. Here, one can also use a yes/no scale or a Likert scale. The same applies for the surprise factor.

3.4.2. Interview

Instead of drawing a line on a paper in order to capture the participant's expectations one can ask the participant in an interview to orally describe what they think the robot will do next. By this means, one can measure the predictability of almost all robot behaviors and is not restricted to motion trajectories.

Furthermore, one can ask the participants in a structured interview regarding the legibility measures predictability, surprise, expectation-fulfillment, effectivity, and efficiency, similarly to the questionnaire based methods.

Another interview-like technique we want to mention is the thinking-aloud method [108]. The participant is encouraged to constantly orally comment to the interaction. From the unstructured data, one have to find the indicators for legibility like finding words related to surprise in order to identify surprising behavior.

However, interview methods are rarely used in HRI experiments, due to the difficulties in the evaluation of the interview data, in particular for unstructured interviews. For a detailed discussion regarding the issues of interviews we refer to Bethel et al. [13]

3.4.3. Task Performance Metrics

By using task performance metrics, we can measure the efficiency and effectivity of a cooperative human-robot interaction task. One can use standardized task performance metrics like the NASA-TLX[1] [50] (used by [15]). Alternatively, one could ask the participants to rate the task performance of an interactive task [41]. In the following we present different task performance metrics to measure the legibility properties predictability, efficiency, and effectivity. For detailed information regarding common task performance metrics for human-robot interaction, we refer to Steinfeld et al. [103].

3.4.3.1. Prediction-Time

In order to measure predictability in a more scaled way one can capture the time a participant needs to predict the robot's intention. Therefore, we follow the assumption, that the sooner a participant can predict the right intention, the higher is the legibility of a behavior. For this purpose, we combine the questionnaire method with a task performance metric. One approach is to measure the time the

[1]NASA-TLX: http://humansystems.arc.nasa.gov/groups/TLX/index.html

participant needs to predict the future robot behavior. Another approach is to ask the participant to indicate as soon as possible once they feel sure to predict the robot behavior in the near future and let the participant then predict the future behavior. For an analysis, the time measure and the prediction should be combined (see also [31]). According to Dragan et al. [31] we suggest assigning a score of 0 for a wrong prediction and a high score for low prediction times and a low score for a high prediction time in order to get a comparable legibility score. For example, let $t_{activity}$ be the time the complete behavior needs and if all other ratings are on a 5 point scale (0-4) we can convert the prediction-time t_p and the rating to a legibility score by calculating

$$legibility = \begin{cases} 0 & \text{,if prediction is wrong} \\ \frac{t_{activity} - t_p}{t_{activity}} \cdot 4 & \text{, otherwise} \end{cases}$$

Examples for this experimental method are to be found by in [31, 40].

3.4.3.2. Distance

Another possibility to measure legibility of a motion trajectory by using something similar to the prediction-time is to measure the distance the robot has covered so far once the participant feels sure to predict the robots' intention. This method is only suitable for evaluating motion trajectories. The experimental method is the same as described above for prediction-times with the sole difference that one measures the length of the trajectory the robot has covered or one can measure the distance to the human in a navigation experiment. In Basili et al. [8] one can find an example.

3.4.3.3. Reaction Time

The time one needs to respond in a cooperative task like handing-over an object is another possibility to capture how legible the robot's behavior is for the participant. One example is a hand-over task. Both the participant and the robot are sitting at a table. The robot's task is to hand over an object the participant. The object is placed before the robot and the participant is told to lay his/her hands on the table during the idle times. For example, one can measure the time the participant needs to react by using a motion capturing system. For a detailed description regarding a reaction-time design we refer to [53].

Furthermore, one can measure the time-to-completion of a task [103] to compare different behaviors regarding their efficiency. Thereby, one follows the obvious assumption that a shorter task-completion time indicates that the task was completed in an efficient way. This measurement was used, for example, by Lu et al. [78] to compare different navigation methods and the influence of gaze behavior in a path crossing scenario.

(a) Windmill (b) Railway signal (c) Tower (d) L shape

Fig. 18: Objects that have to be assembled by the human participant and the robot in a cooperative task. (picture taken from [41])

3.4.3.4. Counting Failures

The effectivity of a task can be measured by counting the failures in an interactive situation. This method is applicable for complex interactive tasks and highly depends on the task type. For example, if we want to investigate the legibility of a robot in a cooperative assembling task, as it was performed in an experiment by Giuliani et al. [41] where the participants had to build an object (see Figure Fig. 18) together with the robot. One can realize the counting failures method by counting the trails where the object was assembled correctly. In a more sophisticated way, one can count how often the participant has to rearrange something or one could count the steps already needed and see if the number is higher or lower a previously defined number of "optimal" steps.

3.4.4. Psychophysiological Measures

First of all, we want to mention, that is it not possible to measure legibility by itself using psychophysical measures, it is only possible to measure correlated properties like arousal caused by surprise or other correlated properties which can be evaluated using psychophysiological measuring instruments. In the following, we elaborate how can we measure legibility correlated factors using psychophysiological measurements. For a detailed description on how to design such experiments we refer to Bethel et al. [12] and for a survey regarding psychophysiological measurements applied to HRI we refer to Bethel et al. [14].

Following Bethel et al. [14] the most common measures in HRI are:

- cardiovascular system (heart rate variability (HRV), respiratory sinus arrhythmia (RSA), cardiac output, interbeat interval (IBI), blood pressure (BP))
- electrodermal activity (skin conductance activity (SCA), skin conductance response (SCR))
- respiratory system (breaths per minute, respiration volume)
- muscular system (electromyography (EMG))
- brain activity (electroencephalography (EEG) and imaging)

We want to add to this list from Bethel et al. [14] the visual system, where we can measure the pupil size, fixations, and saccades of the eye [16, 95].

It is experimentally verified that legibility is correlated with surprise [72] and perceived safety [75], which is the opposite of anxiety. We also assume that legibility is correlated with the stress level and arousal of a person, because it is very obvious that an unpredictable robot behavior will cause stress in an interactive situation. Most of the mentioned psychophysiological measurements can be used to measure arousal, stress, and anxiety. To this end, we can use them to measure legibility. For example, the stress level of a person can be evaluated by measuring the heart rate and blood pressure of a participant like Rani et al. [93]. Kulic et al. [69] measured the anxiety level of their participants in a human-robot interaction experiment by using cardio measurements and EMG. They showed that skin conductance is an effective measure for arousal and anxiety. Furthermore, Preuschoff et al. [92] showed that the pupil dilation reflects surprise and Bradley et al. [16] showed that one can use the pupil size as a measure for arousal. To this end, by using an eye tracking system, we can use changes in pupil size as an indicator for surprising, thus ineligible, robot behavior.

As Bethel et al. [14] pointed out there is only very limited research in the HRI community investigating psychophysiological measures. [14]. Within our review, we found only one experiment using psychophysiological measures in a legibility experiment. It was conducted by Dehais et al. [28] to evaluate different arm movements regarding legibility, comfort, and safety. They used the psychophysiological measures skin conductance, EMG, visual fixations and saccades of the eye and an additional questionnaire. They found significant differences for all measurements and their results let us assume that the used psychophysiological measures are correlated with legibility.

Taken together, due to a lack of further investigations regarding psychophysiological measurements and legibility we are not able to convincingly show how to use psychophysiological measurements, but we highly assume that it is possible to use such measures as indicators for legibility.

3.4.5. Behavioral Measures

According to Bethel et al. [13], behavioral measures are the second most common method of evaluation in HRI studies. However, especially for legibility no experiment was conducted so far using behavioral measures. To this end, we propose some examples of how we think that one could measure legibility by using behavioral measures. As commonly used in HRI (e.g., in [23, 83]) one could video tape the interaction and let independent raters judge, for example, how much the participant was surprised by the robot's behavior or how effective and efficient the interaction was. We think that it is difficult for an external observer to rate how much the participant's expectations was met by the robot behavior. In order to propose more sophisticated behavioral measures we present in the following three implicit feedback measures, which can be directly measured, without the use of an external observer.

One behavioral measure we recommend is to identify hesitations during the human-robot interaction. In general,in hesitation is defined as follows: *"Hesitation happens when you feel uncertainty or doubt.*

It can be a pause in speech, a faltering moment before you act, or a silent second of indecision. If your friend's dog is growling and staring you down, but she tells you he won't bite, that feeling, that something is not feeling right, that stops you from petting him is hesitation. The Latin root of hesitation is haesitationem, which means irresolution or uncertainty."[2] In psychology Doob [30] defines hesitation as *"the time elapsing between the external or internal stimulation of an organism and his/her internal or external response."* Taken together, when a human is uncertain about the situation, then one can observe a time laps in his/her behavior. Ineligible behavior causes uncertainty and doubt, because the interactor is not able to convincingly predict the future behavior. To this end, we suggest using hesitations as an indicator for illegible behavior, more precisely for unpredictable behavior. The other way around, using hesitation gestures to signal uncertainty is presented by Moon et al. [82]. In their human-robot interaction scenario, they used hesitation gestures performed by the robot to communicate that the robot is uncertain about the situation. Furthermore, the results of a human-robot path crossing pilot study evaluated by us and conducted by Dondrup and Hanheide [29] showing promise that it is possible to use variations in the velocity profile like sudden decelerations as measure for unpredictable behavior.

Another behavioral measurement we suggest is gaze behavior. We assume that the more often or the longer the robot is watched by a participant, the less legible and thus predictable is the robot's intention (see [74]). Therefore, we conclude that the focus is a possible implicit measure. Gaze behavior is used and discussed by Dautenhahn et al. [26] as a quantitative technique for analyzing robot-human interactions. No further research has been done so far regarding gaze behavior and legibility. We recommend investigating further the correlation between gaze and legibility and assume that it is a feasible measure for legibility.

Facial expressions are already used as feedback signals in HRI experiments [6, 71] and we recommend the use of facial expressions as a measurement for legibility. For example, one can identify a surprised expression as a negative feedback signal. However, before using facial expressions as feedback measure for legibility, we have to investigate which expression is correlated with illegible behavior. To this end, one can use a similar approach as Lang et al. [71] who identified negative and positive facial communicative signals as feedback measure for a human-robot teaching scenario.

To conclude, we suggest using hesitation signals, gaze behavior, and facial expressions as behavioral measurement methods for illegible robot behavior. However, at first all these measures have to be further investigated regarding their correlation with legible robot behavior before using them as evaluation method for human-robot interaction.

Inverse Wizard of Oz One experimental method we developed in order to capture the participants expectations is the "*Inverse Wizard of Oz* (I-WoZ)" method. The idea of the I-WoZ method is to let the participant steer the robot during an interactive task and capture the robot behavior. This approach is based on the assumption that the way someone steered the robot is not far from its own expectations in the situation. It is somehow similar to the previously proposed method of drawing the expected trajectory. However, referring to the timing of legibility measurements the "I-WoZ" method is like the other behavioral measures a real-time measuring method that allows us to measure during the

[2] http://www.vocabulary.com/dictionary/hesitation

interaction instead of before or afterwards. For a more detailed description we refer to chapter 5. on page 79 and to [76, 77].

3.5. Experimental Setups

In the following, we want to present and discuss experimental setups for the two major legibility factors predictability and expectation-fulfillment. The proposed setups are either derived from our own used setups or the work carried out by other researchers. We will focus on the timing factors described in section 3.3. on page 31 meaning we present setups to measure predictability before the behavior happened and expectation-fulfillment afterwards (see Figure Fig. 16 on page 31). We will give hints and examples for most of the previously proposed evaluation methods.

First of all, we want to point out that a legibility experiment can either be conducted as a video-based or a real life experiment. It offers the opportunity to evaluate different types of behavior in a within-subject experiment (e.g., [31, 73, 75, 106]) or to test the effect of one or more factors in a between-subject-design (e.g., [34]). For example, one can test different algorithms controlling the robot behavior, one can investigate if additional motions/gestures like gaze behavior would increases legibility [8], or if some other factors like an anticipated use of the robot at home [34], influences legibility.

Furthermore, we highly recommend using not only a single evaluation method but rather three or more methods together in one experiment. It has the advantage to get more accurate and reliable results (see [13]). To this end, we will point out how to combine different evaluation methods.

3.5.1. Predictability

Previously, we proposed different methods to measure predictability. For example, one can use a questionnaire, conduct an interview, use task performance metrics like measure the prediction or reaction time, or use behavioral measures like hesitation signals or gaze behavior. Furthermore, we discussed that timing is crucial in Section 3.3. on page 31. One have to measure predictability before the defined robot action is completely performed, but one have to ensure that the participants had enough time to imagine the future behavior. Furthermore, each of the proposed methods poses particular challenges regarding the setup. In the following, we present different experimental setups implementing the previously proposed methods to measure predictability of robot behavior.

3.5.1.1. Stop the Behavior

If we want to ask the participant to predict the future behavior of a robot by using a questionnaire, conducting an interview, measuring the prediction or the reaction time, one has to stop the interaction.

Depending on the used method the stop has to be predefined by the experimenter or triggered by the participant.

Here we want to draw attention to the fact, that for stopping-setups a video-based setting is easier to handle because it is much simpler to stop a video every time at the same point than stopping a robot during its activity in always the same way. Therefore, we do not suggest to use a real-live design in an interactive human-robot interaction, because it is not only hard to stop the robot during an interaction safely, one would also regularly interrupt the interaction, which is very irritating for the participant and can cause side effects. For example, one would stop the robot during a collaborative clearing the table task. It would not only irritate the participant, but also it might even be that the participant is currently not looking at the robot and do not recognize the stop.

Furthermore, in a real interaction, the robot behavior depends on the interaction partner and is very uncontrolled, due to interaction effects. Therefore, the behavior is not identical in each trial, thus not comparable. In short, we suggest using an observer-design where the participant is not involved in the interaction. Alternative, one could videotape the interaction using a head-mounted camera and provide a first-person perspective of the interaction (see also 3.5.3. on page 41).

Predefined Stop In order to ask the participant to predict the future robot behavior we have to stop the activity at a predefined point. The determination of the stopping point is based on the individual setting and the individual research question. For example, if the question is how predictable are different navigation algorithms in a human-robot path crossing setting one has to show the participant the navigation behavior of the robot up to a point before the paths of robot and human already cross. In our experiments, we used videos showing the participant the robot navigation behavior. One can use videos and show participants either the interaction from a third person view or one could videotape the interaction that has to be investigated using a head mounted camera and ,therefore, provide a first person view. After stopping the video one can ask the participant to predict the robot's behavior in the immediate future using a questionnaire (see Figure Fig. 17 on page 32), conducting a structured interview or one could ask the participant to draw the movement trajectory. Furthermore, one can measure the time the participant needs to predict the future behavior. An example of how to combine judgement and prediction time is presented above (see Figure 3.4.3.1. on page 33).

For detailed examples we refer to [31, 73, 75, 106].

Let the Participant Stop In order to measure the reaction time one have to show the participant the interaction and ask him/her to indicate once he/she feels certain to predict the future behavior. Like the prediction time method, this is also combinable with a questionnaire method and after the participant stops the robot one can ask the participant to predict the future behavior. A video-based example is presented by Dragan et al. [31]. They let a robot's arm move to one of two objects and ask the participant to press a button once they feel certain to predict the robot's goal. When the participant pressed the button the video stopped and the participants are asked to predict the robot's goal. Here one can also combine judgement and time measure (see 3.4.3.1. on page 33). For this setup, we strongly recommend using a video-based setup. Nevertheless, Basili et al. [8] let the participants observe a real robot moving to one of three goals and ask them to indicate once they feel certain to judge the robot's goal. However, they did not stop the robot. Like them, one could realize such a combination of reaction time and goal prediction by providing one button for each goal or by letting the participant say the predicted goal.

For detailed examples we refer to [8, 31]

3.5.1.2. Measure During Interaction

If one will evaluate a real-life interaction we suggest using behavioral or psychophysiological measurements, which did not interrupt the interaction. For example, if one will use hesitation signals one can videotape the interaction and manually annotate the video afterwards. Or, one can use a motion capturing system like Dondrup et al. [29]. They measured the participant's movements in a human-robot path crossing scenario and measured hesitation signals by identifying hesitation signals in the velocity profiles. In order to evaluate the gaze behavior of a participant one can use an eye tracking system to determine, for example, how often the participant looked at the robot in an interactive task like clearing the table together. For psychophysiological measurements, one can also use a variety of mobile devices like bracelets to measure heart rate and skin conductance during a human-robot interaction.

Moreover, the thinking allowed method is also applicable in real life interaction scenario. Here we also recommend videotaping the interaction and annotate later.

3.5.2. Expectation-Fulfillment and Surprise

Referring to section 3.3. on page 31, we measure the expectation-fulfillment factor after the robot behavior is accomplished. Therefore, a setup for expectation-fulfillment, when using questionnaires and interviews, psychophysiological measurements, is easy to build up and similar to classical HRI setups, where the participant are also asked afterwards. However, the "Inverse Wizard of Oz method" differs from all classical evaluation methods.

By using questionnaires or structured interviews, one can either show the participant the robot behavior via video or let the participant interact with the robot and ask the participant afterwards to judge if, or to rate how much, the actual robot behavior had fulfilled the participants expectations. Furthermore, we can combine predictability and expectation-fulfillment setups. After stopping the behavior and asking the participant to predict the robot's future behavior, one can start the robot's activity again and participant is able to observe the complete robot behavior. Afterwards, one can ask the participant regarding expectation-fulfillment and surprise and one gets data to all legibility factors. In this case, we highly recommend using video's. We refer to [75] for an example.

If one want to use psychophysiological measurements we suggest letting the participant interact with the real robot and capture psychophysiological data during the interaction and analyze the measures afterwards. Here, a combination of measures for predictability and expectation-fulfillment is very easy. For example, one can use an eye tracking system to measure the gaze behavior for predictability evaluations and simultaneously one can measure the pupil size to evaluate if the participant is surprised. We refer to [28] for an example.

The "Inverse Wizard of Oz" (iWoZ) method differs from the commonly used methods where one let the participant interact or observe the robot. By using iWoZ, the participant has to steer the robot and an instructed confederate interacts with the robot. To this end, one needs an easy to operate remote controller to steer the robot behavior. For example, if one wants to investigate robot navigation behavior one needs a remote controller to steer the robots movements (see [76]). We suggest restricting the functionalities as much as possible to not overcharge the participants, but one have to ensure that all possible robot behaviors are mapped. The development of an easy to handle controller is crucial for this type of evaluation method. In our study, where we wanted to investigate navigation in a human-robot path crossing we first tested different types of remote controller like a Joystick, a PS2 controller, and a marked keyboard regarding their usability. We highly recommend this pre-step of testing different types of remote control. Furthermore, to ensure that the participants are able to handle the remote controller we recommend training the participant before starting the experiment. Also different from common HRI setups is that the interaction is predefined, meaning that one have to instruct a confederate how to interact. The experimental conditions are defined by the different types of interaction the confederate is performing. For example, in a human-robot path crossing setup where one want to investigate the steered navigation behavior one can implement different conditions like different crossing angles or distances by instructing the confederate to cross the robot's path in the predefined conditions. By analyzing the steered robot behavior, you get results regarding the question which behavior is expected in the implemented setup. If one wants to analyze motions we recommend using video recordings and motion capturing system to capture the movements.

For an example of the "Inverse Wizard of Oz" (iWoZ) method we refer to [76, 77]

3.5.3. Video or Real-Live?

One question came frequently up during the planning and conducting experiments phase of this thesis: "Which is the better experimental method, video-based or real live interaction?" It is either possible to conduct a video based (e.g., in [31, 73, 75, 106]) or a real live experiment (e.g., in [8, 28]), in the following we discuss advantages and drawbacks of video and real live experiments:

One important advantage of video based experiment is that every participant sees exactly the same behavior. Therefore, video-based experiments are more controlled, more reliable and replicable. Robotic control methods are producing slightly different behavior. Sometimes, the environmental conditions are hard to control. These are the drawbacks of real live experiments. Most, one want to evaluate a prototype of an algorithm, or test only a design concept by using a wizard-of-oz technique.

Another advantage for video-based experiments are safety reasons or the possibility to conduct an online survey.

By using videos as well as in real live experiments it is possible that the participant is either an uninvolved observer, by using a third person view in the video and by letting the participant only observe the behavior, or an involved interactor, by providing a first person view (see [75]).

Woods et al. [114] convincingly argued that videotaped trials are a feasible approach for pilot studies. Instead of a video from a real robot one could also use a simulated robot, using ,for example, the MORSE Simulator [3] [33]. Takayama et al. [106] demonstrated that the use of simulated or animated robot behavior is a feasible way of testing interactive robot behavior. Also Dragan et al. [31] used a video based approach for their legibility experiments.

Nevertheless, the main drawback of a video-based method is that feelings like anxiety are less distinctive than in real life interaction. When using psychophysiological methods we suggest not to use video-based techniques. Furthermore, to evaluate an interaction like a hand-over or a cooperative task like a cooperative construction task one cannot use video. The same applies, when using behavioral methods. One has to interact directly with the robot to show signals like hesitation. Furthermore, video can never substitute all ranges of sensory information of a real interaction and a video screen is always limited.

To conclude, it is easier to conduct a video-based experiment and notably in respect of reliability and safety, video experiments have advantages. However, even when one will use psychophysiological measures or behavioral measures the real live experience is crucial. Finally, we can say that it is due to the type of interaction and the used evaluation method, whether to choose video or real-life methods.

[3] `http://www.openrobots.org/wiki/morse`

4. Measuring Legibility of Robot Navigation Behavior

After theoretical elaborations regarding legibility, we conducted several experiments and investigated legibility of robot navigation behavior in path crossing situations. In the following chapter we present the designs, results and conclusions of the experiments. We conducted experiments to evaluate state-of-the-art navigation algorithms as well as to assess which navigation behavior is the most legible behavior in a human-robot path crossing scenario.

4.1. Introduction

In the previous two chapters, we defined the term legibility and proposed approaches how to evaluate robot behavior regarding legibility. Thereby, we did not consider any special behavior. However, in the following chapter we will concentrate on legible robot navigation as previously mentioned in the introduction and investigate human-robot path crossing scenarios. To this end, we conducted different experiments in order to figure out what kind of navigation behavior is perceived as legible by humans. Additionally, we analyze correlations to other HRI properties like perceived safety and likeability.

We start our investigations with the evaluation of state-of-the-art navigation algorithms. Generally speaking, for robot navigation one find two popular kinds of concepts for moving the robot from one point to its goal. The older concept is to find a safe path for the robot and calculate motor commands to move the robot towards its goal [37, 39, 81]. Here, everything in the room (or outside) is treated as an obstacle whether it is a table or a human, the goal is to move the robot efficiently and safely, where safe means in this context that the robot did not collide with an obstacle, which could damage the robot or the obstacle. The second approach, which arises with the emerging field of Human-Robot-Interaction (HRI) is to take the human into account considering social and cultural constraints [58,68,78]. The objective of our first investigations is to compare two specific implementations of the previously mentioned concepts of current navigation algorithms regarding their legibility. To this end, we evaluated in our first two experiments a standard navigation approach, which does not take into account moving humans with a human-aware navigation approach, which takes into account moving humans, regarding legibility. After that, we investigated navigation behavior patterns in order to find which behavior is perceived as most legible.

Parts of the following chapter are published in [72, 73, 75].

Figure Fig. 19 on the next page shows a brief overview of the research questions, methods, and dependent measures of the conducted experiments.

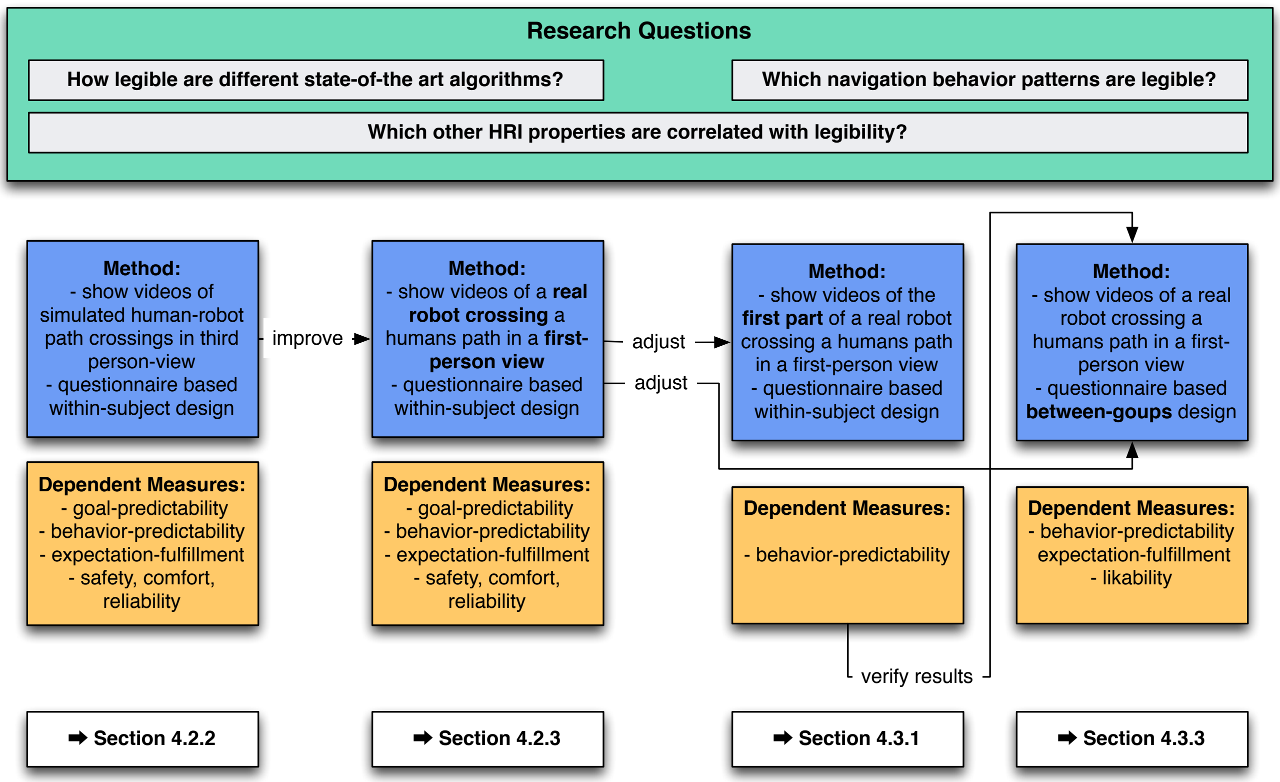


Fig. 19: Brief overview of the research questions, implemented experimental methods, and dependent measures of our cunducted experiments.

4.2. Evaluation of state-of-the-art Navigation Algorithms

We conducted two different experiments in order to examine legibility of different state-of-the-art navigation algorithms. One objective of the experiments was to figure out which common approach leads to high legible navigation behavior. Is it necessary to take the human interactor into account or it is sufficient enough to treat everything as an equal obstacle regardless of whether it is a table or a human? To this end, we compared the Human-Aware navigation method [66], which differentiates between obstacles like tables and humans, with a very common state-of-the-art navigation method Move-Base [81]. Additionally to our main objective we also want to investigate if legibility has an association with other HRI properties like safety, comfort, and reliability. Furthermore, we assess the association between the expectation-fulfillment factor and the HRI property surprise to support our assumption that both are highly correlated (see Section 3.2. on page 30).

We implemented two different variations of a within-subject questionnaire-based experimental design (see 3.5. on page 38). In both experiments we let a human and a robot cross each others path. The different navigation methods are producing different behaviors in a path crossing scenario, which we investigate regarding legibility, safety, comfort, and reliability in our experiments.

Previous work related to our research regarding legibility is done, for example, by Takayama et al. [106]. They also measured legibility[1] and other HRI properties like perceived intelligence and

[1] in their work called readability

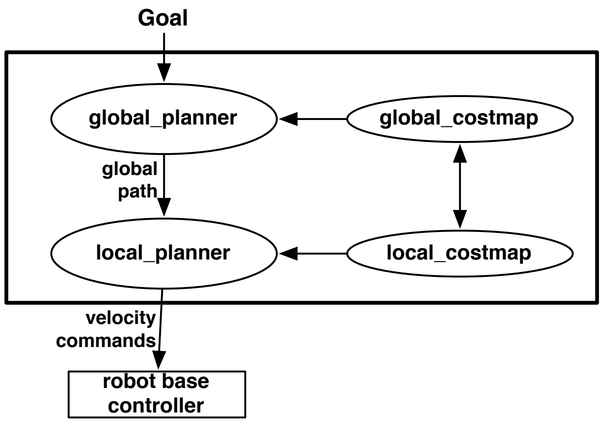


Fig. 20: Concept of navigation methods using a global and a local planner. (Picture based on www.ros.org/wiki/move_base)

safety of different robot behaviors. Opposite to our navigation experiment they showed the participant simulated videos of more complex robot behaviors like opening a door or delivering a drink. They found support for their hypothesis that a robot showing forethought, which should increase the legibility, increases the legibility and improves people's subjective ratings of the robot. Another experiment regarding legible robot behavior was conducted by Dehais et al. [28]. They conducted a real live experiment, where the robot hands over an object to the participant and measured legibility as well as safety and comfort. Similar to our experiment they compared a human-aware motion planner with a non human-aware planner. They found that the human-aware approach revealed higher legibility ratings as well as higher comfort and safety ratings. Based on the findings of Takayama et al. [106] and Dehais et al. [28] we expect to obtain a higher legibility for the human-aware approach and also to find correlations between legibility, safety, and comfort.

4.2.1. Navigation Methods

Before describing the experimental designs, we give a brief introduction into the used navigation methods in order to clarify the different approaches. In our experiments we compared two variations of the human-aware navigation (HA), designed by Kruse et al. [66], with two variations of the state-of-the-art navigation method Move Base[2]. All four navigation methods consist of a global and local planner (see Figure Fig. 20). The global planner uses an A* algorithm to generates the complete path to the goal. The local planner is seeded by the global plan and generates velocity commands to control the robot. Differences between the global planners result from the used cost function. The used local planners are differing in the method that determines the velocity commands. In the following, we describe the different global planner cost functions and the different local planners.

Move Base Global Planner The cost function of the move base global planner (MB) is based on a 3D voxel grid. An obstacle causes infinite costs with descending costs in its surrounding to propagate

[2]We used the Move Base implementation from the ROS navigation stack (www.ros.org/wiki/navigation)

them from obstacles out to a user-specified radius. For further information we refer to [81] and `http://www.ros.org/wiki/navfn`.

Human-Aware Global Planner The cost function of the human-aware global planner (HA) takes the human motions and all obstacles into account. In addition to the infinite costs of obstacles it increases the costs around a human and differentiates between a standing and a moving person:

- Moving: higher costs in front of the human to avoid moving in his/her direction of motion.
- Standing: higher costs behind the human to avoid moving behind his/her back.

Those "social costs" should ensure a comfortable navigation behavior. Further information can be found in [66].

Dynamic Window Approach The dynamic window approach (DWA) is a real-time collision avoidance strategy developed by Fox et al. [37] (`www.ros.org/wiki/dwa_local_planner`). The DWA computes local controls by first determining a target trajectory in position or velocity space (usually a circular arc or another simple curve), then inverting the robot's dynamics to find the desired velocity commands that produce that trajectory [39].

Waypoint Follower Local Planner The waypoint follower (WF) is designed to work with the human-aware global planner. This local planner projects its own motion and the human motion into the future and selects a speed that avoids predictable collisions. The human motion is predicted linearly assuming constant speed and direction while the robot motion is predicted using the global path returned by the human-aware global planner.

Trajectory Planner The trajectory planner (TP), developed by Gerkey et al. [39] (`www.ros.org/wiki/base_local_planner`) is based on different paradigm then DWA. Instead of searching for a feasible velocity commands to a trajectory the TP samples possible velocity commands and simulates the resulting trajectory. The TP algorithm chooses the best simulated trajectory by taking obstacle avoidance, goal distance and distance to the optimal path into account.

4.2.2. Measuring Legibility of state-of-the art Navigation Methods in a Human-Robot Path Crossing Scenario

In our first experiment we compared the four different navigation algorithms, the two variations of the Human-Aware navigation [66] with the two variations of the state-of-the-art navigation method Move-Base [81].

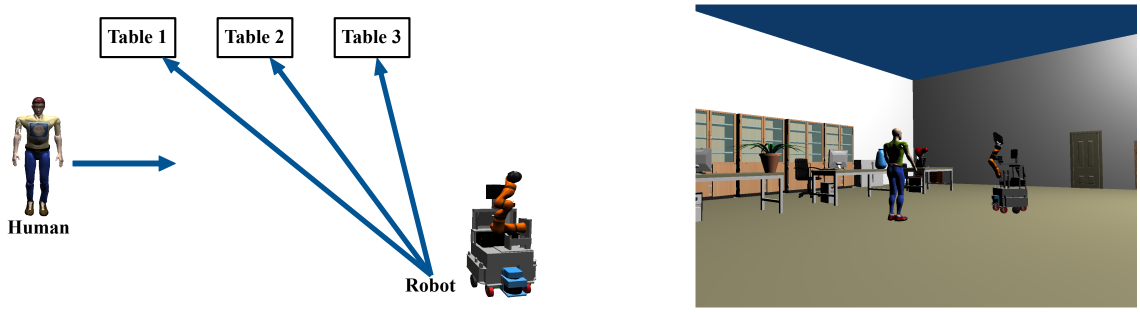


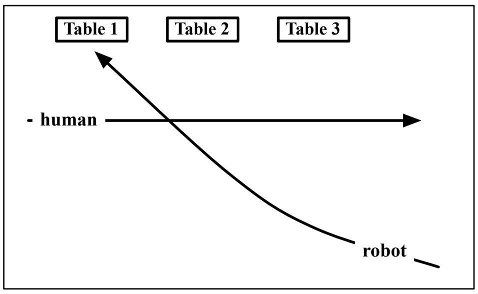
Fig. 21: Experimental design of the simulator based navigation experiment. The robot moves to one of the three tables and the human is crossing the robot's path.

Research Questions The questions we want to address with our experiment is how legible are different state-of-the-art navigation algorithms to a human observer? We assume that a human-aware navigation approach leads to higher legibility. Additionally to our main questions, we evaluated other HRI properties like safety, reliability, and comfort in order to answer the question if legibility is correlated with the HRI properties safety, reliability, and comfort ? Besides the other questions, we want to investigate if the expectation-fulfillment factor is highly and inverse correlated with surprise.

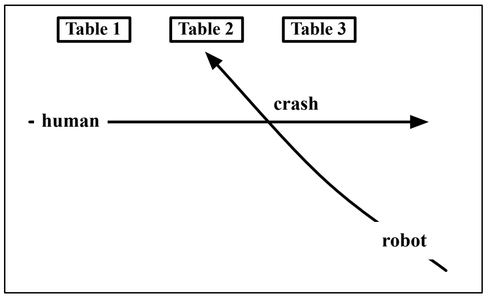
4.2.2.1. Experimental Method

We implemented a within-subject questionnaire-based experimental design (see 3.5. on page 38) where we showed the participants videos of a robot crossing the path of a human and measured legibility as well as the HRI properties safety, comfort, and reliability. Due to safety reasons, we used a simulator for the videos. Woods et al. [114] have convincingly argued that videotaped trials are a feasible approach for pilot studies like this one and also Takayama et al. [106] used a similar simulator based experimental design to determine the legibility.

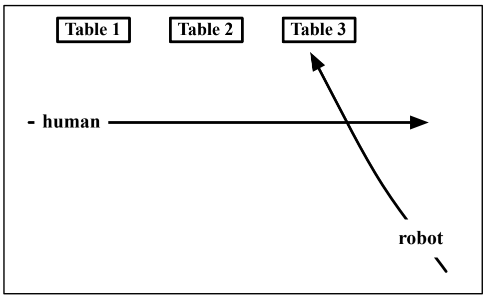
Design We recorded short movie sequences using the MORSE simulator [33] where a human is crossing the robot's path in an office environment (see Figure Fig. 29 on page 58). Our virtual environment shows three tables in an office scenario (see Figure Fig. 29 on page 58). In order to simplify the nomination of the tables we put three different objects with different colors (green plant, blue vase, red lamp) on each of the three tables. Our robot has to deliver a folder to one of them while a person is crossing its way as shown in Figure Fig. 29 on page 58. The robot and the human are starting from a fixed point in each of the clips. We provided a third person perspective in order to generate the feeling of a noninvolved observer. In order to avoid the potentially directing effect of the robot's gaze behavior we choose a robot without any artificial head (see Figure Fig. 29 on page 58). Basili et al. [8] has shown the influence of gaze behavior on predictability in a navigation experiment. Therefore, we eliminated the robot's "gaze" by removing all visible cameras and sensors to concentrate on the pure navigation behavior. We compared two variations of the Human-Aware navigation [66] with two variations of the state-of-the-art navigation method Move-Base [81] taken from the ROS navigation stack (www.ros.org/wiki/navigation). More precisely we used:



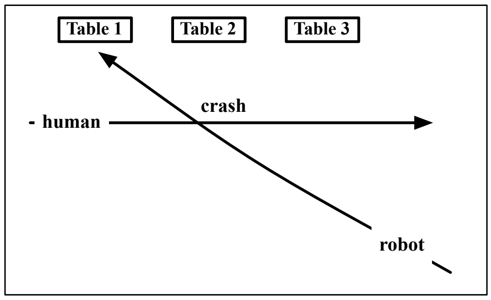
(a) MB-DWA/Table1: The human crosses the robots' path in front of the robot



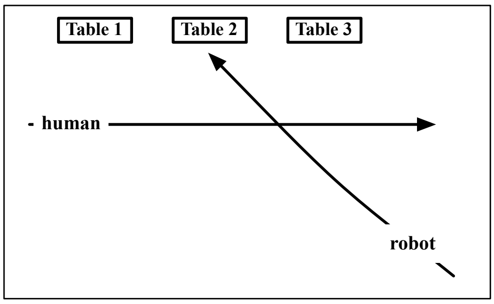
(b) MB-DWA/Table2: The human collides with the robot



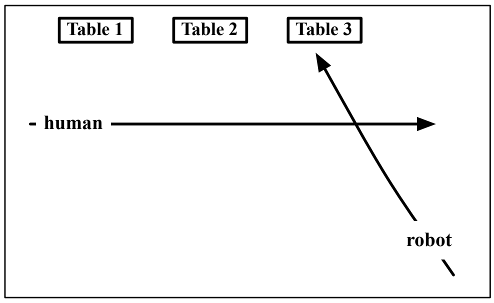
(c) MB-DWA/Table3: The human crosses the robots' path after the robot



(d) MB-TP/Table1: The human collides with the robot



(e) MB-TP/Table2: The human crosses the robots' path after the robot



(f) MB-TP/Table3: The human crosses the robots' path after the robot

Fig. 22: Overview of the MB conditions. The pictures are depicting the path of the robot and the human and the subcaption describes how they cross each others path.

- **MB-DWA:** The move base global planner with the dynamic window local planner.
- **MB-TP:** The move base global planner with the trajectory planner as local planner.
- **HA-WF:** The human-aware global planner with the waypoint follower local planner.
- **HA-TP:** The human-aware global planner with the trajectory planner as local planner.

The navigation behavior of each algorithm is depicted in Figure Fig. 22 and Figure Fig. 23 on page 54.

As depicted in Figure Fig. 22 and Figure Fig. 23 on page 54 in three of the 12 clips you can see the robot colliding with the human (see Figure Fig. 24 on page 55). The local planning algorithm causes the crash because it takes the moving human not into account (MB-DWA, MB-TP, HA-TP).

The video clips were divided into two parts. The first part of each video shows the robot behavior before the paths of human and robot cross up to the point where robot is only at a short distance to the human. The second part shows the robot behavior during the crossing event, like crossing before or behind the human, colliding with the human, spinning around, or waiting until the human pass. The second part ends when the robot reaches its goal.

Conditions With three tables as possible goal positions of the robot and four navigation methods, we tested $(3 \times 4) = 12$ different observation tasks.

Dependent Measures In order to quantify legibility we measured goal-predict-ability, by asking to predict the robot's goal, trajectory-predictability, by asking to predict the future direction of the robot, if the robot met the participants expectations (expectation-fulfillment) by using a yes/no scale, and how surprised the participant was by using a five-point Likert scale. In order to estimate the quality of the prediction, we also measured the confidence about the goal and direction judgement. Furthermore, we measured the HRI properties safety, comfort, and reliability also with a five-point Likert scale. For all ratings we used a five-point Likert scale on which '1' stands for not at all and '5' for very. The used questionnaires are to be found in the appendix (Fig. 50 on page 100)

Participants We recruited 16 participants with the average age of 26.6 years — thereof three women and 13 men. Eight participants have regular contact to robots, three from time to time and five rarely or no contact to robots. All participants gave their written consent to the experiment.

Procedure After the participants were welcomed and introduced we presented three videos to familiarize the participants with the virtual environment and the robot. Here we showed how the robot would move towards each table without the presence of a human. Afterwards, the experiment started and we showed the participants each two-part video clip once in random order. After the first part, we asked the participant to judge which table the robot is aiming at (goal-predictability) and if it will change its direction (trajectory-predictability). Additionally, participants were asked to rate their confidence on this judgment. Then we showed the second part and the participants were able to observe the actual behavior of the robot. Afterwards, the participants were asked to tell whether the robot's actual behavior was expected and if not to rate how surprising this was. Additionally we asked for ratings on the perceived safety, comfort, and reliability of the robot.

4.2.2.2. Results

We performed the data analysis in SPSS. Due to the non-parametric character of our data, differences of frequencies were analyzed with Pearson's Chi-Square tests and assessed ratings with a Friedman's ANOVA. For post hoc analysis of the gathered results, we applied Wilcoxon tests. Furthermore, a

Bonferroni correction was applied and so all effects for post hoc tests are reported at the 0.0167 level of significance.

Legibility Legibility was quantified using the measures goal-predictability, be-havior-predictability, confidence, expectation-fulfillment, and surprise. After the video stopped for the first time, we asked the participants to predict the goal the robot is heading to (goal-predictability) as well as the future direction (trajectory-predictability). Figure Fig. 25 on page 55 shows the results of correctly predictions. Most correct answers regarding the goal were given in scenarios where MB-DWA (83.3%) and HA-WF (79.2%) were applied compared to MB-TP (75.0%) and HA-TP (66.7%). Results of the correct predicted directions (trajectory-predictability) are showing a slightly different ranking regarding the navigation methods. The most correct answers regarding the direction were given in scenarios where the robot navigates with MB-DWA (68.8%) and MB-TP (64.6%) followed by HA-WF (60.4%) and HA-TP (43.8%). Pearson's Chi-Square tests did not reveal any significant association between type of planner and number of correct responses for goal-predictability $\chi^2 = 4.00$, $p = .261$ and marginally significant association between type of navigation and trajectory-predictability $\chi^2 = 7.17$, $p = 0.067$. After participants were able to watch the actual behavior of the robot they were asked if this matched their expectations. Results of this question are shown in Figure Fig. 25 on page 55. In 50.52% of the trials expectation was met. No difference in expectations was found between MB-DWA (64.6%) and HA-WF (64.6%) or MB-TP (58.3%), both $p > 0.05$. However, expectations when HA-TP (14.6%) was applied differed significantly from expectations during observation of HA-WF, $\chi^2 = 25.09$, $p < 0.001$, and MB-TP, $\chi^2 = 19.83$, $p < 0.001$. This indicates that the robots performance when navigated by HA-TP is really not legible. The average confidence rating ranged from $M_{HA-TP} = 3.57$ to $M_{MB-TP} = 3.85$, $\chi^2_F(3) = 2.98$, $p > 0.05$, indicating that all navigation methods equally allowed for confidence rating above average. For the 49.48% of trials in which the robot behavior did not meet the expectations, we asked participants to rate their surprise (see Figure Fig. 26 on page 55). The ratings differed significantly between the navigation methods, $\chi^2_F(3) = 8.91$, $p < .05$. However, after correction post hoc tests only reveal marginal difference between HA-WF ($Mdn = 3.00$) and HA-TP ($Mdn = 4.00$), $T = 13.00$, $p = 0.04$, and no significant difference between HA-WF and MB-DWA ($Mdn = 3.25$) or MB-TP ($Mdn = 4.00$), both $p > 0.0167$.

Safety, Comfort, Reliability Regarding the safety rating (see Figure Fig. 27 on page 56) there was a significant difference between navigation methods, $\chi^2_F(3) = 13.41$, $p < 0.01$. Post hoc tests reveal that safety was perceived higher with the HA-WF ($Mdn = 3.83$) compared to HA-TP ($Mdn = 3.33$), $T = 11.00$, $p < 0.01$, and also marginally higher compared to MB-TP ($Mdn = 3.33$), $T = 24.00$, $p = 0.02$. We found no significant difference regarding perceived safety between HA-WF and HA-DWA ($Mdn = 3.00$), $p > 0.0167$. Also the rating of comfort (see Figure Fig. 27 on page 56) showed a significant difference between navigation methods, $\chi^2_F(3) = 7.97$, $p < 0.05$. Here HA-WF ($Mdn = 3.33$) was perceived as more comfortable than HA-TP ($Mdn = 2.83$), $T = 9.50$, $p < 0.01$ and marginally more comfortable than MB-TP ($Mdn = 3.00$), $T = 14.50$, $p = .017$. Again no difference was found between HA-WF and HA-DWA ($Mdn = 2.83$), $p > 0.05$. The reliability of navigation methods (see Figure Fig. 27 on page 56) also differed significantly, $\chi^2_F(3) = 8.83$, $p < 0.05$. Post hoc Wilcoxon tests showed that HA-WF ($Mdn = 3.67$) was perceived more reliable than HA-TP ($Mdn = 2.67$), $T = 6.00$, $p < 0.01$. No difference in reliability was found for HA-WF

	goal-p.	traj-p.	expected	surprise	safe	comfort	reliable
goal-p.	1	$.157^{*}$	$.372^{**}$	$-.260^{**}$	-.082	.035	.001
traj.-p.	–	1	$.539^{**}$	$-.525^{**}$	$.243^{**}$	$.235^{**}$	$.277^{**}$
expected	–	–	1	$-.851^{**}$	$.300^{**}$	$.362^{**}$	$.407^{**}$
surprise	–	–	–	1	$-.323^{**}$	$-.395^{**}$	$-.457^{**}$
safe	–	–	–	–	1	$.898^{**}$	$.868^{**}$
comfort	–	–	–	–	–	1	$.869^{**}$

*. $p < 0.05$, **. $p < 0.001$

Table 4.1: Correlation coeffecients for all measurements of the experiment. Note, we calculated the point–biserial correlation coefficient, r_{pb} for the discrete dichotomy variables (goal-predictability, trajectory-predictability, expectation-fulfillment) and the standard Pearson's correlation coefficient r for all other variables.

compared to MB-DWA ($Mdn = 3.17$) nor MB-TP ($Mdn = 3.33$), both $p > 0.0167$.

Correlations In order to investigate the relationship between the measurements we calculated the Pearson's correlation coefficient r for all measurements. More precisely, we calculated the point–biserial correlation coefficient, r_{pb} for the discrete dichotomy variables (goal-predictability, trajectory-predictability, expectation-fulfillment) and the standard Pearson's correlation coefficient r for all other variables (see Table 4.1). One can see that all measurements are significantly correlated with each other, except the legibility factor goal-predictability with the HRI-properties safety, comfort and reliability. The correlation values for trajectory-predictability are considerably higher than for goal-predictability. Furthermore, we could prove our assumption from the previous chapter, that the expectation-fulfillment factor is highly and inverse correlated with surprise ratings.

Reliability of the used Questionnaire In order to evaluate the reliability of our questionnaire we calculated the Cronbach's α for the items safety, comfort, and reliability. It revealed that all had high reliabilities, all Cronbach's $\alpha = .957$.

4.2.2.3. Discussion

With the experiment at hand, we evaluated four different robot navigation methods in a simulated environment. A human observer had to watch video clips from a third-person perspective, in which a robot was moving to an unknown target, while a human was crossing its path.

Legibility The video was stopped before the human and the robot crossed paths and the observer had to predict the future goal and direction of the robot. Results of comparing legibility in the form of correct goal prediction and correct behavior prediction between navigation methods did not show any significant differences. This is also mirrored in the reported level of confidence. Participants were equally confident in rating all navigation methods which means that all ratings and judgments on the different navigation methods were performed with comparable quality. Nevertheless, when assessing the numerical differences between numbers of correct prediction (goal-predictability) we found that the legibility of the robots behavior was worse for TP-local planners — independent of the applied global planners. While HA-WF and MB-DWA resulted in more correct predictions no striking differences between them were observed. However, the numerical differences between correct direction predictions (trajectory-predictability) were more distinctive and almost significant. Here the classical navigation algorithms MB-DWA and MB-TP received the best results. Both algorithms treat the human as a static obstacle and do not change the robots behavior in terms of directional or velocity changes like HA-WF and HA-TP. The second judgment for accessing legibility was requested after participants had watched the video to the end and were thus able to see the whole performance of the robot. Expectations regarding the robot's behavior was met in over 60% of the trials applying HA-WF and MB-DWA respectively. Also during navigation by MB-TP expectations were met in more than half of the trials. However it is striking how few trials, namely less than 20%, were met during application of the HA-TP algorithm. One possible explanation for this finding is that if one decides to apply human-aware navigation in the global context, it is also essential to apply human-aware local planning strategies. It almost seems as if the global human-aware planner is useless regarding predictability of behavior when it is combined with a non human-aware local planner. This also goes in line with better results for MB-TP. Here both global and local planner have no human-aware navigation strategies — which might indeed be a strategy that is attributed to a robot being a machine. Predictions and expectations regarding a robot behavior might be guided by that impression and higher ratings for HA-WF and MB-DWA might be a hint towards the fact that human-aware strategies (HA-WF) or other dynamic approaches (MB-DWA) result in a higher legibility. Taken together, we could not support our assumption that a human-aware approach leads to a higher legibility, as it was shown by Dehais et al. [28].

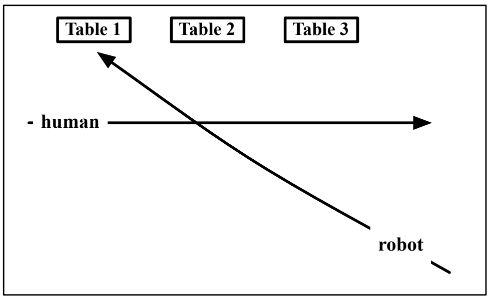
In order to gain information from the trials in which expectations were not met, we asked participants also to rate their surprise about the actual behavior of the robot. Proofing the results from the analysis of the expectation-fulfillment factor, we found that the robot's behavior during HA-TP caused highest surprise ratings. HA-WF caused lowest surprise rates and was also not significantly different from MB-DWA, the latter resulted in numerically higher ratings. One main difference between HA-WF and MB-DWA is that MB-DWA does include the possibility of crashes with the human because the navigation method treats the human as a moving obstacle. One of the MB-DWA videos showed this possibility, which potentially caused higher surprise rates. The same reason can of course also add to the lower legibility of the TP- local planner and cause higher surprise ratings during observation of MB-TP and HA-TP. In other words, this means that the navigation with HA-WF was least surprising and thus best legible because crashes were never possible.

Safety, Comfort, Reliability Additionally, we asked the participants to rate the perceived safety of the robot, the comfort they would feel when interacting with it and the perceived reliability of its behavior. All three measures were rated best for HA-WF with the most striking difference to HA-TP. This shows again that both local and global planner should be human-aware by showing that navigation when combining a human-aware global planner with a non human-aware local planner is not perceived as safe in behavior nor reliable and does not result in a comfortable interaction. Analyzing safety and comfort ratings, we also found that HA-WF was rated safer and more comfortable compared to MB-TP. Regarding perceived safety one has to keep in mind that HA-WF doesn't allow for crashes at all which is most probably the reason for higher ratings. Regarding comfort it is assumed that people feel more comfortable when the responding system acts in a familiar way. The human-aware navigation includes human collision avoidance strategies in form of "social costs" [66] which might cause higher comfort ratings compared to merely technical collision avoidance. Nevertheless, no difference in reliability was observed between HA-WF and MB-TP or MB-DWA. Nowadays people are used to interact with technical systems on a daily basis and we experience that those systems gain robustness with exponential velocity. It is thus not surprising that also systems following a more technical behavior are perceived as reliable.

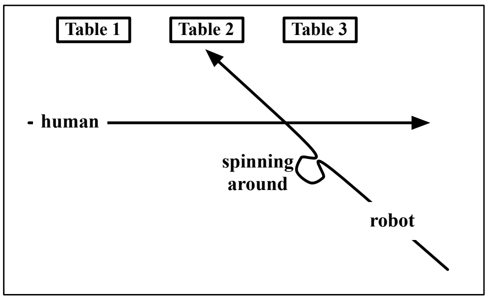
Correlations We found highly significant correlations for all measurements except for goal-predictability and the HRI properties safety, comfort, and reliability. As assumed the legibility factors goal-predictability, trajectory-predictability are significantly correlated with expectation-fulfillment, and surprise , but with considerably higher correlation coefficients for trajectory-predictability. This fact let us assume that the trajectory-predictability, which is furthermore also highly correlated with the HRI properties safety, comfort, and reliability, is a more important property. It is obvious that for safety reasons a human wants to know in which direction the robot will drive instead of knowing its goals, which in turn influences the other properties reliability and comfort. Our findings regarding the correlated properties are in line with results from Takayama et al. [106] and support our overall hypothesis that legibility is an important factor for HRI. One not really new finding is the correlation of the HRI properties safety, comfort, and reliability, which are not only highly significantly correlated, but also with strikingly high correlation values, which shows the connection between these properties.

Limitations A small number of participants (16) can be the reason for the non-significant results regarding the legibility factors goal- and trajectory-predictability. Another Limitation is the third-person view which is helpful in terms of goal predictions, but it can make it difficult to judge directions. This fact is mirrored by the lower number of correct behavior-predictions in contrast to the number of correct goal-predictions.

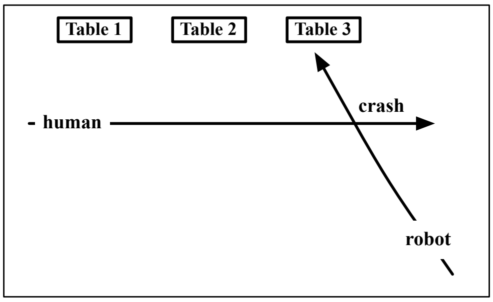
Conclusion Taken together, HA-WF had a higher legibility compared to both approaches with TP local planners and a comparable legibility with the state-of-the-art non human-aware planner MB-DWA. Regarding the HRI properties safety, reliability and comfort, HA-WF resulted in higher perceived safety and increased feeling of comfort compared to HA-TP and MB-TP as well as a higher attribution of reliability compared to MB-TP. In summary, we can ,therefore, assume that a high legibility measured by correct goal- and trajectory-predictions, expectation-fulfillment and surprise leads to high values regarding perceived safety, felt comfort and attributed reliability during the interaction with a robot. Both HA-WF and MB-DWA allow for a high legibility. However, if it comes to real world applications one should consider that only HA-WF does not allow collisions while navigating in the close surrounding of humans.



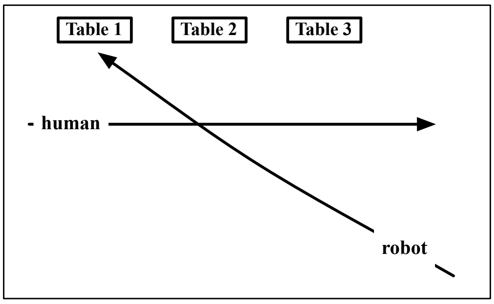
(a) HA-TP/Table1: The human crosses the robots' path in front of the robot



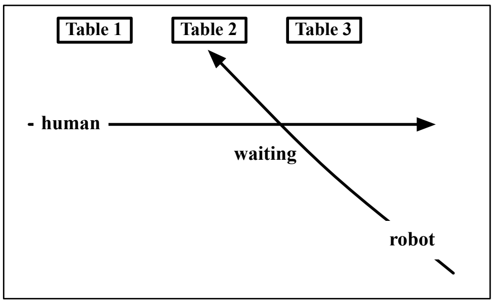
(b) HA-TP/Table2: The human crosses the robots' path in front of the robot, while the robot is spinning around



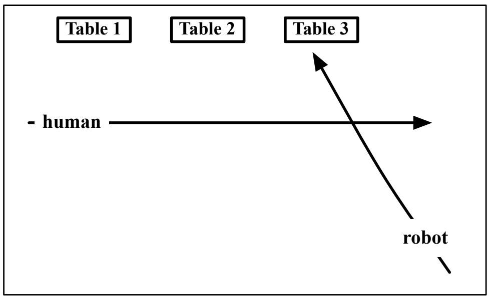
(c) HA-TP/Table3: The human collides with the robot



(d) HA-WF/Table1: The human crosses the robots' path in front of the robot



(e) HA-WF/Table2: The human crosses the robots' path in front of the robot, while the robot is waiting



(f) HA-WF/Table3: The human crosses the robots' path after the robot

Fig. 23: Overview of the HA conditions. The pictures are depicting the path of the robot and the human and the subcaption describes how they cross each others path.

Fig. 24: Example of a crash with the human as shown in the video clip. We added a "BANG!" logo to the video in order to make clear that the robot is colliding with the human.

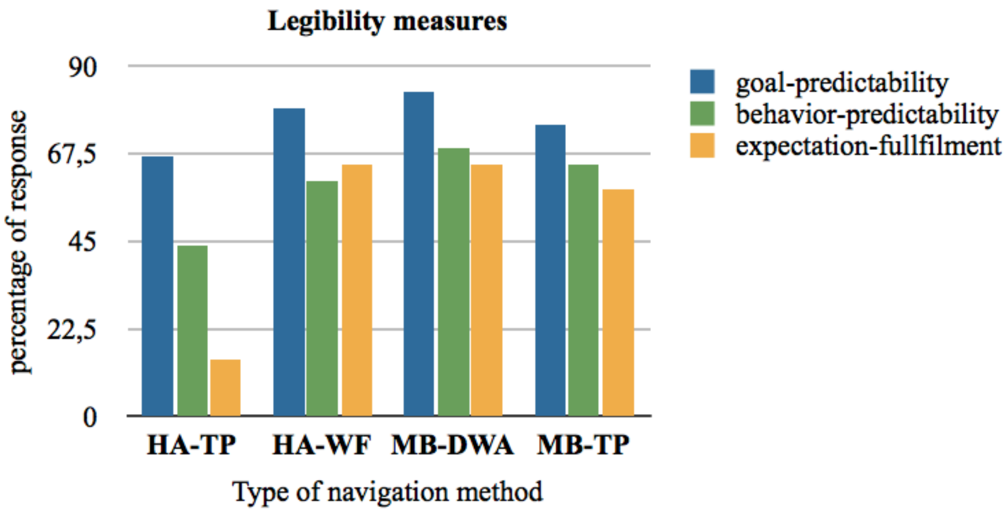


Fig. 25: Frequency of correct predictions for goal- and trajectory-predictability and how often the robot behavior meets the participants expectations

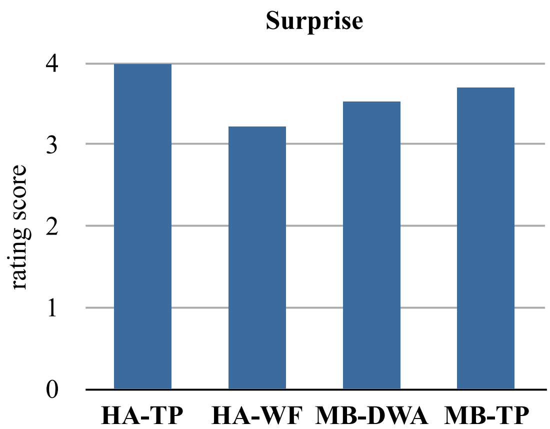


Fig. 26: Mean of the surprise property rated on a five-point Likert scale.

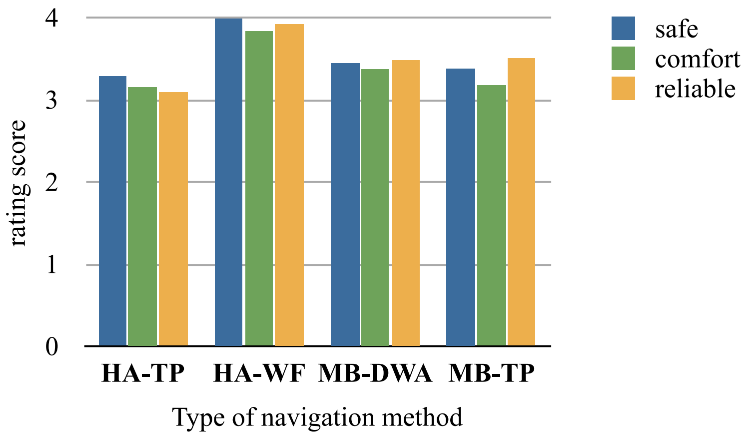


Fig. 27: Mean values for safety, comfort and reliability rated on a five-point Likert scale

4.2.3. Influence of Legibility on Perceived Safety

With a follow-up experiment, we wanted further to investigate how legible the different navigation approaches are perceived by an interacting person. Furthermore, we wanted to improve our experimental design. To this end, we eliminated the former limitation of the third-person view and designed a similar experiment, where we provided a first-person view. Furthermore, we switched from simulated robot behavior to a real-world scenario with a real robot and a real human crossing the robot's path. However, due to reliability reasons (see section 3.5.3. on page 41) we videotaped the behavior.

Additionally, we further investigate the association between the legibility-factor trajectory-predictability and perceived safety. In order to measure the perceived safety in a way that is more comparable to the work of other researchers in the HRI community we used the common Godspeed questionnaire [7] for perceived safety instead of the formerly used own designed questionnaire. We assume to support our former findings of a correlation between legibility and perceived safety.

Research Question With the experiment at hand, we want to address the question of (1) how legible different navigation algorithms are to a human during human-robot path crossing and (2) how legibility increases the perceived safety during human-robot path crossing.

4.2.3.1. Experimental Method

We implemented a within-subject questionnaire based experiment where we showed the participants videos in first-person perspective of a robot crossing the path and used a similar questionnaire-based method as in the previous experiment to measure legibility and perceived safety. Due to safety and practical reasons we used a video based design again (see 3.5.3. on page 41), but this time we videotaped a real interaction.

Technical Equippment The platform used in this experiment is the omnidirectional RWI B21 robot with a four wheel synchronous drive (see Figure Fig. 28(a) on the next page). The robot has a height of 1.22 m and a footprint of 0.53 m in diameter. For localization and navigation it has a SICK laser scanner. For accessing the position and direction of the human, we utilized a fixed Kinect sensor, which was installed in the lab environment.

Basili et al. [8] showed that the gaze has a significant influence on participants' ability to infer the direction of an approaching person or robot. However, in the study at hand we decided to avoid the potentially directing effect of the robot's gaze in order to find out whether and to what extend the behavior of an autonomous agent is legible if the only available cue is its motion. Therefore, we chose the B21 robot without any artificial head or eyes, allowing us to concentrate on the legibility of the robot's navigation behavior. To provide first person view videos we used a head-mounted goPro camera (see Figure Fig. 28(b) on the following page).

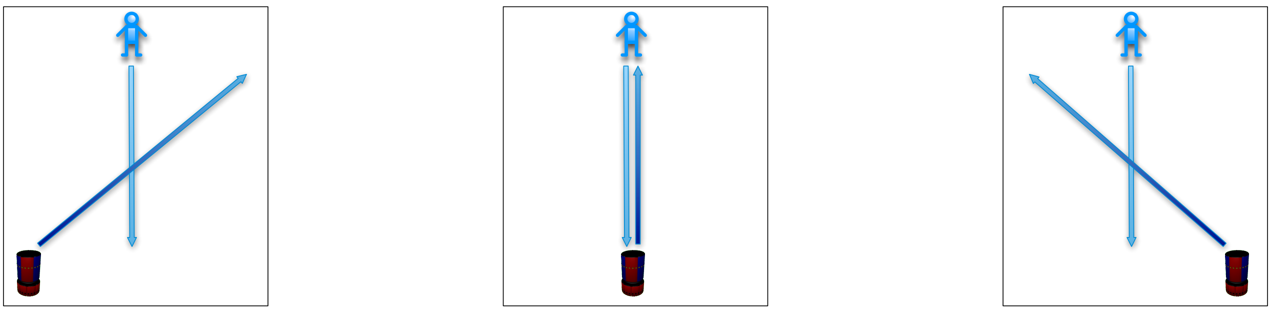
(a) B21 RWI robot

(b) Confederate wearing a head-mounted GoPro camera

(c) Picture of one of the video clips providing a first person view

Fig. 28: Supplementary material and environment for video recording: (a) B21 RWI robot, (b) confederate with head mounted camera, (c) lab area with Kinect sensor as seen from the head mounted camera.

Design For the preparation of the experiment we recorded short video sequences in which a robot is crossing a person's path in our lab environment, see Figure Fig. 29. We used a head mounted goPro camera to provide videos in first person perspective. In order to guarantee an unbiased interaction, the confederate for video recording was a person not professionally involved in developing navigation algorithms. In all video clips the human interactor starts from one defined position while the robot started from three different positions, see Fig Fig. 29. The interactor starts walking when the robot starts moving.



(a)

(b)

(c)

Fig. 29: Experimental design. The pictures are showing the approach directions of the robot (a) from the left, (b) frontal and (c) from the right.

For the experiment we divided the video into two parts. In the beginning of the video clips both navigation methods (MB-DWA and HA-WF) produce the same robot navigation behavior. The video was cut at the point where the navigation methods are starting to differ in their behavior. When the the robot perceived the human either as an obstacle (MB-DWA) or as a human (HA-WF) the navigation method starts reacting. Therefore it is not necessary to keep the speed of the interactor constant.

Nevertheless the interactor attempted to walk with a constant speed.

For our study, we used the following two navigation methods: MB-DWA is the move base global planner with the dynamic window local planner; HA-WF is the human-aware global planner with the waypoint follower local planner. We compare the navigation method MB-DWA with the human-aware navigation method HA-WF, to find out which concept is the best regarding legibility and perceived safety. In the frontal approaching videos (see Figure Fig. 29 on the preceding page (b)) the robot controlled by MB-DWA first stops and then moved to the left side. The robot controlled by HA-WF moved smoothly to the right side. In the right and left side approach (see Figure Fig. 29 on the facing page (a) and (c)) the robot controlled by MB-DWA crosses the path in front of the human. The robot controlled by HA-WF stops and lets the human pass its way.

Conditions With three approach directions of the robot and two navigation methods, we tested $3 \times 2 = 6$ different observation tasks. Each observation task was displayed as a video clip once per participant in random order. To avoid learning effects, each video was only displayed once per participant.

Dependent Measures In order to quantify legibility we measured several legibility factors (see Section 3.2. on page 30). We ask the participants to predict the navigation behavior of the robot in the immediate future (right, left, ahead, back, stop) to measure trajectory-predictability. We quantify the velocity-predictability, by asking if the robot will decelerate, accelerate, stop or maintain its velocity. We used a five-point Likert scale to determine how much the robot behavior meets the participants expectations. Furthermore, we measured the perceived safety by using the Godspeed-V questionnaire [7] in which the participant has to rate his/her emotional state regarding anxiety, agitation, and surprise on a five point Likert scale (5 = very, 1 = not at all) (see also [69]). The used questionnaires are depicted in the Appendix Fig. 51 on page 101.

Participants We recruited 18 participants with the average age of 28 years — thereof six women and 12 men. One participant has regular contact with robots, two from time to time and 15 have rarely or no contact with robots.

Procedure First, participants watched four videos to familiarize with the environment and the robot. These introductory videos showed examples of how the robot would move through the room from every direction without the presence of a human and one video of a human walking through the room without the robot. They were presented in a third-person perspective to allow participants to get a feeling for the room. Afterwards, the experiment started. We measured legibility in each video at two points. First we showed the first part of the video and asked the participants to predict the behavior of the robot in the immediate future. Here two categories were provided: velocity (faster, slower, equal, stop) and direction (to one side — right or left, straight forward, stop). After participants had marked their prediction, the video was started again and participants were able to observe the complete behavior of the robot. After the robot leaves the camera field of vision, the video ended. Participants were asked to rate whether the robot's actual behavior met their expectations on a 5 point Likert scale. Additionally we asked for their perceived safety utilizing the Godspeed-V questionnaire [7]. After the experiment, we debriefed the participants verbally. We asked them to describe their impressions and whether they believed that the robot could collide with a person and if they would be afraid of the robot.

4.2.3.2. Results

Legibility In order to assess the level of legibility of the two navigation methods, participants were asked to predict the robot's behavior at the video break. After watching videos in which the navigation method was HA-WF 29.6%, after videos with MB-DWA 22.2% of predictions were correct. Answers were only considered as correct if both direction and velocity were judged correctly by the participant. However, Pearson's Chi-Square test did not reveal a significant association between navigation method and number of correct predictions, $p > .3$. After watching the video to its end, participants had to rate if the robot's behavior matched their expectations. Here we also measured the level of legibility. These ratings were analyzed with a 2×3 repeated measures ANOVA with the within subject factors *navigation method* (HA-WF, MB-DWA) and *approach direction* (left, frontal, right). No significant difference was observed between the navigation methods HA-WF (Mean: $M = 2.69$) and MB-DWA ($M = 2.59$), $p > .7$. However we found a significant main effect for the factor approach direction, $F(2, 34) = 3.56$, $p < .05$. Contrasts show that expectations were met to a higher extend when the robot was approaching the person from the front ($M = 3.08$) than when the robot was approaching the human from the left side ($M = 2.17$), $F(1, 17) = 6.93$, $p < .05$. No difference was observed between frontal approach or approach from the left and approach from the right ($M = 2.67$), both $p > .09$.

Perceived Safety For assessing the perceived safety, we asked participants to rate their emotional state regarding the three items anxious, agitated and surprised after watching the videos to their end. According to Bartneck et al. [7], the combination of these measures assesses the perceived safety of a person during interaction with a robot. Note that due to the arrangement of the scale in the Godspeed V questionnaire, a lower rating score stands for a higher perceived safety.

In order to verify possible differences, a $3 \times 2 \times 3$ repeated measures ANOVA with the within subject factors *item* (anxious, agitated, surprised), *navigation method* (HA-WF, MB-DWA) and *approach direction* (left, frontal, right) was calculated. It was found that items differ significantly from each other, $F(2, 34) = 17.43$, $p < .001$. Contrasts show that the surprise was on average rated higher ($M = 2.39$) compared to the agitation ($M = 1.90$), $F(1, 17) = 17.81$, $p < 01$, and the anxiety ($M = 1.68$), $F(1, 17) = 33.41$, $p < .001$, with the latter not being different from each other, $p > .1$.

Ratings for navigation method were significantly lower for HA-WF ($M = 1.77$) compared to MB-DWA ($M = 2.21$), $F(1, 17) = 6.58$, $p < .05$. As this result is a conglomeration of all measured items, it also provides an impression on how safety was perceived after watching the videos with a different navigation methods. Note again, that due to the arrangement of the scale in the Godspeed

V questionnaire, a lower value indicates a higher perceived safety. Thus, perceived safety was higher for HA-WF than for MB-DWA, see also Figure Fig. 30. Regarding approach direction, no significant differences were observed, $p > .4$. Also no interaction effect was significant, all $p > .2$.

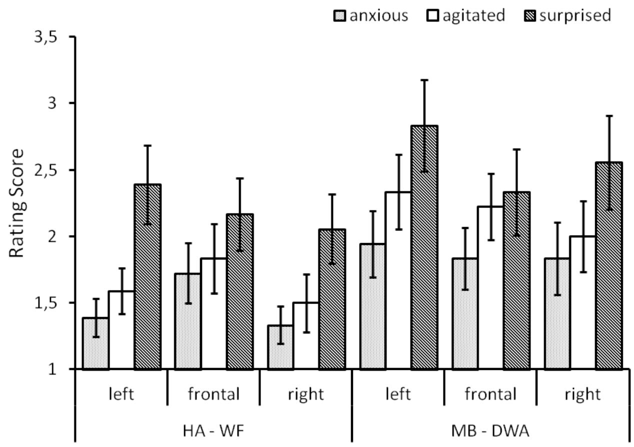


Fig. 30: Godspeed-V-scores of the different items of perceived safety. The perceived safety is higher after watching HA-WF videos. Note: due to the scale in the questionnaire a lower value means higher perceived safety.

Correlation Between Legibility and Perceived Safety In order to find out if legibility (measured as correct or incorrect responses) caused a higher perceived safety we performed a correlation analysis, see Figure Fig. 31. For this purpose, we calculated the perceived safety as the average score of the three items on a trial basis. The analysis revealed that there was a significant relationship between the correct answer and perceived safety, $r_{pb} = -.22$, $p < .05$.

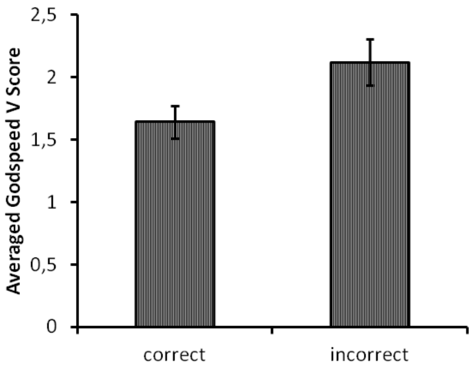


Fig. 31: Relation between correct/incorrect answer and perceived safety. If the answer was incorrect, the average rating score increased which is equivalent with a lower perceived safety.

In the next step, we also correlated the expectation-fulfillment ratings with the value obtained for perceived safety. Also here we found a significant relation, $r_s = -.29$, $p < .01$, showing that the more expectations were met, the higher the perceived safety was, which is expressed in lower Godspeed-V-scores, see Figure Fig. 32 on the next page.

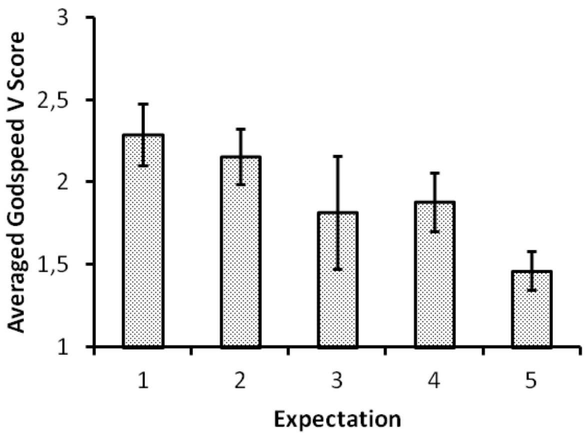


Fig. 32: Relation between perceived safety and expectation score. If expectations are fully met (score = 5), the Godspeed-V-score is low, which expresses higher perceived safety.

Debriefing After the experiment participants were debriefed verbally. No participant was afraid of the robot. Almost all (excluding the one who has regular contact to robots) believed that a robot generally has a built-in safety system which makes sure that the robot never collides with or hurts a human. Most participants (10 of 18) believed that the robot would register them as a human being. Here a frequent comment was: "The robot sees me." To sum up, we can say that participants had a basic trust in the robot's safety.

4.2.3.3. Discussion

Legibility and Perceived Safety In our experiment, legibility was measured at two different times. Both measures of legibility were analyzed independently of each other. Results show that the HA-WF navigation led to a numerically slightly higher number of correct predictions. Nevertheless, there was no significant difference to the responses with MB-DWA navigation, which is not surprising as videos were stopped at that point in time when the algorithm has detected human presence and before it was able to respond accordingly (see Section 4.2.3.1. on page 58). Also, no significant differences were observed regarding the legibility of navigation methods after the video was stopped and people had to rate to what extend the robot's behavior matched their expectations. Nevertheless, we found that expectations were met to a higher extent when the robot was approaching the human frontally. By looking at the results from the debriefing, we conclude an explanation for this fact. It becomes clear that people expect the robot to recognize them. Furthermore, they have a general trust in the fact that the robot would never hurt them, i.e., collide with them. Thus, when the robot is approaching them frontally they expect the robot to react in any way in order not to collide. In this situation, most participants predicted that the robot would stop, which was correct for HA-WF. Many participants predicted that the robot would slow down and turn sidewards, which was not correct for any navigation method.

In general, the number of correct predictions was pretty low. This is also mirrored in the rating of expectations. On average ratings for the two navigation methods are below three, which means that expectations were rather not met with both methods and both resulted in a robot behavior that was not perfectly legible. However, as no differences were observed regarding the legibility of the navigation methods at both judgment times, it is reasonable to combine the data of the navigation methods for correlation analysis while still keeping the two measures independent. This allows for a general remark on the relation between legibility and perceived safety. Regarding the latter, we found a significant correlation between legibility and perceived safety for both judgment times. The average Godspeed-V-score obtained after correct responses was lower than after incorrect responses, which means that the perceived safety was higher when people were able to predict the robot's behavior correctly. Regarding the rating of the expectation-fulfillment factor it was found that the more the expectations were met, the lower the average Godspeed-V-score was and thus, the higher the perceived safety. Therefore, we can assume that the legibility of the robot's behavior is clearly one prerequisite for perceived safety. If the robot's behavior is legible, a person will feel safer during interaction – also if interaction only means crossing paths.

Perceived Safety of Navigation Methods Although, we observed no difference regarding legibility between navigation methods, we found differences regarding the perceived safety. If the robot navigated with HA-WF, a lower Godspeed-V-score was obtained compared to the score obtained after navigation with MB-DWA. Although, the human-awareness is not the only difference between the studied navigation algorithms. Our results provide the first hint towards the possibility that if a human is included into the robot's navigation strategy, like the navigation method HA-WF does, the perceived safety on the human side is higher. It is worth mentioning that the perceived safety was in general pretty high (expressed in a low Godspeed-V-score, see Figure Fig. 30 on page 61). This can first be caused by the experimental situation in which people were only watching a video where the robot is interacting with a person: there was never a real danger of collision with the participant her-/himself. Furthermore, the debriefing showed that people have a general trust in the safety mechanisms of the robot. However, if this basic trust would remain during real world interaction is subject to further investigation.

Limitations One limitation of this experiment is the use of videos instead of confronting the participants with real robot navigation behavior. We discussed this issue in detail in the previous chapter (see 3.5.3. on page 41). Another point is the training effect caused by the repeated measures design. Every participant sees every situation twice and we assume that the first seen behavior influences further judgements..

Conclusion Legibility was rather low in both navigation methods with a slightly higher accuracy for HA-WF. To this end, we can conclude that both methods do not generate legible navigation behavior in a path crossing scenario. As a consequence more research is necessary in order to determine which navigation behavior is legible in a path crossing scenario. However, a significant higher perceived safety was measured for HA-WF compared to MB-DWA. However, we proved our hypothesis that legibility is correlated with perceived safety of the robot's behavior. We found that the higher the legibility the higher the perceived safety.

4.3. Identify Legible Navigation Behavior in a Path Crossing Scenario

Based on our previous finding, that both previously tested state-of-the-art navigation methods scored poorly regarding legibility, we shift our attention from testing state-of-the-art methods towards the question that navigation behavior is perceived as highly legible in a path crossing scenario. Researchers like Kruse et al. [64] or Guzzi et al. [45] assume that human-like behavior is perceived as highly legible. To this end, they implemented a human-like navigation behavior. Again, Sisbot et al. [99, 101] assume that the motion has to be as visible as possible to be legible. They all follow assumptions, but only barely evaluate if their implemented methods are generating legible behavior in terms of predictability and expectation-fulfillment. The human-aware motion planner was successfully evaluated by Dehais et al. [28] and Dragan et al. [31, 32] also evaluated their arm-motion approach, but both are motion planning methods for hand-over and reaching motions for a robotic arm and not navigation methods. Therefore, we concentrate on navigation behavior. With the work at hand, we want to find which behavior is perceived as legible in a human-robot path crossing scenario. To this end, we generated different kinds of behavior patterns using the wizard-of-oz technique and evaluated them in a video-based experiment. We used a similar human-robot path crossing setup as in our previous experiment. The experimental design is comparable to our former designs and also to Dragan et al. [31] and Takayama et al. [106]. We show the participant's videos of different human-robot path crossing scenarios and use a questionnaire to determine legible robot behavior. Hüttenrauch et al. [54] concluded that Hall's personal distance [48] was mostly preferred by participants in a HRI spatial user trial. Therefore, we investigated distances regarding Hall's personal space.

Based on the findings of a human-human path crossing experiment conducted by Basili et al. [9] we expect that a straight motion with a collision avoiding strategy by manipulating the velocity (slowing down, stop) would be the most legible navigation behavior.

Research Question We designed two coherent experiments with the objective (1) to identify the navigation behavior that is most legible in terms of trajectory-predictability and expectation-fulfillment in the first experiment, and (2) to evaluate the findings of the first experiment regarding legibility in a follow-up experiment. Additionally, we investigate the (3) association between legibility and the HRI property likability.

4.3.1. Experiment to Identify Legible Navigation Behavior

The objective of the first experiment is to identify legible navigation behavior in a human-robot path crossing scenario. Moreover, we want to investigate if the distance influences the expectations. To this end, we show participants videos in the first person perspective where a robot is crossing the persons path and stopped the videos at different distance points.

4.3.1.1. Experimental Method

We implemented a within-subject questionnaire-based experiment where we showed participants videos in a first-person perspective of a robot crossing the person's path. We asked the participants to predict the future robot behavior in terms of direction and velocity in order to determine the most legible navigation behavior. To avoid any biasing effect, which we previously mentioned as a limitation (see 4.2.3.3. on page 63), we showed the participants only the robot behavior before the crossing occurs. We tested three different distances to the crossing point for the robot (40 cm, 90 cm, 130 cm), the mid-distance 90 cm is within Hall's personal space [48] (45 cm to 120 cm) and the two outside-distances, 40 cm and 130 cm, are slightly behind and before the personal space. 40 cm is within the intimate space and 130 cm is within the social space. The used questionnaire is presented in the Appendix (see Figure Fig. 53 on page 103).

Design For the experiment at hand, we used the same technical equipment (B21 Robot and a head-mounted camera) as in the previous experiment (see 4.2.3.1. on page 57). We also used a similar path-crossing setup with three different approaching angles (right $45°$, frontal, left $45°$) as in the previous experiment. In order to evaluate the influence of angle and distance we used, additionally to the different approaching angles, three different stopping distances for the robot (40 cm, 90 cm, 130 cm) and two different stopping distances for the confederate (70 cm, 120 cm), see Figure Fig. 33 on the following page. Due to our setup we recorded 18 video clips in which our robot is crossing the confederate's path in our lab environment (see Figure Fig. 34 on the next page). In order to guarantee an unbiased interaction, the confederate for video recording was a person not professionally involved in developing navigation algorithms. In all video clips the human interactor and the robot are starting from the defined positions see Figure Fig. 33 on the following page. The interactor starts walking when the robot starts moving and the video ended when both are reaching their defined stopping positions (see Figure Fig. 33 on the next page). Figure Fig. 34(a) on the following page shows an example of the first person view.

Conditions Due to our setup with three different angles, three different robot distances, and two different person distances, we tested $3 \times 3 \times 2 = 18$ approaching scenarios. Each scenario was displayed as a first-person perspective video clip once per participant in random order.

Dependent Measures In order to determine the most legible robot navigation behavior we measured the trajectory-predictability, by asking the participants to predict the robot navigation behavior in the immediate future. We asked if the robot will change its behavior with a yes/no scale and if the answer was yes, how the robot would change its behavior. Here we asked for direction (stop, avoid to one side, backwards) and velocity changes (stop, accelerate, decelerate). Due to cultural differences for the preferred avoiding side, like, for example, left side for english people and right side for german people, we did not specify the side for an avoiding movement.

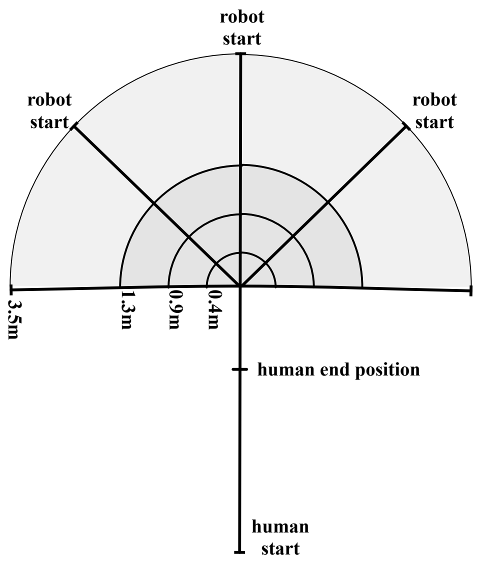


Fig. 33: Experimental design. The pictures are showing the approach directions and the respective points for robot and confederate where we stopped the video.

Participants We recruited 31 participants with an average age of 23.33 - thereof eight female and 22 male. Two participants had regular contact with robots, five from time to time and the remaining 23 had rarely or no contact with robots.

Procedure In order to familiarize the participants with the robot and to get a feeling for the room we showed them two introductory videos from a third person perspective (see Figure Fig. 34(b)) presenting a person and a robot crossing the person's path. The person in the video is the same gender as the participant, with this we foster the personification with the confederate. After seeing the introductory videos, we ask the participants to imagine that they would be the person in the scene

(a) First person view as provided in the videos

(b) Confederate and the robot from a third person perspective.

Fig. 34: Screenshots of the used videos. The first picture (a) shows the first person view towards the robot. The second picture (b) shows the confederate and the robot in our lab enviroment as shown to the participants in the introduction videos.

and encounter the robot and introduced the questionnaire. Afterwards, the experiment started. Every of the 18 video was shown once per participant in random order, each video was followed by three questions where we asked the participant to predict the robot's navigation behavior in the immediate future. First, we asked the participants to judge if the robot will change its behavior. If the participant answered "yes" we asked the participants to predict how the robot will change its behavior (stop, avoid to one side, drive backwards). We also asked if the robot will change its velocity (faster, slower, no change). After the experiment, participants were orally debriefed.

4.3.1.2. Results

We used SPSS to perform the data analysis. Due to the non-parametric character of our data, differences of frequencies were analyzed with a Pearson's Chi-Square test and the correlation analysis by calculation the point–biserial correlation coefficient.

Change Behavior First, we asked the participants to judge whether the robot will change its behavior or not. All results regarding this judgement are presented as bar charts in Figure Fig. 35 on the following page. In only 95 (17%) of the 558 observation tasks, the answer was no. We found a significant associations between the side (left, right, middle) and the judgement, $\chi^2(2) = 6.115$, $p < .05$, and also for the robot distance and the judgement whether the robot will change its direction or not, $\chi^2(2) = 11.747$, $p < .05$. We found no significant association for the variable distance human. By looking at Figure Fig. 35(b) on the next page one can presume that the higher the distance of the robot to the crossing point the higher is the number of the answer that the robot will not change its behavior. A correlation analysis confirmed that the distance of the robot to the crossing point is significantly correlated with the change behavior judgment, $r = -.124$, $p < .001$, showing us the lower the distance the higher is the probability that the participant wants the robot to change its behavior. No significant correlation was found for the distance of the human to the crossing point, $r = -.072$, $p = .091$.

Direction Within the second question we asked all participants, who answered yes to the first question ($n = 463$), to judge how the robot will change its navigation behavior regarding the direction (avoid to one side, stop, drive backwards). All results regarding the direction are presented as bar charts in Figure Fig. 36 on page 69. We found a significant association between the robot distance and the predicted direction (avoid to one side, stop, backwards), $\chi^2(4) = 63.21$, $p < .001$. As depicted in Figure Fig. 36(a) on page 69 one can see that for the nearest robot distance (40 cm) the most frequent answer was backwards (35.8 %), which is as compared to the other two distances (90 cm = 10.6%, 130 cm = 3.6%) quite high. We also found a significant association between the approach direction (left, frontal, right) and the predicted direction, $\chi^2(4) = 28.69$, $p < .001$. When the robot approaches from the left or right side the most frequent response was stop (left=52.8%, right=49.7 %) whereas for the frontal approach most participants wanted the robot to avoid to one side (55%), see Figure Fig. 36(b) on page 69. The distance of the human to the crossing point revealed no significant association with the predicted direction, $\chi^2(2) = 4.93$, $p = .08$. By considering the distributions in

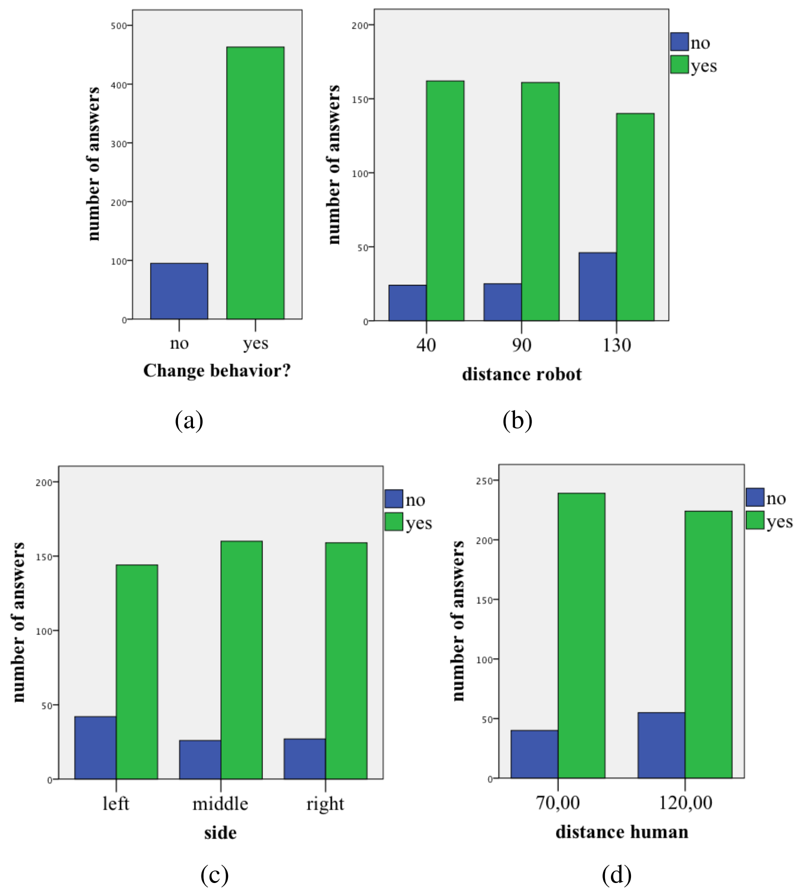
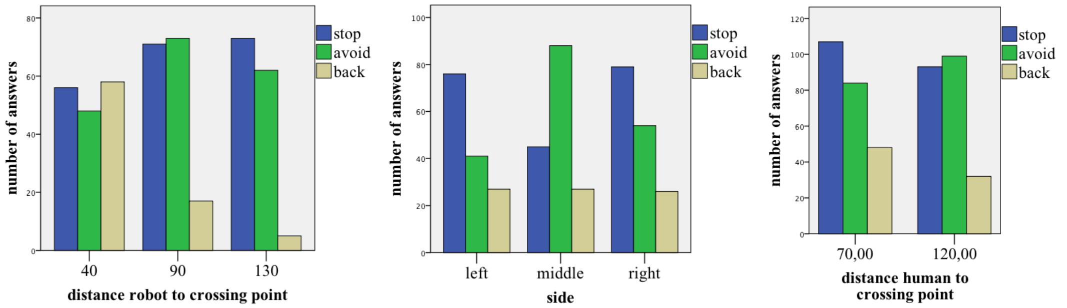


Fig. 35: Bar charts displaying the number of answers regarding the question whether the robot will change its behavior regarding side and distance.

the side chart (see Figure Fig. 36(b) on the next page) one can presume that the most legible navigation behavior for a frontal approach is to avoid the human by moving to one side. For a side approach (left and right) to stop and let the human pass. Furthermore, considering the distribution of the predicted navigation behavior regarding the robot distance (Figure Fig. 36(a) on the facing page) and in particular the high frequency of the answer backwards for the 40 cm distance one can conclude that this near distance is perceived as too near. The participants want the robot to go away from them. In order to further investigate the influence of the independent measures side and distance on the predicted navigation behavior we performed a correlation analysis. We found a significant correlation for the distance of the robot to the crossing point, $r = -.280$, $p < .01$, as well as for the side, $r = 145$, $p < .01$. Due to the symmetry of left and right side approach we merged right and left side for the correlation analysis. Based on the results at hand, we can derive that the most legible robot behavior in a human-robot path crossing scenario is for a frontal approach of the robot to move to one side at a distance of 90 cm to 130 cm and for a side approach of the robot to stop at a distance of 90 cm to 130 cm and let the human pass.

Velocity Additionally to the question regarding the direction, we also asked the participants, who answered yes to the first question ($n = 463$), to judge how the robot will change its velocity (equal = no change in velocity, faster, slower). All results regarding the independent measures distance and side are depicted in Figure Fig. 37 on page 70. We considered only the ratings for the participants who



(a) Type of direction regarding the distance of the robot to the crossing point.

(b) Type of direction regarding the navigation behavior

(c) Type of direction regarding the distnce of the human to the crossing point

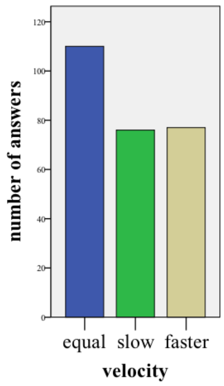
Fig. 36: Bar charts displaying the number of answers regarding the question how the robot will change its direction

rated the direction with avoid or backwards ($n = 263$), because, for the navigation behavior stop, the velocity is predefined. Altogether, the most frequent answer was equal (41.8%), see Figure Fig. 37(a) on the following page. As well as for the previous questions, the distance of the human to the crossing point revealed no significant association with the predicted velocity change, $\chi^2(2) = 2.47$, $p = .290$, (see Figure Fig. 37(b) on the next page). We found a significant association for the side value, $\chi^2(4) = 14.797$, $p < .005$, as well as for the distance value, $\chi^2(4) = 25.307$, $p < .001$. Equal and faster scored almost identical (equal = 40%, faster = 39.1%) for the frontal approach, see Figure Fig. 38(d) on page 71.

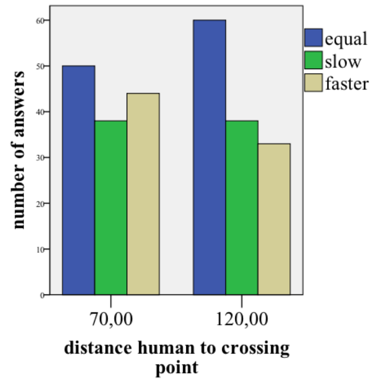
In order to figure out which combination of velocity and direction is the most legible navigation behavior, we analyzed the velocity ratings regarding the predicted directions avoid and backwards. As also in the former analysis we excluded the stop rating due to the predefined velocity of a stop. The most frequent answer was "no change in velocity" when the direction was judged as avoid (46.4%), see Figure Fig. 38(a) on page 71. We found a marginally significant association between the predicted navigation behavior and the velocity judgement, $\chi^2(2) = 5.34$, $p = .069$. Furthermore, by also considering the distance of the robot to the crossing point we found the robot was predicted to be faster mainly for the nearest distance of 40 cm (see Figure Fig. 38(b) on page 71). The most frequent combination for the 40 cm distance was to move faster and avoid to one side (24.5%), followed by move faster and drive backwards (20.8%). This supports our assumption that the 40 cm distance is way to far for a comfortable interaction and the participants want the robot to move away as quickly as possible. For the 90 cm and 130 cm distance the most frequent combination was to avoid with a constant velocity (90 cm = 43.3 %, 130 cm = 53.7 %). We decided not to take into account the human distance to the crossing point due to the minor effect on the ratings before.

4.3.2. Conclusion

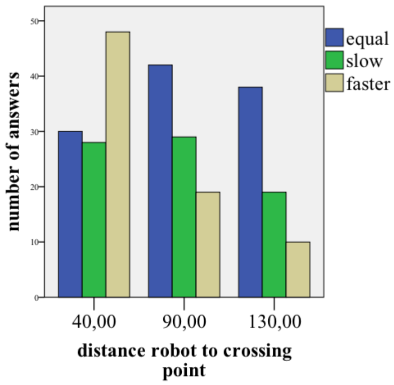
Based on the result, that a high majority of participants predicted a change in the robot's behavior, we can conclude that the robot has to be polite and change its behavior in a crossing scenario. This result is in line with our former finding (see 4.2.3.2. on page 61) that a robot has to be safe and make sure not to collide with a human. This result is in line with a finding of Dautenhahn et al. [27] regarding



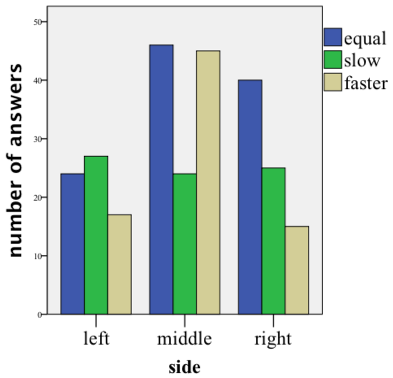
(a) Overall velocity ratings.



(b) Velocity rating regarding the distance of the human to the crossing point



(c) Velocity rating regarding the distance of the robot to the crossing point

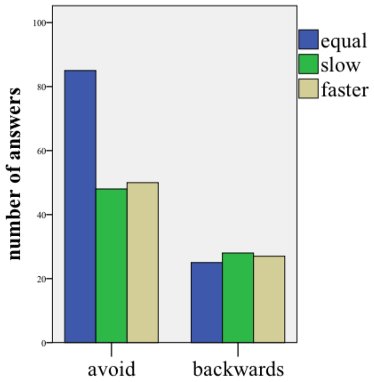


(d) Velocity rating regarding the side

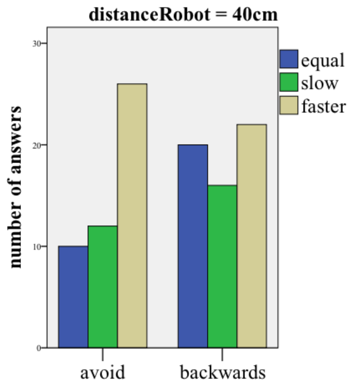
Fig. 37: Bar charts displaying the number of answers regarding the question how the robot will change its velocity

robot movements. In their experiment, most of the participants also wanted the robot to be polite and give way to them.

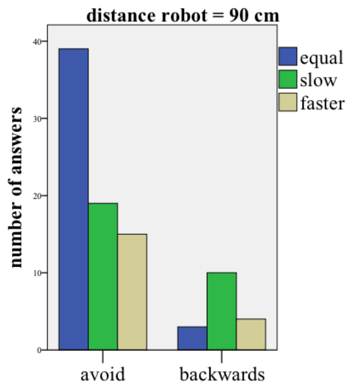
In order to determine the most legible behavior we asked the participants to predict the future navigation behavior of an approaching robot in terms of direction (backwards, stop, avoid to one side) and velocity (slow, equal, faster). We manipulated the approach direction and the distance of the robot to the crossing point and the distance of the human to the crossing point. We found that the most legible navigation behavior for a frontal approach was to avoid the human by driving to one side with equal velocity at a distance of 90 cm to 130 cm. For a side approach, the most legible behavior was to stop at a distance of 90 cm to 130 cm and let the human pass. Furthermore, we found that most participants wanted the robot to drive backwards at a distance of 40 cm, which is within the private space of a human [48]. We assume that this result show that the participants do not want to have the robot within their private space. Additionally to our results regarding the most legible navigation behavior we made another interesting finding. We found no significant association of the human distance to the crossing point with one of the three judgements, as opposed to this the distance of the robot to the crossing point has a significant association with every judgement. The same applies to the side value. To this end, we can assume that the approach direction (left, right, frontal) and the robot's distance can be used as predictors for the expected robot behavior whereas the distance of the human to the



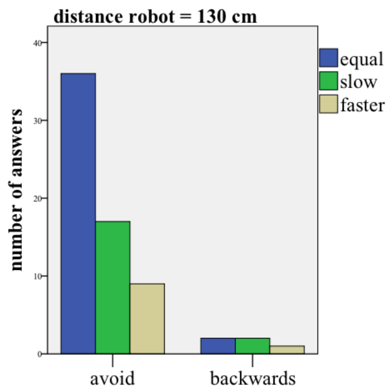
(a) Velocity rating regarding the predicted navigation behavior



(b) Velocity rating regarding the predicted navigation behavior for the robot distance 40 cm



(c) Velocity rating regarding the predicted navigation behavior for the robot distance 90 cm



(d) Velocity rating regarding the predicted navigation behavior for the robot distance 130 cm

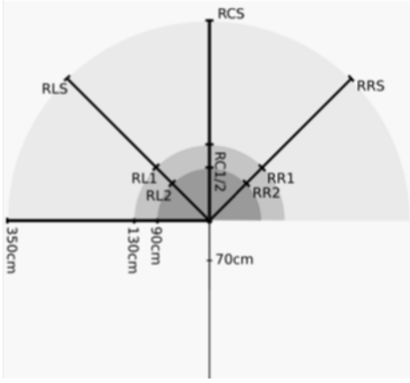
Fig. 38: Bar charts displaying the number of answers for the velocity rationgs regarding the direction rating

crossing point has, for our two tested distances, a minor influence.

4.3.3. Experiment to Verify Legible Navigation Behavior

The objective of the second experiment is to verify the previous findings. To this end, we adjusted the setup used in the previous experiment. We let the robot approach a human from different directions, like in the previous experiment. We modified the setup by also showing the participants how the robot behaves during the crossing situation like we did it in our former experiments (see 4.2.2. on page 46) to quantify legibility by measuring predictability, as well as the expectation-fulfillment factor. In addition, we measured the HRI property likability to answer the question if legibility is correlated with likability.

(a) Third person view as provided in the videos showing the confederate and the approaching robot.



(b) Experimental design. The picture shows the approach directions and the respective points for robot and confederate where we stopped the video.

Fig. 39: Experimental setup of the follow up experiment.

4.3.3.1. Experimental Method

In order to avoid learning effects, as mentioned as a limitation in the previous experiments 4.2.3.3. on page 63, we implemented a randomized between-groups (independent measures) design where every participant observes every crossing situation only once. We randomized the order of situations as well as the performed robot behavior. Therefore, every participant has a different set of crossing event observations. To achieve the high number of participants, needed for a between-group design, we conducted an online survey. Screenshots of the online questionnaire are presented in the Appendix (see Fig. 54 on page 104, Fig. 55 on page 104, Fig. 56 on page 105, and Fig. 57 on page 105).

Design For the follow-up experiment, we used the same technical equipment (B21 Robot and camera) as in the previous experiment and an adjusted setup. Due to visibility reasons we provided a third person view from above the scene (see Figure Fig. 39(a)). Based on our findings regarding the human and robot distance to the crossing point we used only one distance of the human (70 cm) and the two outer distances of the robot (90 cm, 130 cm) to the crossing point (see Figure Fig. 39(b)). In order to verify our previous results we let the robot perform the following actions: stop, drive backwards, drive forward (= no change in behavior) and avoid, all performed with no change in velocity, except for stop. Due to our design we videotaped the different path crossing scenarios where the robot crosses the human path and performs one of the four behaviors. The robot starts to change its behavior from only forward approaching the human to one of the defined behaviors when both are reaching the defined distances (see Figure Fig. 39(b)). This is also the point where we divided the videos into two parts.

Conditions With 3 approach directions, two different distances of the robot to the crossing point and four different navigation behaviors, we tested $3 \times 2 \times 4 = 24$ crossing scenarios.

Dependent Measures In order to quantify legibility we measured the trajectory-predictability, by asking the participants to predict the navigation behavior of the robot in the immediate future by using a text field, and if the robot met the participants' expectations by using a five-point Likert scale. Furthermore, we measured likability by using the Godspeed-III questionnaire [7] in which the participant has to rate his/her emotional state regarding the properties like – dislike, friendly – unfriendly, kind – unkind, pleasant – unpleasant and nice – awful on a five point Likert scale.

Participants 96 participants took part in our inline survey – thereof 37 women and 59 men with an average age of 29.1. Three of the participants had regular contact to robots, 10 from time to time and 83 had rarely or no contact to robots. We had a minimum of 20 participants per condition.

Procedure First, we presented an introductory video showing an example of the human-robot crossing scenario in the third person view (see Figure Fig. 34(b) on page 66). After that, we introduced the questionnaire and asked the participants to imagine that they would be the person in the scene and encounter the robot. Afterwards, the experiment started. We showed each participant one video for each of the 6 approaching situations in randomized order and selected the robot behavior by chance. After seeing the first part of the video, showing only the approaching up to the defined stopping point, we asked the participant to predict how the robot would behave in the near future. We used a text box for the answer. Afterwards, we showed the second part of the video, showing the robot's actual navigation behavior (stop, drive forward, drive backward or avoid) and asked the participants to judge how much the actual behavior met their expectations on a five point Likert scale. Additionally, we ask to for likeability using the Godspeed-III questionnaire [7].

4.3.3.2. Results

We used SPSS to perform the data analysis. We mapped the text answers to our first question manually to one of our four navigation behaviors (stop, drive backwards, drive forward = no change, avoid to one side). For the discrete dichotomous variable predictability differences of frequencies were analyzed with Person's Chi square test and for the expectation-fulfillment ratings we performed a one-way independent ANOVA. For post hoc analysis, we applied a Tukey's HSD.

Legibility The objective of the second experiment was to verify previous results. In order to determine the legibility of each navigation behavior we measured predictability, as well as expectation-fulfillment. The legibility factor behavior predictability was measured by asking the participants to predict the future robot behavior. As a measurement for the predictability of each behavior, we count the number of correct predictions. Results are shown in Figure Fig. 40 on the following page and Fig. 41 on page 75. We found a significant association between the performed navigation behavior and predictability for all approach directions, $\chi^2(3) = 58.55$, $p < 0.001$, and also for the frontal approach, $\chi^2(3) = 40.51$, $p < 0.001$, and the side approach, $\chi^2(3) = 38.11$, $p < 0.001$. Furthermore, we measured the expectation-fulfillment factor using a four-point Likert scale. We per-

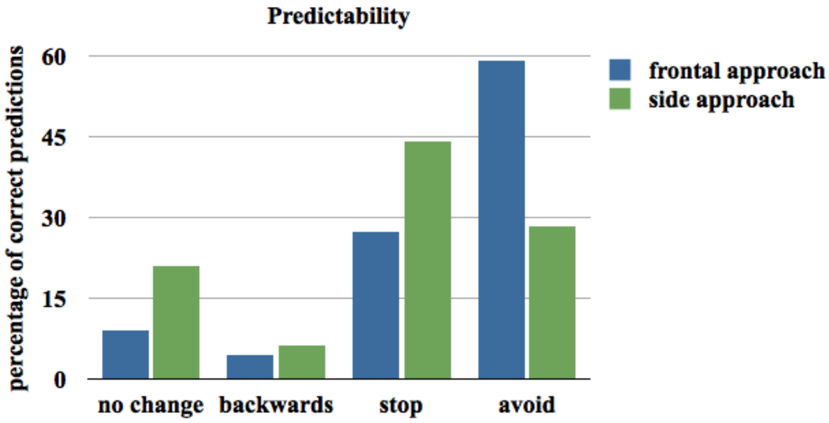


Fig. 40: Frequency of correctly predicted navigation behvior.

formed a found a significant effect of the presented navigation behavior on the level of how much the behavior met the participants expectations, $F(3, 572) = 26.50$, $p < 0.001$. By only considering the frontal approach data, we also found a significant difference between the expectation-fulfillment ratings, $F(3, 188) = 18.10$, $p < 0.001$. The same applies to the side approach data, $F(3, 188) = 14.46$, $p < 0.001$. A post hoc test revealed that for the frontal approach data the expectation-fulfillment ratings are significantly higher than all other ratings when the robot performs an avoiding behavior ($M = 1.94$). For a side approach, the stopping behavior ($M = 1.8$) revealed significantly higher expectation-fulfillment ratings than backwards or forward, but no significant difference between stop ($M = 1.8$) and avoid ($M = 1.45$) behavior.

The distance revealed no significant influence on predictability as well as on the expectation-fulfillment ratings. However, when we split the data based on the approach directions we found a significant association for the distance. Predictability was significantly higher for the avoiding behavior when the robot approaches frontally at a distance of 130 cm (90 cm = 54.5 %, 130 cm = 63.6 %), $\chi^2(3) = 40.51$, $p < 0.001$. The same applies for the side approach and the stopping behavior (90 cm = 42.6 %, 130 cm = 46.3 %), $\chi^2(3) = 38.11$, $p < 0.001$. This shows us that regarding predictability the robot behavior has to change at a distance of 130 cm. Nevertheless, the distance revealed no significant association on the expectation-fulfillment ratings.

To conclude, the results regarding trajectory-predictability and the ratings on how much the behavior meets the participants expectations verify our previous findings. For a frontal approach, the most legible robot navigation behavior is to avoid the human and for a side approach to stop and let the human pass.

Correlation Legibility – Likability Likability was measured using the Godspeed-III questionnaire [7]. We calculated the average value of the five likability measurements (like, friendly, kind, pleasant, nice) to get one likability measurement (0 to 3 4-point Likert scale). In order to investigate how the HRI property likability and legibility are correlated we calculated the Person's correlation coefficient for the expectation-fulfillment factor and the point–biserial correlation coefficient for the dichotomous legibility rating. Results are shown in Table 4.2 on the facing page. One can see that

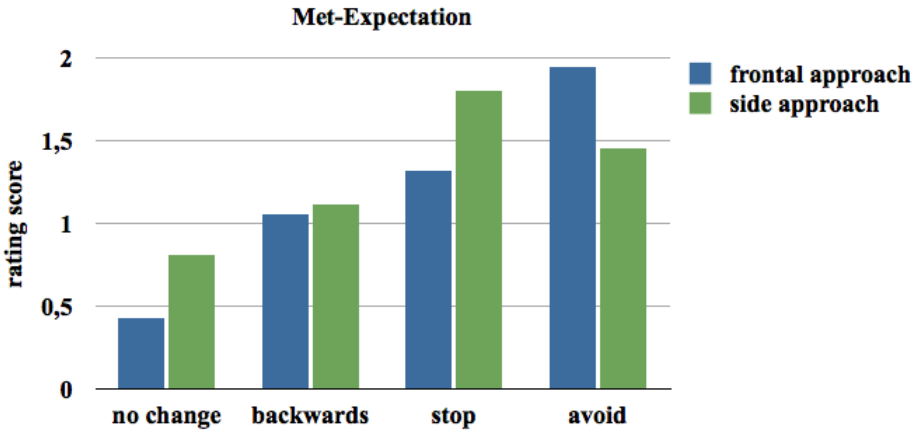


Fig. 41: Average rating score of the expectation-fulfillment factor for each navigation behvior. We used a four-point Likert scale from 0 to 3.

	expect.	likability	like	friendly	kind	pleasant	nice
predict.	.684	.255	.281	.235	.252	.224	.170
expect.	1	.425	.430	.395	.401	.364	.361

all $p < 0.01$

Table 4.2: Correlation coeffecients for all measurements of the experiment. Note, we calculated the point–biserial correlation coefficient, r_{pb} for the discrete dichotomy variable predictability and the standard Pearson's correlation coefficient r for all other variables.

expectation-fulfillment is significantly correlated with legibility, $r = .425$, $p < 0.01$, as well as the measured trajectory-predictability with legibility, $r = .425$, $p < 0.01$. Furthermore, Figure Fig. 42 on the next page shows the mean likability value for each performed navigation behavior. The navigation behavior stop ($M = 2.07$) and avoid ($M = 2.22$) are revealing higher likability ratings than backwards ($M = 1.79$) or drive forwards ($M = 0.66$), meaning no change in behavior. By considering each approach direction (frontal, side) separately we found for the frontal approach that likability differed significantly between the four navigation behaviors, (avoid: $M = 2.18$, stop: $M = 1.82$, backwards: $M = 1.86$, forward: $M = 0.60$), $F(3, 188) = 65.75$, $p < 0.001$. Post hoc tests revealed that the ratings for an avoiding behavior are significantly higher than for forward and stop ($p < 0.05$), and marginally higher for backwards, $p = 0.09$. For the side approach we found significant differences between the performed navigation behavior (stop: $M = 2.2$, avoid: $M = 2.25$, backwards: $M = 1.76$, forward: $M = 0.69$) regarding likability ratings, $F(3, 380) = 111.06$ $p < 0.00$. However, post hoc tests revealed significant differences for all ratings except for stop and avoid. The distance revealed no significant association on the likability ratings. To conclude, we found a significant correlation between legibility and likability and the results are confirming our former assumption that a polite robot behavior is preferred.

Reliability of the used Questionnaire In order to evaluate the reliability of our questionnaire we calculated the Cronbach's α for the Godspeed-III likability scale and the expectation-fulfillment rating. It revealed that all had high reliabilities, all Cronbach's α = .916.

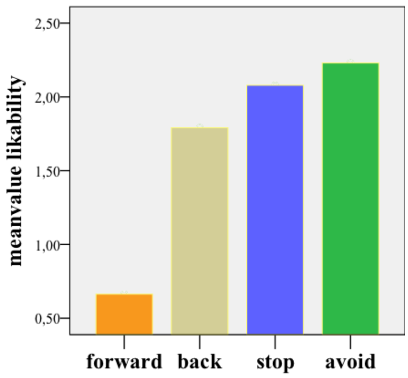


Fig. 42: Bar chart showing the mean likability values for each performed behavior

4.3.4. Discussion

The objective of the two conducted experiments was to identify which robot navigation behavior is the most legible in a path crossing situation. To this end, we tested a variety of possible robot behaviors in three different crossing situations each one differing in the approaching angles (frontal and side 45°). In the first of our two coherent experiments, we found that the most legible robot navigation behavior in a human-robot path crossing scenario is to be polite and give way to the human that is in line with results from Dautenhahn et al. [27]. We also found that it is most predictable, when the robot stops in a side approach situation, and when the robot approaches frontally to avoid the human by moving to one side. Furthermore, our results let us assume, that the spatial relationship theory investigated by Hüttenrauch et al. [54] in a home tour scenario is also applicable for path crossing situations. Hüttenrauch concluded that most participants preferred a distance within the interpersonal space (45 cm to 120 cm) for the interaction. With our finding that most of the participants want the robot to move away from them at a distance of 40 cm, which is within the private space, we could support the assumption that the distance within the personal space and not in an intimate space is more preferred in a path crossing setting. At this point, we want to point out that contrary to Hall's proxemics theory [48] we use the crossing point (see Figure Fig. 33 on page 66) as the center point for the personal spaces. Particularly, we found that the distance of the human to the crossing point do not have a significant influence. Therefore, we assume that the participants' response to our questions is based on a projection of themselves to the crossing point. However, this interesting hypothesis about the proxemics of crossing situations has to be further investigated in the future.

Our findings were evaluated in a follow-up experiment and we could verify our former conclusions. It reveals, that the most legible robot navigation behavior is to stop within the personal space in a side approach scenario and to avoid the human by moving to one side in a frontal approach scenario. Also the follow-up results are confirming the previously mentioned findings of Dautenhahn et al. [27], that the robot has to be polite and give way. Furthermore, it is also comparable to the findings of the human-human path crossing experiment conducted by Basili et al. [9]. Namely, that a human prefers to go straight and decelerate in order to avoid a collision in a side approach situation. This relation of similar findings in human-robot and human-human experiments supports the assumption that human-like behavior tends to be legible. Furthermore, our finding also supports that a high number of navigation methods especially for frontal approaches, which are forcing the robot to move to one side [58, 78, 115], producing legible robot navigation behavior.

Additionally, we found a significant correlation between legibility and likability, which is in line with findings from Takayama et al. [106]. A legible robot is not only perceived as safe it is also perceived as more likable, which supports our previous assumption, that legibility is an influencing HRI property and it has to be taken into account.

Limitations One limitation of our experiment is the small number of distances. We tested only three different robot distances to the crossing point and two different human distances. We suggest further to investigate farther distances in order to see where exactly is the preferred distance for a robot reaction and is there is a distance where no reaction of the robot is more legible. As we investigate human-robot spatial interaction we also have to take into account that human-robot proxemics is also influenced by the robot appearance [63] or the participants' personality [107]. One limitation is that we hold appearance factors constant and did not further investigate the influence of personality factors.

4.4. Conclusion

Our first experiments where we evaluated state-of-the-art navigation algorithms, following different approaches how to take into account interacting humans, showed us, that legibility is rather low in current navigation methods. However, the results let us assume that taking the human into account as a particular moving object and following social norms makes the robot navigation not only more legible but were also perceived as safer, more comfortable and more reliable. The second series of experiments, where we evaluated different navigation behavior patterns showed us that polite and humanlike behavior is the most legible behavior. Stopping for side approaches and avoiding for frontal approaches were measured as highly legible and our results showed that for human-robot spatial interaction the personal space is the preferred interaction distance.

Predictability and Expectation-Fulfillment First of all, our experiments reveal a strong and significant correlation between predictability and expectation-fulfillment. This supports our theoretical definition of a complex property legibility consisting of the factors predictability and expectation-fulfillment with experimental results.

In all our experiments, we found that the expectation of the participants was met with a higher extent as the behavior was predicted correctly. Furthermore, what is also very interesting that the correlation values between the additionally measured HRI properties safety, comfort, reliability, and likability were always higher for the expectation-fulfillment factor than for the behavior- or goal-predictability. These findings let us assume that humans expect more than one behavior. We hypothesize that humans expect different kinds of behaviors with different probabilities and different values like likeliness or safety. Alternatively, it is one abstract behavior like "avoid" for a frontal approach crossing situation and the way the robot performs this behavior influences other properties, which has already been investigated by other researchers [23, 58, 78, 88]. One possible opportunity to investigate our assumption is to repeat one of our experiments or design a similar one with one adjustment. Instead of letting the participants predict one distinct behavior one can provide different kinds of possible behaviors and let the participant rate how much he/she would expect the behavior.

Correlations Within our conducted experiments, we found significant correlations between legibility and safety, comfort, reliability, and likability. This result shows us that legibility is an important HRI property and has to be taken into account when developing algorithms for human-robot interaction.

5. Determine Human Expectations of Robot Navigation Behavior

In this chapter, we first present our self-developed "Inverse Wizard of Oz" method to capture human expectations of robot navigation behavior in a real-life scenario. After that, we show how we used our ”Inverse Oz ofWizard” method by letting participants steer a robot in a scenario in which an instructed person is crossing the robot’s path. We investigated two aspects of robot behavior: (1) What are the expected actions? (2) Can we determine the expected action by considering the spatial relationship?

5.1. Introduction

The experiments in the previous chapter are questionnaires-based and focus on measuring legibility of predefined robot navigation behavior. We identified which the evaluated robot navigation behaviors are the most legible. One limitation we pointed out in the previous chapter was the very limited number of spatial relationships in the crossing scenarios. For example, we could not answer the question if the expectations will change when the robot is farther away from the crossing person. Furthermore, all previously conducted experiments were video-based and we are not able to confirm, that the results regarding legible robot navigation behavior is transferable to a real-life scenario. In order to overcome this limitation with the study at hand we (1) will not predefine the robot’s behavior, (2) analyze a high number of spatial relationships, and (3) use a real-life interaction scenario. Furthermore, we also want to go one step towards the methodical implementation of a legible robot navigation method and investigate if it is possible to predict what kind of robot action is the best choice regarding legibility. Due to the infinite number of crossing situations, which can occur in real-life situations, the robot needs a decision function, that uses the robot’s sensor input, to determine the most legible behavior. Initially, we have to figure out what kind of sensor input the robot can use and what are the significant predictor variables? Therefore, we have two main objectives within the study at hand: (1) identify legible robot behavior in real-life path crossing-situations and (2) find predictor variables in order to build up a system that can decide how to behave in arbitrary crossing situations.

Related Work A large body of research is dedicated to investigate several aspects of robot navigation behavior [18, 23, 54, 88]. For example, Butler et al. [18] analyzed the influence of different factors like speed, distance and design on robot motion patterns (frontal approach, passing by, non-interactive navigation). They evaluated how the navigation is perceived by humans regarding the level of comfort. Pacchierotti et al. [88] also tested the conditions speed, and distances in a passing by situation. Proxemics [48] is also widely studied in human-robot interaction [84, 107]. In a controlled experiment Dautenhahn et al. [23] identified preferable robot approaching motions by testing different strategies. Former research shows that the spatial relationship influences how the robot motion is perceived. Furthermore, navigation is based on the spatial relationship between the robot and its environment. Therefore, we decided to investigate if we can use the spatial relationship of the robot and the crossing human as predictor to determine which behavior is the best choice in terms of legibility.

However, the aforementioned research presents controlled experiments testing how motion is perceived regarding different conditions like speed, distance and orientation like we investigated legibility in our previous experiments (see chapter 3. on page 29). However, what is only rarely investigated is what humans expect from a robot in an unstructured study without predefined behavior patterns. Hüttenrauch et al. [54] presents one study towards expected robot motion patterns. They investigated the spatial behavior (distances according to Hall [48] and orientation according to Kendon [57]) of a human towards a robot during a "Home Tour" study. In their experiment, the robot behavior is steered but the participants are moving freely. Dragan et al. [31] presents another method to capture the participant's expectations of robot movements. They ask the participants to draw the expected trajectory.

5.2. The Inverse Wizard of Oz Method

In order to achieve our objectives, we developed the "Inverse Wizard of Oz" method. It is a new type of the commonly used "Wizard of Oz" method [44]. The idea is to let the participant steer the robot. Consequently, the participant took over the role of the "Wizard" and the interacting person is a confederate, therefore, we call our method "**Inverse** Wizard of Oz." We assume that the way someone steered the robot reflects his/her own expectations regarding the robot in the situation. Following the categorization proposed by Steinfeld et al. [104] with our "Inverse Wizard of Oz" method we (1) "simulate" the human by instructing a confederate with behavioral rules, (2) the robot behavior is controlled by the participant, and (3) the robot behavior is evaluated, in order to capture the participants expectations regarding the robot's behavior. It is similar to the drawing method proposed by Dragan et al. [31] with the advantage that we can evaluate interactive HRI scenarios and are not limited to a two dimensional drawing of a line. The "ghost in the machine" paradigm presented by Loth et al. [77] is also very similar. In their bartender experiment, they also let the participants control the robots actions in an interactive scenario. However, different to our method is that they predefine the robot's actions and the participant have only the limited robot sensor information available. By letting the participant freely steer the robot with all possible capabilities in a human-robot path crossing scenario, we do not predefine the robot's actions and thus, our "Inverse Wizard of Oz" method fulfills our objectives.

5.3. Mathematical Formulation of Robot Behavior and Spatial Relationship

Regarding the implementation of our findings into a robot navigation algorithm and for structured analysis of our results we mathematically formalize the terms "robot behavior" and "spatial relationship," because one needs machine processable data to make further use of our results in coded algorithms.

Robot Behavior First, we mathematically formalize the term "robot behavior." The term "behavior" is defined as the range of **actions** done by organisms, or artificial entities that are the **response** of the organism, or artificial entity to various **stimuli**. Alternatively, as Arkin shortly states in his book, *"behavior is a reaction to a stimulus"* [5]. Based on this general definition and according to Arkin's behavior-based robotics theory we formalize robot behavior as an **action *a*** performed by a robot, which is **caused** by a **stimulus *s*** , i.e., it exist a function f so that $f(s) \mapsto a$. In our navigation scenario, possible actions can be driving, stopping or driving a curve and the spatial relationship is a stimulus.

Definition 3 (Robot Motion Pattern RMP)**.** *A Robot Motion Pattern RMP is defined as a simple* ***action*** $a \in A$ *performed by a robot, which is* ***caused*** *by a* ***stimulus*** $s \in S$*, i.e., it exist a function* f *so that* $f(S) \mapsto A$*.*

For example, if the robot is waving its arm as a response to an approaching human, then waving the arm is the action a_w and the stimulus s_w is the approaching human which can be formalized ,for example, as spatial relationship.

Spatial Relationship When talking about the spatial relationship in HRI scenarios we specify how the robot is located in space in relation to one (or a group of) human(s). Different from pure mathematical applications where mostly the distances and angles are of interest in HRI distances are used in proxemics [48, 84, 107], but other social and interaction aspects are also important information. For example, the way people arrange each other in a conversation [4, 57] or how they body and head are oriented towards a robot bartender [38]. With spatial relations humans communicate intentions [38, 90]. In particular, for navigation tasks it is important to know if a human is passing the robot path or not and whether it is a head-on encounter or a 90 degree crossing. Hanheide et al. [49] presents one approach to represent the spatial relationship in human-robot path crossing scenarios. They adapted a representation for moving objects in geographical systems [111] to HRI.

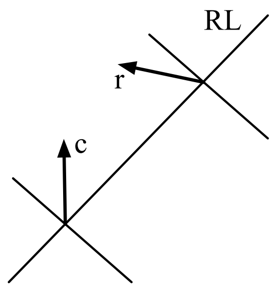


Fig. 43: QTC_c double cross (reproduced from [49])

Calculating QTC_c Features: QTC_c is a compact representation of spatial relations between two moving objects. It represents the relative motion, with respect to the reference line RL that connects them, as shown in Fig. Fig. 43. As Hanheide et al. [49] proposed, we use a simple version of QTC_C (see [111]) that considers only the distance and the relative direction of each point with respect to the other. In this case, four qualitative relations are defined as follows:

(1) movement of the robot r with respect to the *customer* c:

- 1 : r is moving towards c
+1 : r is moving away from c

(2) movement of c with respect to r
same as (1), but with r and c swapped

(3) movement of r with respect to RL:
-1: r is moving to the left-hand side of RL
0 : r is moving along RL or not moving at all
+1 : r is moving to the right-hand side of RL

(4) movement of c with respect to RL
same as (3), but with r substituted by c

All these features are easily and fast to calculate using the dot and cross product of the vectors c and r. Therefore, these representation is applicable in online methods. In our experiments we investigate crossing situations. To this end, we need a technique to identify a crossing situation in the captured spatial data. By using QTC_C we can easily determine crossing situations. Given the above definition a crossing situation is given when:

$$(1) = -1, (2) = -1, (3) = -1 \cdot (4) \text{ or } (3) = (4) = 0.$$

According to the proxemics investigations [84,107] we investigate the distances in crossing situations. However, different from the classical distance measuring in proxemics where only the direct distance between human and robot is considered we consider also the distances of human and robot with regard to the crossing point and the crossing angle (see Fig. Fig. 44).

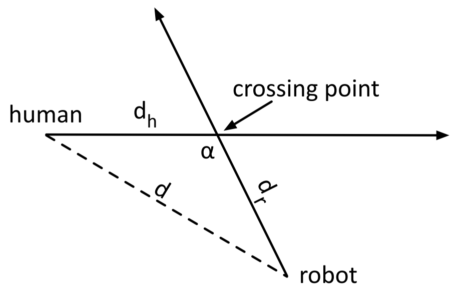


Fig. 44: Distance measures used for spatial relationship investigations of crossing situations. With d we denote the distance between human and robot, with d_h the distance of the human to the crossing point, with d_r the distance of the robot to the crossing point and with α the crossing angle.

With the study at hand, we investigate how these spatial features influence the human regarding his/her reaction to a crossing situation.

5.4. Determine Legible Robot Navigation Behavior

The objective of the study at hand is to determine legible robot motion patterns (RMP) in real-life path crossing-situations. Based on our previous video-based experiment (see 4.3. on page 64) we expect that stopping the robot and let the human pass is a legible navigation behavior. Different to our previous experiments we will not predefine the robot actions. To this end, we use the "Inverse Wizard of Oz" method and let the participant steer the robot in a grocery store scenario where a human crosses the robots path by chance. Due to our scenario, where the crossing situations are completely random we can investigate a huge amount of different situations regarding crossing angle and distances.

Research Questions With the study at hand we want to identify legible actions a in a human-robot path crossing scenario and determine the stimulus s. We expect that the stimulus for action in a human-robot path crossing situation lies in the spatial relationship of human and robot. Furthermore, we want to verify that it is possible to predict the expected action, based on the spatial relationship. I.e. a function f exist so that $f(s) \mapsto a$.

5.4.1. Method

We implemented a within-subject study using the "Inverse Wizard of Oz" method (see 3.4.5. on page 37) and let the participant steer the robot's movements in a grocery store scenario. A confederate crossed the robot path by chance. Therefore, we investigate how the robot should behave in a path-crossing situation and if the spatial relationship influences the reaction.

5.4.2. Participants

We recruited 46 participants with an average age of 28 years - thereof 26 women and 20 men. 89% of the participants had rarely or no contact to robots and 11% had regular contact to robots. The participants gave their written consent to the study.

5.4.3. Technical Setup

5.4.3.1. Robot

The platform used in this study was the BIRON (BIelefeld Robotic CompaniON) robot [46] (see Fig. Fig. 45(a) on the following page). It is a combination of a two-wheeled PatrolBot™and GuiaBot™. BIRON has an overall size of approximately 0.5 m (w) x 0.6 m (d) x 1.3 m (h). Besides two wheels, BIRON has two rear casters for balance and is constructed with a differential drive (2 degrees of freedom: translation and rotation). This enables it to turn on the spot and drive curves while driving forward, but BIRON is not able to drive directly to its left or right (no holonomic drive).

(a) BIRON

(b) Wireless keyboard used as robot remote control

(c) Person steering the robot by using the remote control.

Fig. 45: Technical setup used in our study. The pictures are showing the used robot and the remote control that we used to steer the robot.

5.4.3.2. Robot Remote-Control

We used a wireless keyboard to steer the robot (see Fig Fig. 45(c)) and marked the commands of how to steer the robot on the keyboard with arrows (see Fig. Fig. 45(b)). Five keys corresponded to the five ways of moving the robot: straight forward, rotate around its own axis in a clockwise direction, in an anti-clockwise direction, drive and turn left or right in an arc. The robot only moved by holding down a particular key and the robot stopped by releasing the key. These motions map the actual movement abilities of the robot BIRON. Therefore, the actions we investigate in this scenario are drive, stop, curve, and rotate. Due to safety and usability reasons there was no possibility to accelerate the robot as it was driving at its full speed of 0.7 m/s.

In a brief pre-study, we tested different types of remote controls. A Joystick, a PS2 Controller, and a wireless keyboard. It turns out that the keyboard were the easiest to handle remote control, in particular for non-computer science people.

5.4.3.3. Motion Capturing System

To capture the movements of the robot and the confederate we used a VICON motion capturing system (`www.vicon.com`) with 10 (6 x T10, 4 x T20) infrared VICON cameras. We captured with a frame rate of 150 Hz. The robot and the person were equipped with rigid body markers constructed by B. Brüning and colleagues [91], therefore, we were able to track the position and also the orientation robot and person. We recorded additional video data with an HD camera. For synchronizing video and motion capturing data we used a clapper board equipped with markers.

5.4.4. Study Design

5.4.4.1. Cover Story

In order to make the scenario realistic all participants were told the same cover story about a grocery store that uses a robot (BIRON) to refill the shelves with goods from the storing place. The participants were asked to navigate the robot from the storing place to a shelf (see Fig. 1(c)) by steering the robot with a wireless keyboard. Furthermore, they were told that the robot might encounter customers in the store.

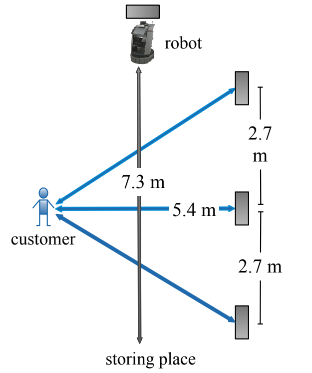
5.4.4.2. Setup

According to the grocery store cover story we built up a store scenario with four shelves and one storing place (see Fig. 1(b),1(c)) in a laboratory that measures approximately $133m^2$. Three shelves were placed at the wall on the right side of the storing place with a distance of 2.7 m between them (see Fig. 1(c)). One shelve was placed 7.3 m opposite to the storing place (see Fig. 1(c)). The shelves and the storing place were filled with typical grocery store products. The robot, steered by a participant, had the task to bring items from the storing place to the opposite shelve (see Fig. 1(c)). One experimenter took the role of a *customer*. The *customer* had the task to walk from a fixed point (see Fig. 1(c)) randomly to one of the three shelves at the wall and put an item into his/her basket. In addition to the three randomized aims the *customer* walked randomly in three different walking velocities slow (0.6 to 0.8 m/s), normal (1.2 to 1.5 m/s), and fast (1.9 to 2.1 m/s). The *customer* had to go straight and maintain the velocity even if the robot would crash into them. To avoid eye contact with the participant, the *customers* wore sunglasses. Due to the arrangement of the shelves the robot and the customer coincidentally met each other in $45°$ and $90°$ angles (see Fig. 1(c)). Thus, the setup was designed to create completely random and unforeseeable crossing events. Hence, we had nine different crossing conditions and a no-crossing control condition in random order. Two other experimenters helped the robot to sort and to put away the good.

5.4.5. Procedure

First, we welcomed the participants and explained the cover story (see Section 5.4.4.1.). In order to familiarize the participant with the setup and with steering the robot BIRON the participants received an introduction to the robot BIRON and an extensive practice of how to steer the robot. Only after the participants managed to drive around obstacles and felt capable of steering the robot, the study began. The participant were told to carry 15 items (only one item per time) from the storing place to the opposite shelve (see Fig. 1(c)) and then go back to the storing place. Therefore, the robot moves 30 times (two times per item) straight through the room. The *customer* crosses the robots path randomly as described in Section 5.4.4.. We captured the movements of the robot and the *customer* by using motion capture system and a video camera (see Section 5.4.3. on page 83). Once the participant completed the task they were debriefed about the purposes of the study and discussed the study with

(a) The picture of the laboratory shows a scene during the study. The robot is steered by a participant and moves fom the shelf to the storing place while the *customer* goes to one shelf and croses the robots path.



(b) The diagram depicts the paths the confederate is instructed to go.

Fig. 46: Grocery store study setup.

the experimenters. Demographical data were recorded within the debriefing.

5.4.5.1. Conditions

Due to our study design we manipulated the variables distance, velocity, and crossing angle of the *customer* with regard to the robot. Due to our setup, which produces random crossing scenarios regarding the independent variables distance, velocity and crossing angle, we have no fixed condition regarding the three factors distance, velocity, and crossing angle.

5.4.5.2. Dependent Measures

Regarding our research question to identify legible robot navigation patterns in a human-robot path crossing scenario, we measured the steered robot's actions $a \in \{\text{drive, stop, curve, rotate}\}$ and the spatial relationship s (see 5.3. on page 81) of robot and confederate (*customer*) during the interaction.

5.5. Results

We carried out the data analysis in two steps. First, we identified robot actions $a \in \{\text{drive, stop, curve, rotate}\}$ in the video by observing its motions as well as in the motion capturing data by analyzing the path and velocity. After that, we analyzed the spatial relationships of robot and *customer* using the motion capturing data in order to identify the stimulus for a specific action. We only consider crossing situations for our analysis. A crossing situation is defined as a situation where 1) the paths of both, BIRON and the *customer*, will cross and 2) both are located before reaching the crossing point (see Figure Fig. 47(b) on page 88).

5.5.1. Identified Actions a

By analyzing the video data, we identified four different navigation patterns performed by the robot in crossing situations. The first motion pattern, which the participants performed the most (76.7%), was driving straight towards the goal (shelve or storage place) and stopping (or stuttering) when the *customer* came close to the crossing point. In 75% of these situations the distance of the *customer* to the crossing point was between 0.58 1.8 mm and 1.8 m (median: 1.13 m) and BIRON stopped within a distance between 1.14 m and 2.13 m (median: 1.55 m) to the crossing point. 44 of 46 (95.6%) participants performed this pattern. The second motion pattern we identified was to drive along, passing the *customers* path far before or behind (18%). There is no risk of a collision in these crossing situations. All participants showed this pattern. The third identified motion pattern was to drive a curve in order to avoid a collision with the *customer* (3.7%). Similar to the first motion pattern BIRON was driving straight towards the goal and when the *customer* came close to the crossing point they started to drive a curve. 21 participants showed this behavior. The motion pattern "collision with the *customer*" (1.6%) was only shown by two participants. The participants were driving straight towards the goal without considering the *customer*.

1. stopping before the crossing point (76.7%)
2. driving straightforward and passing behind or in front of the *customer* (18%)
3. driving a curve (3.7%)
4. collision with the *customer* (1.6%) (only shown by two participants)

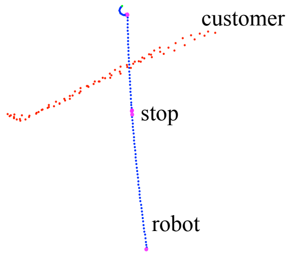
The action "rotate" was never performed in a crossing situation and "driving a curve" was mostly performed by a participant in his/her first trials. The overall navigation strategy, which we can conclude from our observation, was driving straight towards the goal (shelve or storage place) and either stop when both, the *customer* and the robot, are close to the crossing point, or otherwise drive on towards the goal and pass the path of the *customer* without colliding. This strategy was performed by almost all (44 of 46) participants. We assume that the participants anticipate if a collision happens or not and either stop or drive on. To conclude, from our video analysis we can derive that the actions drive and stop are the preferable actions and we can assume that the stimulus lies in the spatial relationship.

5.5.2. Identify Robot Motion Patterns

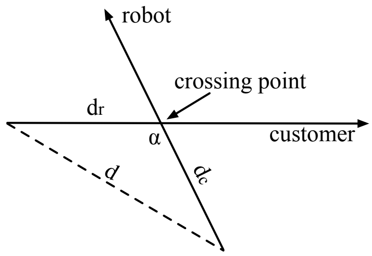
The raw data from the motion capturing system contains the position of robot r and *customer* c (see Fig. Fig. 47(a) on the following page) captured with a frame rate of 150 Hz. In order to describe the spatial relationship between robot and *customer* we calculated the following spatial features using Matlab (see also Figure Fig. 47(b) on the next page):

- QTC_c according to Hanheide et al. [49] to determine a crossing situation

- distance between *customer* and robot d
- distance between *customer* and the crossing point d_c
- distance between robot and the crossing point d_r
- angle robot- *customer* α
- velocity *customer* v



(a) Path of customer and robot with identified stopping point



(b) Distance features

Fig. 47: Spatial features of a crossing situation

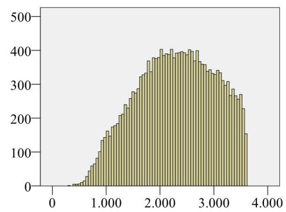
Calculating Spatial Features: For the purpose of reducing the amount of data, we calculated one feature vector for every 15 frames. Thus, we get 15 feature vectors for one-second of recorded data. Additionally to the aforementioned spatial features we determined the action a (drive, stop, curve, rotate) the robot is performing (see Fig. Fig. 47(a)). Thus, we transformed position data points into action related spatial feature vectors containing the action a, distance between customer and robot d, distance between *customer* and the crossing point d_c, distance between robot and the crossing point d_r, angle robot-customer α, velocity customer v, and the QTC_c values.

$$(a, d, d_c, d_r, \alpha, v, QTC_c)$$

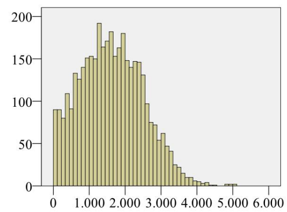
In order to concentrate only on the main strategy we excluded all feature vectors with curve and rotate actions. We identified the crossing situations by using the QTC_c [49] information and excluded all non-crossing situations. We also excluded all feature vectors where the robot is outside of the *customer's* social space ($d > 3.6m$) [48]. Note that we have far more than one feature vector per crossing situation, because different to the video analysis, where we only count the reaction of a crossing situation, we now consider every data point of a crossing situation. The robot drives before it stops, therefore, we have more feature vectors for the action drive than for the action stop.

Statistical Data Analysis The aim of the statistical data analysis, which was performed with SPSS, was to find support for our hypothesis: "can we predict the action based on the spatial features".

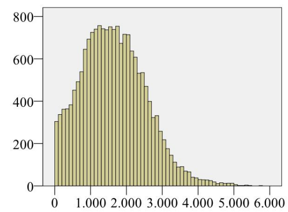
First of all, we show the distributions for the distance values (d, d_c, d_r) in Fig. 5.5.2. on the next page. The histograms in Figure 5.5.2. on the facing page show, that there is trend for greater distances for



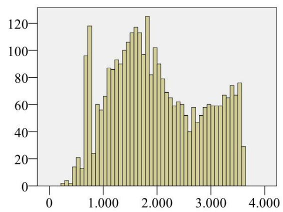
(a) d for action = drive
$Mdn = 2.32m$



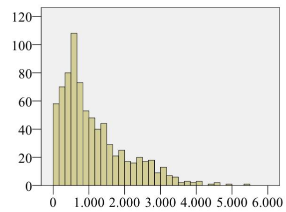
(b) d_c for action = drive
$Mdn = 1.58m$



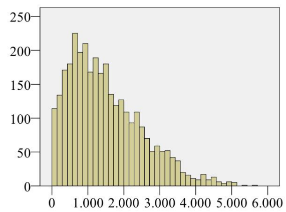
(c) d_r for action = drive
$Mdn = 1.59m$



(d) d for action = stop
$Mdn = 1.85m$



(e) d_c for action = stop
$Mdn = 0.85m$



(f) d_r for action = stop
$Mdn = 1.29m$

Fig. 48: Histograms of the distance values for action = {stop, drive} measured in mm

the drive action and that most of the participants stop the robot within a distance of approximately 0.7 m to the crossing point whereby the *customer* has a distance of approximately 0.67 m to the crossing point.

By performing inferential statistics we found support for our hypothesis that the spatial features are correlated with the action. Due to the dichotomous action variable we calculate the point-biserial correlation coefficients. The action of the robot was significantly related to the distance between *customer* and robot d, $r_{pb} = .149$, $p < .01$, to the distance between *customer* and the crossing point d_c, $r_{pb} = .200$, $p < .01$, to the distance between robot and the crossing point d_r, $r_{pb} = .063$, $p < .01$, to the angle robot-textitcustomer α, $r_{pb} = .027$, $p < .01$, and to the *customer's* velocity v, $r_{pb} = .054$, $p < .01$.

As a next step to support our hypothesis that we can predict the action based on the spatial features, we performed a logistic regression on the normalized spatial feature values. Results are shown in Table 5.1 on the next page. Only the angle has no significant influence on the classification model ($B = .06, p = .396$). All other variables have a significant influence and the model is able to predict 86.7 % of the actions correctly.

Additionally to the logistic regression we trained a Support Vector Machine SVM with a RBF Kernel [22] in order to show that we can precisely predict the expected action based on the spatial features. The results of a ten-fold-cross-validation, performed with LibSVM [19] are shown in Table 5.2 on the following page. The very good prediction results also support our hypothesis.

Value	B (SE)	Lower	Odds Ratio[a]	Upper
distance d	-1.023** (.086)	.304	.359	.425
velocity v	-.515** (.037)	.556	.598	.643
angle α	.060 (.71)	.924	1.062	1.221
distance robot d_r	.648** (.081)	1.630	1.912	2.243
distance customer d_c	-.255** (.071)	.674	.775	.892
constant	-2.131 (.059)		.119	

** $p < 0.001$, [a]95% CI for Odds Ratio

Model Statistics

$R^2 = .15$ (Cox & Snell), .25 (Nagelkerke)
$\chi^2(8) = 106.097$, $p < 0.001$ (Hosmer & Lemeshow)
accuracy 86.7% (goodness-of-fit)

Table 5.1: Results of the logistic regression, performed with SPSS

accuracy	99.9527%	f-score	0.999764
precision	100%	recall	99.9527%

Table 5.2: Results of the ten-fold-cross-validation of the SVM model performed with LibSVM [19]

5.6. Discussion

The study at hand was conducted to identify expected robot navigation behavior, which is consisting of an action a and a stimulus s, which causes the action, so that a function exist which maps the stimulus to action $f(s) \mapsto a$. In the video analysis, we found a prominent robot behavior. Driving straight towards the goal and, when a crossing situation occurs, either stop and wait until the customer passes the robot's path or drive on and pass the path of the person before or behind the person. Thus, the expected actions a are $drive$ towards the goal and $stop$. We hypothesized that the stimulus lies in the spatial relationship. Therefore, we use the motion capturing data to identify the stimulus s and calculated several spatial features. We used the QTC_c [49] representation to identify a crossing situation and extended these purely relative representations with distance measures, the crossing angle, and velocity information. We found support for our hypothesis that the spatial relationship is significantly correlated with the action. Furthermore, we could show that it is possible to predict the expected action based on our spatial features whereby we found that the distance measures are the most influential values. Additionally, we could show that it is possible to precisely predict the action, by using a Support Vector Machine. Thus, it becomes apparent that the variables are not clearly linear correlated or the noise in the data affects the linear regression model. This has to be investigated in further research.

The results of the study at hand are in line with our previous findings (see 4.3. on page 64). In our video-based experiment, the participants rated the stop and wait behavior as most legible which is also performed mostly in this study. Moreover, the distance d_r where most of the people stopped the robot is within the social space and very similar to the distance we identified as the most legible distance for a stop in our previous research.

The overall navigation strategy, which we found in our data, is similar to the behavior Basili et al. [8] found in their human-human path crossing experiment. In both studies, the participants were going (driving the robot) straight towards the goal and show a collision avoiding strategy by manipulating the speed. In our case by stopping and in their case by decreasing the speed.

However, the findings of the study at hand are significant, but the correlation values are rather low. The reason can be the rather uncontrolled study design, which causes noise in our data. Sometimes a stop was caused by loosing the connection to the robot. Furthermore, in the study at hand the participants had only a third person view from a fixed point in the room. This fact makes it difficult for the participants exactly to estimate distances, which can also be the reason for the rather low correlation values. Another limitation of the study is the small set of crossing angles and the missing of a frontal approach.

The aforementioned limitations can be the basis for further investigations of robot navigation behavior. For example, by implementing more controlled experiments, based on our findings, one can avoid the noise and find more accurate thresholds for the distance values. Also the first person view by using a camera on the robot can cause more precise results. Furthermore, it could be useful to evaluate the identified robot motion patterns and test how they are perceived by a human in order to verify our hypothesis that we can find out human expectations about robot behavior by using our "Inverse Wizard of Oz (I-WoZ)" study design. This can be done by doing it the other way around in an experiment where the behavior of the robot is scripted and the participants were asked to rate the behavior.

5.7. Conclusion

To sum up, we conducted a study to identify robot behavior patterns in a human-robot path crossing scenario. The overall navigation strategy we can conclude from the data is to drive straight towards the goal and only react (stop) to a crossing human when the stimulus based on the spatial relationship predicts to stop, otherwise drive on towards the goal. The expected action can be predicted by using a standard machine learning method like an SVM trained on our dataset. Based on these findings and by using our SVM model, we can develop a social navigation method, which fulfills human expectations about robot navigation behavior.

6. Discussion

After giving a summary, we relate our findings in the wider context of HRI research and discuss the differences and similarities of our investigations with the insights of other researchers. We point out the limitations of our work and make suggestions for future research by presenting open questions. Finally, we conclude by pointing out considerable factors for legible robot navigation and by drawing final conclusions regarding legibility of robot navigation in path crossing scenarios.

6.1. Summary

The work at hand investigates legibility of robot navigation behavior in human-robot path crossing situations. However, in the first two chapters (chapter 2 and 3) we consider the full range of possible robot behaviors to see the big picture of legibility research. We conducted a comprehensive literature research regarding legibility, where we answered the questions of how we can define, measure, and realize legibility, as well as which HRI properties are correlated. We concluded our findings and proposed a general definition for legible robot behavior. Based on the literature research, our own experiences, and further literature we propose evaluation methods to measure legibility of arbitrary robot behavior. After that, we present the setups and results of our own conducted experiments. We investigated how legible state-of-the-art navigation algorithms are in a human-robot path crossing scenario. Simultaneously, we evaluated the correlation of legibility with the HRI properties safety, comfort, and reliability. We found that all tested navigation methods scored rather low regarding legibility and we found statistically significant associations between legibility, safety, comfort, and reliability. Afterwards, we conducted an experiment with a similar setup with the objective to find the most legible navigation behavior pattern in a human-robot path crossing scenario. We found that moving sideways for a head-on encounter and stopping and let the human pass for a side crossing at a distance of 1.3 m to the crossing point is perceived as most legible. Furthermore, we showed a statistically significant correlation to the HRI property likability. After these series of controlled experiments, we conducted an explorative real-life study to capture human expectations about robot navigation behavior in arbitrary path crossing situations. We could verify our former result of the stopping and "let the human pass" behavior and additionally found that also passing behind or before the human is also expected. We could show that the expected behavior depends on the spatial relationship of human and robot with regard to the crossing point.

6.2. Discussion – Relations to Previous Research

First of all, we discuss the differences and similarities of the thesis at hand regarding the work of Dragan et al. [31, 32] who also using the term legibility. The main finding of their work is that legibility and predictability are different and contradictory properties of motion, which seems to be very different to our findings. They also developed methods to either generate goal directed motion trajectories optimizing either legibility or predictability. According to Dragan et al. [31,32] a legible goal directed motion is a sweeping motion and a predictable motion is straightforward (see Figure Fig. 49) which is also contrary to our results for robot navigation in path crossing scenario's. Our investigations revealed that a straight motion is most legible. However, by taking a closer look at their definitions and

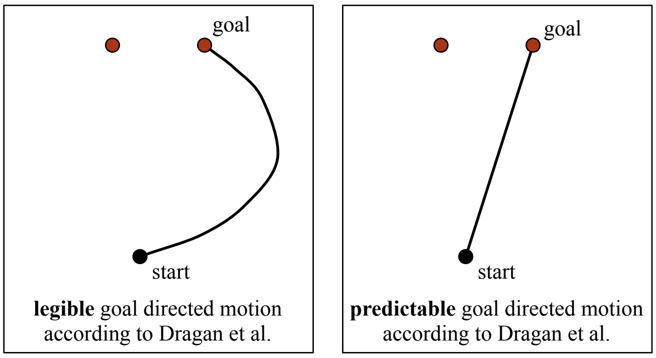


Fig. 49: Legible and predictable goal directed motion according to Dragan et al. [31, 32]

experimental methods we see that we use the same term but we investigate different things and have, therefore, different results. The first and also main difference lies in our different definition of the terms legibility and predictability. In contrast to our own definition of legibility (Def. 2 on page 15) which is in line with [2, 11, 23, 61, 64, 67, 96] Dragan et al. [31] proposes another formalization of the term legibility. They define legibility of goal directed motions as follows:

Legible motion is motion that enables an observer to quickly and confidently infer the correct goal G [31].

For the sake of clarity, we denote legibility according to Dragan et al. [31] as dragan-legibility. Their definition of legibility is very similar to our legibility factor predictability (see 3.2. on page 30) and in particular to goal-predictability. The way they measure their dragan-legibility by showing the participant's videos of different goal-directed arm motions and asking them to stop the video as soon as they are sure to predict the goal is similar to the methods we suggested using to measure goal-predictability. Consequently, we can conclude that we can compare our legibility factor predictability with their results regarding dragan-legibility. Furthermore, Dragan et al. [31] defines the motion property predictability.

Predictable motion is motion that matches what an observer would expect, given the goal G.

In order to avoid any possibility of confusion we will denote predictability according to Dragan et al. [31] as dragan-predictability. By looking at the definition dragan-predictability seems to be somehow equal to our legibility factor expectation-fulfillment, but taking into account the used experimental method we see more similarities to our predictability of motion trajectory (see 3.4.1. on page 32), in short trajectory-predictability. As we now have cleared the differences in our definitions we can compare our experimental results. We conducted only one experiment where we investigated goal-predictability as well as trajectory predictability. We did not found that goal-predictability (dragan-legibility) and trajectory-predictability (dragan-predictability) are contradictory properties of motion. A correlation analysis revealed quite the opposite. We found a weak but significant correlation between the two properties [72]. We explain these contrary results with the fact that Dragan et al. [31] evaluated motion trajectories that are either dragan-legibility or dragan-predictability optimized, which obviously increases either the one or the other property. Our motions were generated by state-of-the-art navigation methods generating more straight and, therefore, dragan-predictable motions. Furthermore, we want to point out that we investigate a different scenario as Dragan et al. [31].

We look at robot navigation in human-robot path crossing scenarios whereas Dragan et al. [31] looks specific at goal directed arm motions without any human interaction. All these differences in the experimental setup can also be an explanation for our different results. In our further experiments we only evaluated trajectory-predictability (= dragan-predictability) and our results are in line with the results of Dragan et al. [31] that a straight motion is highly predictable (see 4.3. on page 64 and 5.4. on page 82). We did not especially investigate dragan-legible motions in our experiments. Therefore, we cannot directly compare our results. To conclude, our research regarding legibility seems to be similar to the work of Dragan et al. [31], but when going into details we see that we are defining and investigating very different things and have, therefore, partly different results.

One important finding of our research is that most of the participants want the robot to react in a path crossing situation in order to prevent a collision. This result is in line with a finding of Dautenhahn et al. [27] regarding robot movements. In their experiment, most of the participants wanted the robot to be polite and give way to them. This describes the behavior we found as most legible in path crossing situations.

The results of our experiments to determine legible navigation behavior in human-robot path crossings (see Section 4.3. on page 64 and 5.4. on page 82) are in line with results of a conducted human-human path crossing experiment [9]. The similarity of the results supports the hypothesis of [11,45,64] which human-like behavior is legible. Moreover, Kruse et al. [65] implemented and evaluated a human-like path crossing behavior and found that the human-like behavior is perceived as less confusing, which is opposite to legibility. To conclude, our results regarding legible robot navigation behavior in path crossing scenarios combined with the results in [9, 65] indicates that human-like behavior is highly legible.

By investigating the proxemics of human-robot path crossing situations we found similar results to Basili et al. [9] in a human-human path crossing experiment. The participants in the experiment conducted by Basili et al. [9] decreased their velocity to let the other human pass at a distance of 1.0 m to the crossing that is similar to our findings where participants stopped the robot at a median distance of 1.25 m (see Section 5.4. on page 82). Furthermore, our results regarding the proxemic behavior in path crossing situations supports most of the social navigation approaches which are avoiding to enter the personal space of a human [52, 58, 70].

With our investigations regarding correlated HRI properties we verify the assumptions of [61,96,101] that legibility is correlated with safety and comfort. Furthermore, our results regarding the correlation of legibility and safety (see Section 4.2.3. on page 57) are conform with results from Dehais et al. [28] who evaluated legibility and safety in hand-over tasks.

Finally, we can say that our research results are conform with results of other researchers in the HRI field and extend current knowledge regarding legible human-robot path crossing behavior.

6.3. Limitations

Despite the limitations that we already mentioned referring specifically to each of the conducted experiments we point one overall limitation of the work at hand.

Besides our main objective, to investigate legibility of robot navigation behavior, we also investigated the proxemics of path crossings. For example, we look at distances where the robot has to stop in a side crossing situation or how the human has perceived the path crossing interaction. We know from a variety of other studies [84, 107, 112] that proxemic is highly influenced by other factors in particular in the height and appearance [63] of the robot. In our experiments, we kept most of these things constant. In almost all legibility experiments we used the same B21 robot, due to the lack of available other robots. The influence of appearance on legibility is also to be investigated further.

6.4. Open Questions - Further Research

Within the work at hand, several open questions came up. In the following we want to repeat, conclude, and briefly discuss the questions and ideas for future research in order to set a starting point for further research regarding legibility.

Legibility Metrics In chapter 3. on page 29 we developed metrics to measure legibility. As we already stated (see 3.2. on page 31), the list of measurements and evaluation methods is meant as a first step towards legibility metrics and has to be extended and verified in the future.

Why is the expectation-fulfillment factor higher than predictability? In our experiments, we found that the expectation of the participants was met with a higher extent as the behavior was predicted correctly and the correlation to other HRI properties were stronger for expectation-fulfillment than for legibility. This let us assume that humans expect different kinds of behaviors with different probabilities and different values like likeliness or safety. Alternatively, it is one abstract behavior like "avoid" for a frontal approach crossing situation and the way the robot performs this behavior influences other properties, which has already been investigated by other researchers [23, 58, 78, 88]. We suggest repeating one of our experiments or design a similar one with one adjustment. Instead of letting the participants predict one distinct behavior one can provide different kinds of possible behaviors and let the participant rate how much he/she would expect the behavior.

Which other HRI properties are correlated with legibility? Correlated HRI properties of legibility should be further investigated. Not only the already investigated factors like safety, comfort and efficiency need further considerations, there is also a need to look at other HRI properties like perceived intelligence (see [7]) or indicators of robot acceptance like intention to use, usefulness or enjoyment (see [51]).

Influences on Path Crossing Proxemics As mentioned in the former section we kept some factors like the robot appearance constant, which are influencing human proxemic behavior [107, 112]. It would be very interesting to investigate how the robot appearance influences the legibility in path crossings. We suggest repeating our experiments using different kinds of robots with different heights and appearances.

6.5. Factors for Legible Robot Navigation

In the following, we conclude our findings regarding legible robot behavior in human-robot path crossing scenarios in order to draw some key factors for generating legible robot navigation behavior.

Straight Towards the Goal Based on our research results we concluded one main factor for legible robot navigation behavior. A moving robot should always move as far as possible straight towards its goal and react as smoothly as possible to a human by stopping or moving smoothly to one side in a head-on encounter. Meaning the robot should move using mostly straight lines and does not corner sharply to avoid the human. This finding is in line with [15, 31]. Straight lines towards the goal are also fulfilling the efficiency criteria that we mentioned earlier in our review (Chapter 2) as one factor of legible motion. In addition, our simulator based experiment (Section 4.2.2. on page 46) showed that driving curves or spinning around leads to lower the legibility and confused the participants. Driving on mostly straight lines is also in line with the claim for human-like behavior. In a human-human path crossing experiment Basili et al. [9] found that humans do not swerve. They observed that the participants decreased velocity in order to avoid a collision.

Be Polite! Another crucial factor for legible navigation, which we can conclude from our experimental results, is generating a "polite" behavior. Meaning in a path crossing situation, the robot should react to let the human pass by or make way for the human coming towards the robot. This finding is in line with Dautenhahn et al. [27]. Humans expect a robot to be a polite and attentive assistant. To this end, politeness is one important factor to generate legible motions.

Consider the Spatial Relationship Taking into account the spatial relationship of robot and human in a path crossing scenario (see Section 1.2. on page 2) is crucial to determine the most legible behavior. Depending on the particular spatial relationship the robot should either stop or move on to pass the human's path before or behind the human. Humans do not only expect this behavior it is also efficient, which is correlated with legibility [15].

6.6. Conclusion

The investigated scenarios of the thesis at hand are human-robot path crossing situations. According to our results, we can finally conclude that legibility is a crucial factor for robot navigation. We have found significant correlations to the crucial HRI properties safety, comfort, reliability, and likability showing us the importance of legibility in HRI scenarios. We found coherent results in all our conducted experiments regarding legible robot navigation in path crossing scenarios. Our main finding is a navigation behavior that we identified as highly legible. In a head-on encounter, the robot should move to the side (left or right, depending on cultural norms) and make way for a human. In a side-way crossing, we identified two different behaviors depending on the spatial relationship. Either the robot should pass before or after the human or the robot should stop and let the human pass. All behaviors are polite and give priority to the human and his/her path. Additionally, we investigated the proxemics of a crossing situation. We can conclude that the robot should stay outside of the humans personal space. Furthermore, by using the point where the paths of the robot and human cross as a reference point we also found that it is highly legible when the robot reacts near the outside margin of the personal space to avoid the human.

A Experimental Material

In the following we present the questionnaires used in our experiments to evaluate legibility of robot navigation behavior as well as our demographical questionnaires. We used classical pen and paper questionnaires (Figure Fig. 50 on the following page, Fig. 51 on page 101, Fig. 52 on page 102), computer supported questionnaires (Figure Fig. 53 on page 103), and online survey and embedded videos in the website (Figure Fig. 54 on page 104, Fig. 55 on page 104, Fig. 56 on page 105, Fig. 57 on page 105).

Video:

1. Stop

Where should the robot bring the folder to?

○ table 1 **(with the green plant)**

○ table 2 **(with the blue vase)**

○ table 3 **(with the red lamp)**

○ other:..

Will the robot change its direction?

○ No

○ Yes In which direction do you think?.................................

How confident do you feel about your answer?

Not at all ① ② ③ ④ ⑤ Very

---

2. Stop

Did the robot behave like you expected?

○ Yes
○ No

If you answered no, how surprised have you been about the robots behavior??

Not at all ① ② ③ ④ ⑤ Very

If I had to interact with the robot...

I would feel not safe at all	1	2	3	4	5	I would feel very safe
I would not feel comfortable at all	1	2	3	4	5	I would feel very comfortable
I would perceive it as not reliable at all	1	2	3	4	5	I would perceive it as very reliable

Fig. 50: Questionnaire used to evaluate robot navigation behavior in a path crossing task. The participant sees a robot crossing the path of a human in an office enviroment (see sec:experiment1). We measure the predictability (question 1+2), the confidence about the prediction (question 3), the met-expectation factor (question 4), surprise (question 5), and the HRI factors safety, comfort, and reliability (question 6-8)

1. Video

Klicken Sie wenn Sie sich sicher sind wohin der Roboter als nächstes fahren wird!

Wohin wird der Roboter fahren?

- ○ nach rechts
- ○ nach links
- ○ geradeaus
- ○ rückwärts
- ○ bleibt stehen

Wie sicher sind Sie sich bei Ihrer Antwort?

Gar nicht ① ② ③ ④ ⑤ Sehr

2. Video

Hat der Roboter sich so verhalten wie Sie es erwartet haben?

Gar nicht ① ② ③ ④ ⑤ Sehr

Bitte bewerten Sie Ihren Gemütszustand auf diesen Skalen:

Ängstlich	①	②	③	④	⑤	Entspannt
Unruhig	①	②	③	④	⑤	Ruhig
Nicht Überrascht	①	②	③	④	⑤	Überrascht

Fig. 51: Questionnaire used to evaluate robot navigation behavior in a path crossing task. The participant sees a robot approaching (see sec:experiment2). We measure the predictability (question 1), the confidence about the prediction (question 2), and the perceived safety by using the Godspeed-V questionnaire

Erfassungsbogen VP_______ Datum:______________

Liebe/r Versuchsteilnehmer/in,

vielen Dank, dass Sie sich bereit erklärt haben, an unserer Studie teilzunehmen. Zunächst bräuchten wir von Ihnen noch einige Angaben. Wir versichern Ihnen, dass wir diese vertraulich behandeln und nicht weitergeben.

Alter:________

Geschlecht: ☐ männlich ☐ weiblich

Wie häufig sind Sie mit Robotern in Kontakt?

☐	☐	☐	☐
Regelmäßig	Ab und zu	Selten	Nie

Wie häufig spielen Sie Computer – Spiele bei welchen der Spieler aus der Ich-Perspektive agiert ? (Ego-Shooter)

☐	☐	☐	☐
Regelmäßig	Ab und zu	Selten	Nie

Das war‘s auch schon.

Vielen Dank für Ihr Vertrauen und viel Spaß beim Experiment.

Fig. 52: Demographical data and robot experience aquisition

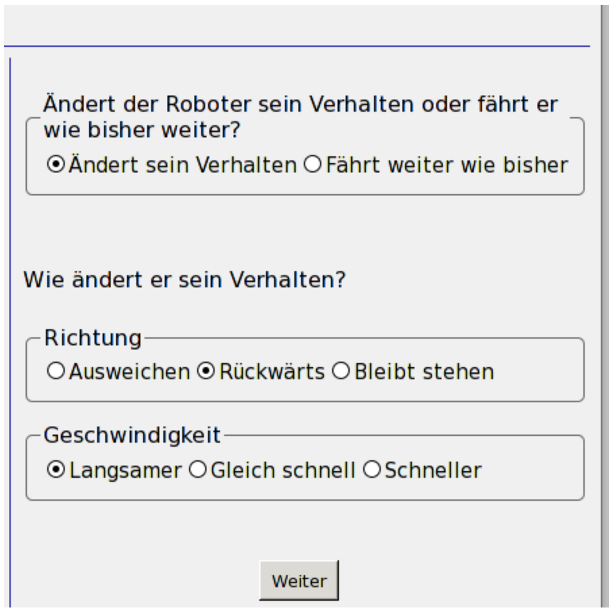


Fig. 53: Survey to investigate predictability

Fig. 54: Start screen of the online survey with explanations.

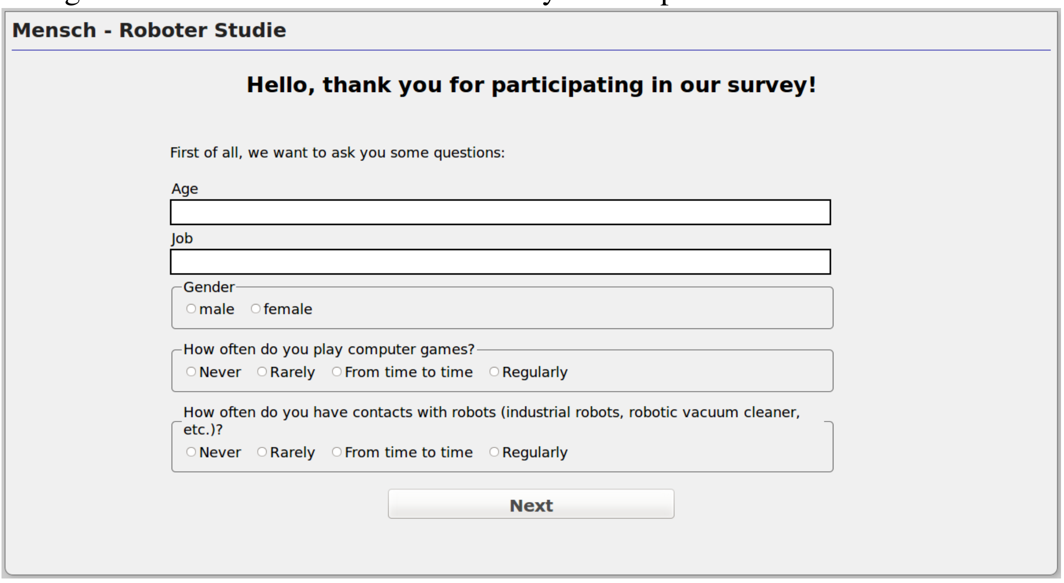


Fig. 55: Demographical questionnaire used in the online survey to investigate legible navigation behavior in a path crossing scenario. See Section 4.3. on page 64

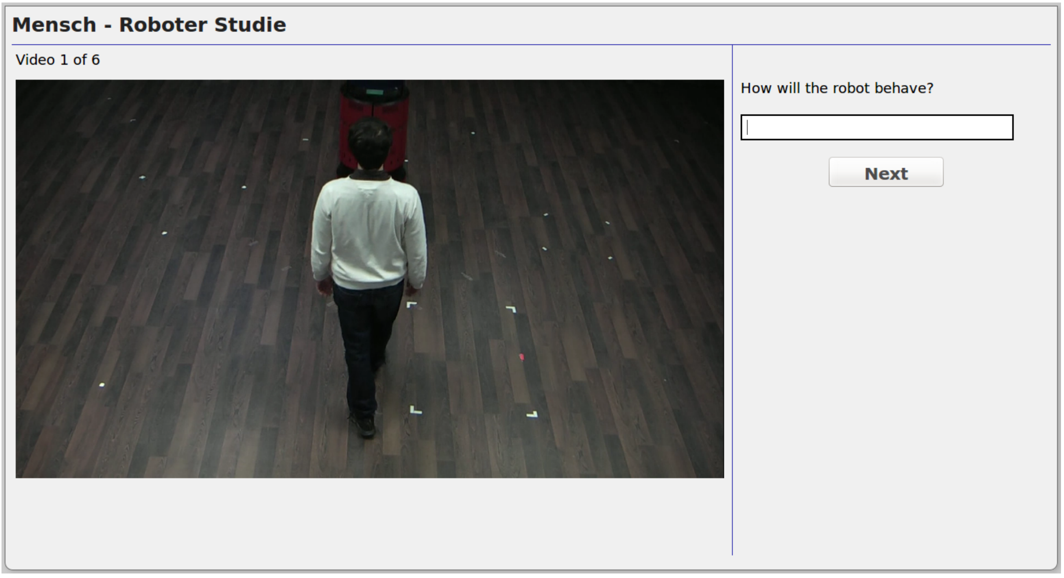


Fig. 56: Survey screen with video showing a robot crossing a human path. The first video shows the robot behavior up to a defined stopping point. After the video, the participants were asked to predict the future behavior of the robot.

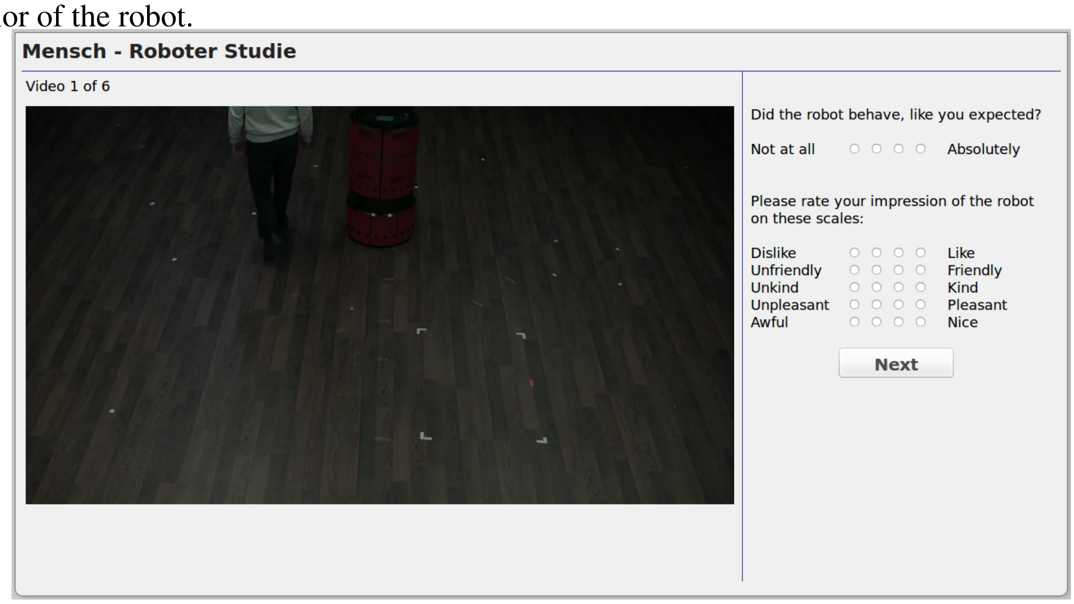


Fig. 57: After the participant has answered the first question, we showed the second part of the video. Afterwards, we asked the participants to rate how mich the robot behavior met their expectations. Furthermore, we used the Godspeed questionnaire [7] to measure the likability of the robot behavior.

Bibliography

[1] Rachid Alami, Aurélie Clodic, Vincent Montreuil, Emrah Akin Sisbot, and Raja Chatila. Task planning for human-robot interaction. In *ACM Conference on Smart Objects and Ambient Intelligence: Innovative Context-Aware Services: Usages and Technologies*, 2005.

[2] Rachid Alami, Aurélie Clodic, Vincent Montreuil, Emrah Akin Sisbot, and Raja Chatila. Toward Human-Aware Robot Task Planning. In *AAAI Spring Symposium: To Boldly Go Where No Human-Robot Team Has Gone Before*, 2006.

[3] Samir Alili, Matthieu Warnier, Muhammad Ali, and Rachid Alami. Planning and Plan-execution for Human-Robot Cooperative Task achievement. In *International Conference on Automated Planning and Scheduling*, 2009.

[4] Philipp Althaus, H. Ishiguro, T. Kanda, T. Miyashita, and H.I. Christensen. Navigation for Human-Robot Interaction Tasks. In *IEEE International Conference on Robotics and Automation*, 2004.

[5] Ronald C. Arkin. *Behavior-Based Robotics*. MIT press, 1998.

[6] Marian Stewart Bartlett, G. Littlewort, I. Fasel, J. Chenu, H. Ishiguro, and J.R. Movellan. Towards Social Robots: Automatic Evaluation of Human-Robot Interaction by Face Detection and Expression Classification. *Advances in Neural Information Processing Systems*, 16, 2003.

[7] Christoph Bartneck, Dana Kulić, Elizabeth Croft, and Susana Zoghbi. Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots. *International Journal of Social Robotics*, 1(1), 2008.

[8] Patrizia Basili, Markus Huber, Omiros Kourakos, Tamara Lorenz, Thomas Brandt, Sandra Hirche, and Stefan Glasauer. Inferring the Goal of an Approaching Agent: a Human-Robot Study. In *IEEE International Workshop on Robot and Human Interactive Communication*, 2012.

[9] Patrizia Basili, Murat Sag, Thibault Kruse, Markus Huber, Alexandra Kirsch, and Stefan Glasauer. Strategies of Locomotor Collision Avoidance. *Gait & Posture*, 2012.

[10] Heidrun Becker, Mandy Scheermesser, Michael Früh, Yvonne Treusch, Holger Auerbach, and Richard Alexander Hüppi. *Robotik in Betreuung und Gesundheitsversorgung*, volume 58. vdf Hochschulverlag AG, 2013.

[11] Michael Beetz, Freek Stulp, Piotr Esden-Tempski, Andreas Fedrizzi, Ulrich Klank, Ingo Kresse, Alexis Maldonado, and Federico Ruiz. Generality and Legibility in Mobile Manipulation. *Autonomous Robots*, 28(1), 2010.

[12] Cindy Bethel, Jennifer Burke, Robin Murphy, and Kristen Salomon. Psychophysiological Experimental Design for Use in Human-Robot Interaction Studies. In *IEEE International Symposium on Collaborative Technologies and Systems*, 2007.

[13] Cindy Bethel and Robin Murphy. Review of Human Studies Methods in HRI and Recommendations. *International Journal of Social Robotics*, 2(4), 2010.

[14] Cindy L. Bethel, Kristen Salomon, Robin R. Murphy, and Jennifer L. Burke. Survey of Psychophysiology Measurements Applied to Human-Robot Interaction. In *IEEE International Symposium on Robot and Human interactive Communication*, 2007.

[15] Dino Bortot, Maximilian Born, and Klaus Bengler. Directly or on detours? How should industrial robots approximate humans? In *ACM/IEEE International Conference on Human-Robot Interaction*, 2013.

[16] Margaret Bradley, Laura Miccoli, Miguel Escrig, and Peter Lang. The Pupil as a Measure of Emotional Arousal and Autonomic Activation. *Psychophysiology*, 45(4), 2008.

[17] Catherina R. Burghart and Aaron Steinfeld. Human-Robot Interaction Metrics and Future Directions. *Metrics for Human-Robot Interaction, a Workshop at ACM/IEEE International Conference on Human Robot Interaction*, 2008.

[18] John Travis Butler and Arvin Agah. Psychological Effects of Behavior Patterns of a Mobile Personal Robot. *Autonomous Robots*, 2001.

[19] Chih-Chung Chang and Chih-Jen Lin. LIBSVM: A Library for Support Vector Machines. *ACM Transactions on Intelligent Systems and Technology*, 2, 2011. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

[20] Aurélie Clodic, Sara Fleury, Rachid Alami, Matthieu Herrb, and Raja Chatila. Supervision and Interaction. In *IEEE International Conference on Advanced Robotics*, 2005.

[21] Aurélie Clodic, Vincent Montreuil, Rachid Alami, and Raja Chatila. A Decisional Framework for Autonomous Robots Interacting with Humans. In *IEEE International Workshop on Robot and Human Interactive Communication*, 2005.

[22] Nello Cristianini and John Shawe-Taylor. *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge university press, 2000.

[23] K. Dautenhahn, M. Walters, S. Woods, K.L. Koay, C.L. Nehaniv, A. Sisbot, R. Alami, and T. Siméon. How May I Serve You? A Robot Companion Approaching a Seated Person in a Helping Context. In *ACM SIGCHI/SIGART Conference on Human-Robot Interaction*, 2006.

[24] Kerstin Dautenhahn. Robots We Like to Live With?– A Developmental Perspective on a Personalized, Life-Long Robot Companion. In *IEEE International Workshop on Robot and Human Interactive Communication*, 2004.

[25] Kerstin Dautenhahn. Methodology and Themes of Human-Robot Interaction: A Growing Research Field. *International Journal of Advanced Robotic Systems*, 2007.

[26] Kerstin Dautenhahn and Iain Werry. A Quantitative Technique for Analysing Robot-Human Interactions. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2002.

[27] Kerstin Dautenhahn, Sarah Woods, Christina Kaouri, Michael L. Walters, Kheng Lee Koay, and Iain Werry. What is a Robot Companion-Friend, Assistant or Butler? In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2005.

[28] Frédéric Dehais, Emrah Akin Sisbot, Rachid Alami, and Mickaël Causse. Physiological and Subjective Evaluation of a Human-Robot Object Hand-Over Task. *Applied Ergonomics*, 2011.

[29] Christian Dondrup, Christina Lichtenthäler, and Marc Hanheide. Hesitation Signals in Human-Robot Head-on Encounters. In *ACM/IEEE International Conference on Human Robot Interaction*, 2014.

[30] Leonard William Doob. *Hesitation: Impulsivity and Reflection*. Greenwood Press, 1990.

[31] Anca Dragan, Kenton Lee, and Siddhartha S Srinivasa. Legibility and Predictability of Robot Motion. In *ACM/IEEE International Conference on Human-Robot Interaction*, 2013.

[32] Anca Dragan and Siddhartha Srinivasa. Generating Legible Motion. In *Proceedings of Robotics: Science and Systems*, 2013.

[33] G. Echeverria, N. Lassabe, A. Degroote, and S. Lemaignan. Modular Open Robots Simulation Engine: MORSE. In *IEEE International Conference on Robotics and Automation*, 2011.

[34] Friederike Eyssel, Dieta Kuchenbrandt, and Simon Bobinger. Effects of Anticipated Human-Robot Interaction and Predictability of Robot Behavior on Perceptions of Anthropomorphism. In *ACM/IEEE International Conference on Human-Robot Interaction*, 2011.

[35] Julia Fink, Séverin Lemaignan, Pierre Dillenbourg, Philippe Rétornaz, Florian Vaussard, Alain Berthoud, Francesco Mondada, Florian Wille, and Karmen Franinovic. Which Robot Behavior Can Motivate Children to Tidy up Their Toys? Design and Evaluation of "Ranger". In *ACM/IEEE International Conference on Human-Robot Interaction*, 2014.

[36] Mary Ellen Foster, Andre Gaschler, Manuel Giuliani, Amy Isard, Maria Pateraki, and Ronald Petrick. Two People Walk into a Bar: Dynamic Multi-Party Social Interaction with a Robot Agent. In *ACM International Conference on Multimodal Interaction*, 2012.

[37] D. Fox, W. Burgard, and S. Thrun. The Dynamic Window Approach to Collision Avoidance. *Robotics & Automation Magazine*, 4(1), 1997.

[38] Andre Gaschler, Sören Jentzsch, Manuel Giuliani, Kerstin Huth, Jan de Ruiter, and Alois Knoll. Social Behavior Recognition Using Body Posture and Head Pose for Human-Robot Interaction. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2012.

[39] B.P. Gerkey and Kurt Konolige. Planning and Control in Unstructured Terrain. In *ICRA Workshop on Path Planning on Costmaps*, 2008.

[40] Michael J Gielniak and Andrea Lockerd Thomaz. Generating Anticipation in Robot Motion. In *IEEE International Workshop on Robot and Human Interactive Communication*, 2011.

[41] Manuel Giuliani and Alois Knoll. Evaluating Supportive and Instructive Robot Roles in Human-Robot Interaction. In *International Conference on Social Robotics*, 2011.

[42] Barbara Gonsior, Stefan Sosnowski, Christoph Mayer, Jürgen Blume, Bernd Radig, Dirk Wollherr, and K Kuhnlenz. Improving Aspects of Empathy and Subjective Performance for HRI through Mirroring Facial Expressions. In *IEEE International Symposium on Robot and Human Interactive Communication*, 2011.

[43] Michael A Goodrich and Alan C Schultz. Human-Robot Interaction: A Survey. *Foundations and trends in human-computer interaction*, 1(3), 2007.

[44] Anders Green, Helge Hüttenrauch, and K Severinson Eklundh. Applying the Wizard-of-Oz Framework to Cooperative Service Discovery and Configuration. In *IEEE International Workshop on Robot and Human Interactive Communication*, 2004.

[45] Jérôme Guzzi, Alessandro Giusti, Luca Gambardella, Guy Theraulaz, and Gianni Di Caro. Human-friendly Robot Navigation in Dynamic Environments. In *IEEE International Conference on Robotics and Automation*, 2013.

[46] A. Haasch, S. Hohenner, S. Hüwel, M. Kleinehagenbrock, S. Lang, I. Toptsis, G.A. Fink, J. Fritsch, B. Wrede, and G. Sagerer. BIRON - The Bielefeld Robot Companion. In *International Workshop on Advances in Service Robotics*, 2004.

[47] Martin Hägele, Nikolaus Blümlein, and Oliver Kleine. Wirtschaftlichkeitsanalysen neuartiger Servicerobotik-Anwendungen und ihre Bedeutung für die Robotik-Entwicklung. *Eine Analyse der Fraunhofer Institute IPA und ISI im Auftrag des BMBF, Fraunhofer Gesellschaft*, 2011.

[48] Edward T. Hall. *The Hidden Dimension*. Anchor Books New York, 1969.

[49] Marc Hanheide, Annika Peters, and Nicola Bellotto. Analysis of Human-Robot Spatial Behaviour Applying a Qualitative Trajectory Calculus. In *IEEE International Workshop on Robot and Human Interactive Communication*, 2012.

[50] Sandra Hart and Lowell Staveland. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Human Mental Workload*, 1(3), 1988.

[51] Marcel Heerink, Krose, Vanessa Evers, and Bob Wielinga. Measuring Acceptance of an Assistive Social Robot: A Suggested Toolkit. In *IEEE International Symposium on Robot and Human Interactive Communication*, 2009.

[52] Kuo-Chen Huang, Jiun-Yi Li, and Li-Chen Fu. Human-Oriented Navigation for Service Providing in Home Environment. In *IEEE SICE Annual Conference 2010*, 2010.

[53] Markus Huber, Markus Rickert, Alois Knoll, Thomas Brandt, and Stefan Glasauer. Human-Robot Interaction in Handing-Over Tasks. In *IEEE International Symposium on Robot and Human Interactive Communication*, 2008.

[54] Helge Hüttenrauch, K Severinson Eklundh, Anders Green, and Elin Anna Topp. Investigating Spatial Relationships in Human-Robot Interaction. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2006.

[55] Takayuki Kanda, Michihiro Shimada, and Satoshi Koizumi. Children Learning with a Social Robot. In *ACM/IEEE International Conference on Human-Robot Interaction*, 2012.

[56] Nikos Katevas. *Mobile Robotics in Healthcare*, volume 7. IOS Press, 2001.

[57] Adam Kendon. *Conducting Interaction: Patterns of Behavior in Focused Encounters*, volume 7. CUP Archive, 1990.

[58] Rachel Kirby, Reid Simmons, and Jodi Forlizzi. COMPANION: A Constraint-Optimizing Method for Person-Acceptable Navigation. In *IEEE International Symposium on Robot and Human Interactive Communication*, 2009.

[59] Alexandra Kirsch. *Integration of Programming and Learning in a Control Language for Autonomous Robots Performing Everyday Activities*. PhD thesis, Technische Universität München, 2008.

[60] Alexandra Kirsch, Thibault Kruse, and Lorenz Mösenlechner. An Integrated Planning and Learning Framework for Human-Robot Interaction. In *Workshop on Planning and Plan Execution for Real-World Systems*, 2009.

[61] Alexandra Kirsch, Thibault Kruse, E.Akin Sisbot, Rachid Alami, Martin Lawitzky, Dražen Brščić, Sandra Hirche, Patrizia Basili, and Stefan Glasauer. Plan-Based Control of Joint Human-Robot Activities. *KI - Künstliche Intelligenz*, 24(3), 2010.

[62] Kheng Lee Koay, Emrah Akin Sisbot, Dag Sverre Syrdal, Mick L Walters, Kerstin Dautenhahn, and Rachid Alami. Exploratory Study of a Robot Approaching a Person in the Context of Handing Over an Object. In *AAAI Spring Symposium: Multidisciplinary Collaboration for Socially Assistive Robotics*, 2007.

[63] Kheng Lee Koay, Dag Sverre Syrdal, Michael L. Walters, and Kerstin Dautenhahn. Living with Robots: Investigating the Habituation Effect in Participants' Preferences During a Longitudinal Human-Robot Interaction Study. In *IEEE International Symposium on Robot and Human interactive Communication*, 2007.

[64] Thibault Kruse, Patrizia Basili, Stefan Glasauer, and Alexandra Kirsch. Legible Robot Navigation in the Proximity of Moving Humans. In *IEEE Workshop on Advanced Robotics and its Social Impacts*, 2012.

[65] Thibault Kruse, Alexandra Kirsch, Harmish Khambhaita, and Rachid Alami. Evaluating Directional Cost Models in Navigation. In *ACM/IEEE International Conference on Human-Robot Interaction*, 2014.

[66] Thibault Kruse, Alexandra Kirsch, E. Akin Sisbot, and Rachid Alami. Exploiting Human Cooperation in Human-Centered Robot Navigation. In *IEEE International Symposium in Robot and Human Interactive Communication*, 2010.

[67] Thibault Kruse, Alexandra Kirsch, Emrah Akin Sisbot, and Rachid Alami. Exploiting Human Cooperation in Human-Centered Robot Navigation. In *IEEE International Workshop on Robot and Human Interactive Communication*, 2010.

[68] Thibault Kruse, Amit Kumar Pandey, Rachid Alami, and Alexandra Kirsch. Human-Aware Robot Navigation: A Survey. *Robotics and Autonomous Systems*, 2013.

[69] Dana Kulic and Elizabeth Croft. Anxiety Detection During Human-Robot Interaction. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2005.

[70] Chi-Pang Lam, Chen-Tun Chou, Chih-Fu Chang, and Li-Chen Fu. Human-Centered Robot Navigation - Toward a Harmoniously Coexisting Multi-Human and Multi-Robot Environment. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2010.

[71] Christian Lang, Sven Wachsmuth, Marc Hanheide, and Heiko Wersing. Facial communicative signals. *International Journal of Social Robotics*, 4(3), 2012.

[72] Christina Lichtenthäler and Alexandra Kirsch. Goal-Predictability vs . Trajectory-Predictability – Which Legibility Factor Counts. In *ACM/IEEE International Conference on Human Robot Interaction*, 2014.

[73] Christina Lichtenthäler, Tamara Lorenz, Michael Karg, and Alexandra Kirsch. Increasing Perceived Value Between Human and Robots—Measuring Legibility in Human Aware Navigation. In *IEEE Workshop on Advanced Robotics and its Social Impacts*, 2012.

[74] Christina Lichtenthäler, Tamara Lorenz, and Alexandra Kirsch. Towards a Legibility Metric: How to Measure the Perceived Value of a Robot. *International Conference on Social Robotics, Work-In-Progress-Track*, 2011.

[75] Christina Lichtenthäler, Tamara Lorenz, and Alexandra Kirsch. Influence of Legibility on Perceived Safety in a Virtual Human-Robot Path Crossing Task. In *IEEE International Symposium on Robot and Human Interactive Communication*, 2012.

[76] Christina Lichtenthäler, Annika Peters, Sascha Griffiths, and Alexandra Kirsch. Social Navigation - Identifying Robot Navigation Patterns in a Path Crossing Scenario. In *International Conference on Social Robotics*, 2013.

[77] Sebastian Loth, Manuel Giuliani, and Jan P. de Ruiter. Ghost-in-the-Machine: Initial Results. In *ACM/IEEE International Conference on Human-Robot Interaction*, 2014.

[78] David V. Lu and William D. Smart. Towards more Efficient Navigation for Robots and Humans. In *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2013.

[79] Jim Mainprice, E. Akin Sisbot, Léonard Jaillet, Juan Cortés, Rachid Alami, and Thierry Siméon. Planning Human-Aware Motions Using a Sampling-Based Costmap Planner. In *IEEE International Conference on Robotics and Automation*, 2011.

[80] Jim Mainprice, E. Akin Sisbot, Thierry Siméon, and Rachid Alami. Planning Safe and Legible Hand-Over Motions for Human-Robot Interaction. In *IARP Workshop on Technical Challenges for Dependable Robots in Human Environments*, 2010.

[81] Eitan Marder-Eppstein, Eric Berger, Tully Foote, Brian Gerkey, and Kurt Konolige. The Office Marathon: Robust Navigation in an Indoor Office Environment. In *IEEE International Conference on Robotics and Automation*, 2010.

[82] AJung Moon, Chris Parker, Elizabeth Croft, and HF Van der Loos. Design and Impact of Hesitation Gestures during Human-Robot Resource Conflicts. *Journal of Human-Robot Interaction*, 2(3), 2013.

[83] Lilia Moshkina and Ronald C. Arkin. Human Perspective on Affective Robotic Behavior: A Longitudinal Study. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2005.

[84] Jonathan Mumm and Bilge Mutlu. Human-Robot Proxemics: Physical and Psychological Distancing in Human-Robot Interaction. In *ACM/IEEE International Conference on Human-Robot Interaction*, 2011.

[85] Dana Nau, Tsz-Chiu Au, Okhtay Ilghami, Ugur Kuter, J William Murdock, Dan Wu, and Fusun Yaman. SHOP2: An HTN planning system. *Journal Artificial Intelligence Research*, 20, 2003.

[86] Chrystopher L. Nehaniv, Kerstin Dautenhahn, Jens Kubacki, Martin Haegele, Christopher Parlitz, and Rachid Alami. A Methodological Approach Relating the Classification of Gesture to Identification of Human Intent in the Context of Human-Robot Interaction. In *IEEE International Workshop on Robot and Human Interactive Communication*, 2005.

[87] Don Norman. *The Design of Future Things*. Basic Books, 2009.

[88] Elena Pacchierotti, Henrik I. Christensen, and Patric Jensfelt. Human-Robot Embodied Interaction in Hallway Settings: a Pilot User Study. In *IEEE International Workshop on Robot and Human Interactive Communication*, 2005.

[89] Elena Pacchierotti, Henrik I. Christensen, and Patric Jensfelt. Evaluation of Passing Distance for Social Robots. In *IEEE International Symposium on Robot and Human Interactive Communication*, 2006.

[90] Annika Peters, Thorsten P. Spexard, Marc Hanheide, Petra Weiss, et al. Hey Robot, Get out of my Way: Survey on a Spatial and Situational Movement Concept in HRI. 2011.

[91] Karola Pitsch, Bernhard Brüning, Christian Schnier, and Holger Dierker. Linking Conversation Analysis and Motion Capturing : How to robustly track multiple participants ? In *Multimodal Corpora: Advances in Capturing, Coding and Analyzing Multimodality*, 2010.

[92] Kerstin Preuschoff, Bernard Marius't Hart, and Wolfgang Einhäuser. Pupil Dilation Signals Surprise: Evidence for Noradrenaline's Role in Decision Making. *Frontiers in Neuroscience*, 5, 2011.

[93] Pramila Rani, Jared Sims, Robert Brackin, and Nilanjan Sarkar. Online Stress Detection Using Psychophysiological Signals for Implicit Human-Robot Cooperation. *Robotica*, 20(06), 2002.

[94] Satoru Satake, Takayuki Kanda, Dylan F Glas, Michita Imai, Hiroshi Ishiguro, and Norihiro Hagita. How to Approach Humans? –Strategies for Social Robots to Initiate Interaction. In *ACM/IEEE International Conference on Human-Robot Interaction*, 2009.

[95] C.M. Schulz, E. Schneider, L. Fritz, J. Vockeroth, A. Hapfelmeier, M. Wasmaier, E.F. Kochs, and G. Schneider. Eye Tracking for Assessment of Workload: a Pilot Study in an Anaesthesia Simulator Environment. *British journal of anaesthesia*, 106(1), 2011.

[96] Emrah Akin Sisbot and Rachid Alami. A Human-Aware Manipulation Planner. *IEEE Transactions on Robotics*, 28(5), 2012.

[97] Emrah Akin Sisbot, Rachid Alami, Thierry Siméon, Kerstin Dautenhahn, Michael Walters, and Sarah Woods. Navigation in the Presence of Humans. In *IEEE-RAS International Conference on Humanoid Robots*, 2005.

[98] Emrah Akin Sisbot, Aurélie Clodic, Rachid Alami, and Maxime Ransan. Supervision and Motion Planning for a Mobile Manipulator Interacting with Humans. In *ACM/IEEE International Conference on Human-Robot Interaction*, 2008.

[99] Emrah Akin Sisbot, Luis F Marin, and Rachid Alami. Spatial Reasoning for Human Robot Interaction. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2007.

[100] Emrah Akin Sisbot, Luis F. Marin-Urias, Rachid Alami, and Thierry Simeon. A Human Aware Mobile Robot Motion Planner. *IEEE Transactions on Robotics*, 23, 2007.

[101] Emrah Akin Sisbot, Luis F. Marin-Urias, Xavier Broquere, Daniel Sidobre, and Rachid Alami. Synthesizing Robot Motions Adapted to Human Presence. *International Journal of Social Robotics*, 2(3), 2010.

[102] Charles Spielberger. *State-Trait Anxiety Inventory*. John Wiley & Sons, Inc., 2010.

[103] Aaron Steinfeld, Terrence Fong, David Kaber, Michael Lewis, Jean Scholtz, Alan Schultz, and Michael Goodrich. Common Metrics for Human-Robot Interaction. In *ACM Conference on Human-Robot Interaction*, 2006.

[104] Aaron Steinfeld, Odest C. Jenkins, and Brian Scassellati. The Oz of Wizard: Simulating the Human for Interaction Research. In *ACM/IEEE International Conference on Human-Robot Interaction*, 2009.

[105] DagSverre Syrdal, Kerstin Dautenhahn, KhengLee Koay, MichaelL. Walters, and WanChing Ho. Sharing Spaces, Sharing Lives – The Impact of Robot Mobility on User Perception of a Home Companion Robot. 8239, 2013.

[106] Leila Takayama, Doug Dooley, and Wendy Ju. Expressing Thought: Improving Robot Readability with Animation Principles. In *ACM/IEEE International Conference on Human-Robot Interaction*, 2011.

[107] Leila Takayama and Caroline Pantofaru. Influences on Proxemic Behaviors in Human-Robot Interaction. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2009.

[108] Thora Tenbrink. The Verbalization of Cognitive Processes: Thinking-Aloud Data and Retrospective Reports. *Segmentation in Discourse*, 2008.

[109] Frank Thomas and Ollie Johnston. *The Illusion of Life: Disney Animation*. Hyperion New York, 1995.

[110] Elena Torta, Johannes Oberzaucher, Franz Werner, Raymond H Cuijpers, and James F Juola. Attitudes Towards Socially Assistive Robots in Intelligent Homes: Results from Laboratory Studies and Field Trials. *Journal of Human-Robot Interaction*, 1(2), 2012.

[111] Nico Van de Weghe, Anthony G. Cohn, Guy De Tre, and Philippe de Maeyer. A Qualitative Trajectory Calculus as a Basis for Representing Moving Objects in Geographical Information Systems. *Control and Cybernetics*, 35, 2006.

[112] Michael Walters, Kerstin Dautenhahn, René Te Boekhorst, Kheng Lee Koay, Dag Sverre Syrdal, and Chrystopher Nehaniv. An Empirical Framework for Human-Robot Proxemics. *New Frontiers in Human-Robot Interaction*, 2009.

[113] Astrid Weiss, Regina Bernhaupt, Manfred Tscheligi, Dirk Wollherr, Kolja Kuhnlenz, and Martin Buss. A Methodological Variation for Acceptance Evaluation of Human-Robot Interaction in Public Places. In *IEEE International Symposium on Robot and Human Interactive Communication*, 2008.

[114] Sarah Woods, Michael Walters, Kheng Lee Koay, and Kerstin Dautenhahn. Comparing Human Robot Interaction Scenarios Using Live and Video Based Methods: Towards a Novel Methodological Approach. In *Proceedings of the International Workshop on Advanced Motion Control*, 2006.

[115] Mitsumasa Yoda and Yasuhito Shiota. The Mobile Robot which Passes a Man. In *IEEE International Workshop on Robot and Human Communication*, 1997.