Validation of an Evaluation Framework for Human-Robot Interaction

The Impact of Usability, Social Acceptance, User Experience, and Societal Impact on Collaboration with Humanoid Robots

Dissertation

zur Erlangung des akademischen Grades Doktor phil.
an der Kultur- und Geisteswissenschaftlichen Fakultät
der Universität Salzburg

Fachbereich Politikwissenschaft und Soziologie

UNIVERSITÄT SALZBURG

Astrid Weiss
Salzburg, Feb. 2010

Erstgutachter: Univ. Prof. Martin Weichbold
Zweitgutachter: Univ. Prof. Manfred Tscheligi

© Copyright 2010 by Astrid Weiss.
All rights reserved.

Acknowledgements

The road to finish this PhD thesis was long and difficult. I would not have been able to complete this journey without the help and support of countless people over the past five years. I must first express my thanks towards my two advisors, Prof. Martin Weichbold and Prof. Manfred Tscheligi. Without the support of Prof. Weichbold, it would not have been possible to write an interdisciplinary PhD thesis like this where sociological theory and methods of empirical social research are combined with robotics. Prof. Tscheligi offered me the possibility to work in the EU-funded FP6 project "Robot@cwe: Advanced robotic systems in future collaborative working environments", which built the framework for my research efforts in Human-Robot Interaction. At this point I would also like to thank all members of the project consortium who supported my evaluation activities during the project. I would especially like to thank Prof. Aude Billard, who always supported me in my belief that it is possible for a woman to work in research and nevertheless have a family.

Furthermore, I wish to thank Dr. Regina Bernhaupt and Prof. Philippe Palanque in particular: The one month I spent at the Institute de Recherche en Informatique de Toulouse (IRIT) Toulouse in the context of an Short Term Scientific Mission (STSM) of cost action 294: Towards the MAturation of Information Technology USability Evaluation (MAUSE) set the starting point for this PhD thesis. They taught me the key elements of a PhD thesis!

Moreover, I would like to thank all the members of the HCI&Usability research group I have worked with at the ICT&S Center of the University of Salzburg, especially (in alphabetical order): Martin Altmaninger, Roland, Buchner, Sabine Hösch, Judith Igelsböck, Florain Förster, Michael Lankes, Bernd Ploderer, Thomas Scherndl, Daniel Schwaiger, and Daniela Wurhofer as well as all of my PhD colleagues who attended all the research seminars with me during the semesters, especially (in alphabetical order): Celina Raffel, Thomas Mirlacher, and David Wilfinger.

I also would like to thank all my Professors at the University of Salzburg, especially Prof. Reinhard Bachleitner who taught me my basic knowledge in methods of empirical research in social science. Without the tools I acquired during my Sociology studies, I would have never had the possibility to work in Human-Computer Interaction and Human-Robot Interaction.

Finally, I thank my family and my partner and my close friends for instilling in me confidence and a drive for pursuing my PhD.

Danksagungen

Der Weg bis zur Vollendung einer Dissertation ist lange und steinig. Es wäre mir niemals möglich gewesen, diese Reise über die letzten fünf Jahre zu einem Ende zu führen ohne die Hilfe und Unterstützung zahlreicher Menschen. Zuerst möchte ich meinen Dank meinen beiden Dissertationsbetreuern aussprechen, Prof. Martin Weichbold und Prof. Manfred Tscheligi.

Ohne die Unterstützung von Prof. Weichbold wäre es mir niemals möglich gewesen, eine interdisziplinäre Dissertation wie diese zu schreiben, in der soziologische Theorie und Methoden der empirischen Sozialforschung mit Robotik kombiniert werden. Prof. Tscheligi ermöglichte es mir, in dem EU subventionierten RP6 Projekt "Robot@cwe: Advanced robotic systems in future collaborative working environments" zu arbeiten, das die Rahmenbedingungen für meine Forschung im Bereich Mensch-Roboter Interaktion geboten hat. An dieser Stelle möchte ich auch allen Mitgliedern des Projektkonsortiums danken, die meine Evaluierungsaktivitäten während des Projektes immer unterstützt haben. Besonders möchte ich mich bei Prof. Aude Billard bedanken, die mich immer bestärkt hat, dass es für eine Frau möglich sei, Forschung und Familiengründung unter einen Hut zu bekommen.

Darüberhinaus möchte ich Frau Dr. Regina Bernhaupt und Prof. Philippe Palanque ganz besonders danken. Das eine Monat, das ich in Toulouse am Institute de Recherche en Informatique de Toulouse (IRIT) im Rahmen einer Short Term Scientific Mission (STSM) von cost action 294: Towards the MAturation of Information Technology USability Evaluation (MAUSE) verbrachte, setzte den Startpunkt für mein Dissertationsvorhaben. Von Ihnen habe ich gelernt, worauf es bei einer Dissertation ankommt!

Des Weitern möchte ich allen Mitgliedern der HCI & Usability Unit des ICT&S Centers der Universität Salzburg danken, mit denen ich zusammengearbeitet habe, im Besondern (in alphabetischer Reihenfolge): Martin Altmaninger, Roland Buchner, Sabine Hösch, Judith Igelsböck, Michael Lankes, Florian Förster, Bernd Ploderer, Thomas Scherndl, Daniel Schwaiger, and Daniela Wurhofer und allen meinen Dissertationskollegen, die mit mir die Forschungs- und Dissertantenseminare über die Semester hinweg besucht haben, im Besondern (in alphabetischer Reihenfolge): Celina Raffel, Thomas Mirlacher und David Wilfinger.

Ich möchte meinen Dank auch allen meinen Professoren an der Universität Salzburg aussprechen, besonders Prof. Reinhard Bachleitner, der mich die Grundlagen der Methoden der empirischen Sozialforschung lehrte. Ohne dieses Handwerkszeug, das ich während meines Soziologiestudiums gelernt habe, hätte ich niemals die Möglichkeit gehabt, in den Bereichen Mensch-Computer Interaktion und Mensch-Roboter Interaktion zu arbeiten.

Schlussendlich, danke ich meiner Familie und meinem Lebensgefährten und allen guten Freunden, dass sie immer wieder Zuversicht und Antrieb in mir geweckt haben, meine Dissertation zu beenden.

Abstract

Industrial robots started to enter the factories in the 1980s. From there on technology matured, however the industrial scenario has not changed a lot since then. Nowadays, with the help of increased computing power and improved navigation technologies, robots should move away from their simple and repetitive tasks they were assigned in the beginning (i.e., the "three D's" - dull, dirty, and dangerous jobs).

While consumers are warming up to robots that vacuum floors, mow the lawn, and serve as a toy for their children, university researchers are developing industrial robots that go beyond the same, definable tasks performed repeatedly in exactly the same fashion. They are working in order to let the robots enter environments which are unpleasant or dangerous for humans to work in, e.g. bomb disposal, work in space or underwater, in mining, and for cleaning of toxic waste. Moreover, they even attempt to realize Human-Robot Collaboration, meaning that the human carries out a task together with the robot, e.g. lifting a heavy object, carrying an object from place A to place B, mounting an object, etc.

However, there is still a lack in analysis and research of future socially acceptable and beneficial collaborative work scenarios, where robots can be integrated as co-workers. As robots are designed to work an integration of advanced robotic within future working contexts is not really surprising. In future working scenarios, robotic systems can be viewed as sophisticated tools that will assist humans. If properly designed, they would be powerful tools that our society could exploit to increase productivity and safety in work procedures.

Thus, this thesis proposes a theoretical and methodological evaluation framework for the assessment of human-robot interaction with humanoid robots in terms of Usability, Social Acceptance, User Experience, and Societal Impact - the USUS evaluation framework. The goal of this thesis is to validate the proposed framework by means of eleven case studies, to analyze the synergies between the proposed evaluation factors, to set up guidelines for the application of the framework, and to propose design implications for future human-robot collaboration scenarios.

In the beginning of this PhD thesis the theoretical and methodological evaluation framework is proposed based on an extensive literature review on existing frameworks and guidelines in Human-Robot Interaction, evaluation approaches already used in Human-Robot Interaction to address usability, social acceptance, user experience, and societal impact and a positioning of the

proposed evaluation framework in the user-centered design cycle. The framework is based on a multi level indicator approach and thus defines all evaluation factors and indicators and the methods with which those can be addressed during the evaluation process. By the means of eleven case studies, which were conducted in conjunction with the EU-funded FP6 project "Robot@CWE: Advanced robotic systems in future collaborative working environments", the feasibility of the framework can be proved. Based on a cross-analysis of the quantitative data gathered in a broad online survey an explanatory reference model on the synergies and influences between the proposed factors and indicators is presented.

Thus, the whole PhD thesis can be conceived of as series of case studies investigating several research questions on how users perceive the usability, social acceptance, user experience, and societal impact of humanoid robots proposed for the collaboration with humans. As a piece of exploratory research, the conducted case studies cannot guarantee a holistic validation of the theoretical and methodological evaluation framework, but the guideline-based application of the evaluation framework can positively influence the design process in Human-Robot Interaction projects.

Zusammenfassung

Bereits in den 1980er Jahren fanden Industrieroboter Einzug in unsere Fabriken. Seit damals ist die technische Weiterentwicklung stark vorangeschritten, jedoch hat sich das Einsatzszenario für Roboter nur wenig verändert. Heutzutage ist es möglich, dank höherer Rechnerleistungsfähigkeit und verbesserter Navigationstechnologie, Roboter nicht mehr nur für einfache, repetitive Aufgaben zu nützen, für die sie ursprünglich gedacht waren (die sogenannten 3Ds - dull (langweilig), dirty (schmutzig) und dangerous (gefährlich)).

Während Konsumenten sich immer mehr mit Robotern anfreunden, die staubsaugen, Rasen mähen oder ein Kinderspielzeug sind, beschäftigt sich die Forschung damit, Industrieroboter zu entwickeln, die mehr können als nur immer wieder dieselbe vordefinierte Aufgabe in gleicher Art und Weise auszuführen. Die Forschung ist daran interessiert, Roboter zu entwickeln, die in unwegsamen und/oder für den Menschen zu gefährlichen Gebieten arbeiten können, wie z.B. Bombenentschärfung, Arbeiten im Weltraum, unter Wasser, auf Minenfeldern oder Entsorgung von Giftmüll. Darüber hinaus wird daran geforscht, Mensch-Roboter Kollaboration zu ermöglichen, d.h. ein Mensch erledigt eine Aufgabe in Kooperation mit einem Roboter, wie z.B. ein schweres Objekt hochzuheben, ein Objekt gemeinsam von Platz A zu Platz B zu transportieren oder ein Objekt gemeinsam zusammenzubauen.

Jedoch mangelt es immer noch an Studien und Analysen von zukünftigen sozial akzeptierten und wünschenswerten kollaborativen Arbeitsszenarien, in welche Roboter tatsächlich als "Mitarbeiter" integriert sind. Roboter wurden entwickelt um zu arbeiten, daher ist es nicht überraschend, dass weiterentwickelte Robotersysteme in die zukünftige Arbeitswelt integriert werden. In einem Arbeitskontext können Roboter als hochentwickelte Werkzeuge gesehen werden, die dem Menschen bei seiner Arbeit assistieren. Wenn Roboter angemessen entworfen werden, können sie mächtige Werkzeuge darstellen, die unsere Gesellschaft einsetzen kann, um die Produktivität und Sicherheit in Arbeitsabläufen zu steigern.

Daher wird in dieser Dissertation ein theoretisches und methodisches Rahmenmodell zur Evaluierung von Mensch-Roboter Interaktion mit humanoiden Robotern vorgeschlagen, welches die Faktoren Usability, Soziale Akzeptanz, User Experience und gesellschaftliche Auswirkungen berücksichtigt - das USUS Evaluierungsmodell. Zielsetzung dieser Dissertation ist es, das vorgeschlagene Rahmenmodell mit elf Fallstudien zu validieren und die Synergie-Effekte zwischen den einzelnen Evaluierungsindikatoren und -faktoren aufzuzeigen. Schließlich sollen darauf basierend Richtlinien für die Anwendung des

Rahmenmodells und für Designimplikationen zukünftiger Mensch-Roboter Interaktionsprojekte gegeben werden.

Zu Beginn der Dissertation wird ausführlich auf existierende Rahmenmodelle und Richtlinien im Bereich Mensch-Roboter Interaktion sowie bereits durchgeführte Evaluierungsstudien zu Usability, Sozialer Akzeptanz, User Experience und zu den gesellschaftlichen Auswirkungen im Bereich Mensch-Roboter Interaktion und auf die Positionierung des vorgestellten Rahmenmodells im Nutzer-zentrieten Designzyklus eingegangen. Das Rahmenmodell basiert auf einem vielschichtigen Indikatoransatz, welcher die vier Evaluierungsfaktoren, ihre einzelnen Indikatoren und die Methoden, mit welchen diese evaluiert werden können, definiert. Das Rahmenmodell wird durch elf Fallstudien validiert, die in Verbindung mit dem EU-subventionierten RP6 Projekt "Robot@CWE: Advanced robotic systems in future collaborative working environments" durchgeführt wurden. Basierend auf den quatitativen Daten einer breiten Onlineumfrage wird eine Queranalyse durchgeführt, um ein Referenzmodell über die Synergieeffekte zwischen den einzelnen Evaluierungsindikatoren und -faktoren zu erstellen.

Somit kann diese Dissertation im Gesamten als eine Serie von Fallstudien gesehen werden, die eine Vielzahl von Teilforschungsfragen beantworten, wie Nutzer die Usability, Soziale Akzeptanz, User Experience und die gesellschaftlichen Auswirkungen von humanioden Robotern wahrnehmen, die für kollaborative Aufgaben bestimmt sind. Da dieser Dissertation ein explorativer Ansatz zugrunde liegt, können die durchgeführten Fallstudien die Validierung des gesamten theoretischen und methodischen Rahmenmodells zur Evaluierung nicht garantieren, jedoch kann die Anwendung dieses Rahmenmodells den Designprozess von Mensch-Roboter Interaktionsprojekten positiv beeinflussen.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 HRI and HRC as Research Areas

Since the fictional play of Josef Capek on Rossmus Universal Robots (RUR) (Capek, 1920) it became popular belief that robots should perform a variety of "dull, dirty, and dangerous" tasks humans would rather not perform themselves. Robots can be used for painting, welding, and assembly of cars. Certainly robots are suitable for these kinds of tasks as they are clearly definable, need to be fulfilled accurately, and must be performed exactly the same every time. As industrial robots entered factories in the 1980ies, the industrial scenario did not change a lot despite the great strides technology made (US Department of Labor, 1994). Technology has matured since then and it became possible for robots to move away from the simple and repetitive tasks they were originally designed for.

It becomes more interesting to extensively use robots in specific working environments. However, for collaborative work scenarios, it is important that robots will be socially accepted as sophisticated tool assisting humans. Consumers are already warming up to robots that vacuum floor, mow the lawn, and serve as companions for their children. Could consumers accept industrial robots that could become co-workers able to perform a variety of tasks even in unstructured environments?

The research field of Human-Robot Interaction (HRI) is concerned with this kind of question. Research within this field is highly interdisciplinary; integrating disciplines like robotics, engineering, computer science, Human-Computer Interaction (HCI), psychology, ethology and social science disciplines like communication science and sociology. HRI is a young research area existing approximately for 20 years. The annual IEEE RO-MAN conference series, originating in 1992 in Japan, has since then traveled across the world, reflecting this emerging new field. The growing interest for this research area can be determined by the inaugural annual ACM/IEEE International Conference on Human-Robot Interaction (HRI 2006).

As the term HRI suggests, the main focus in research is centered on the interplay between a human and a robotic agent, where the interaction can be verbal and non-verbal (Dautenhahn, 2007b). There exists no "official" and commonly agreed on definition of HRI - perhaps because of the different embodiments robotic agents can have, and the different ways in which humans can interact with a robot (e.g. communication via speech, gestures or symbols). Fong et al (2001) define HRI the following: "Human-robot interaction (HRI) can be defined as the study of humans, robots, and the ways they influence each other." This definition goes in line with the understanding of HRI in this PhD thesis. Dautenhahn (2007b) defines three possible "not mutually exclusive" research directions HRI can take.

1. Robot-centered HRI, which highlights the view of a robot as a creature. The robot is seen as an autonomous entity which can fulfill its goals only by interacting with humans, whereas the goals are identified by the designer of the robot and modeled by its control architecture. This indicates an interaction, even if the robot navigates through an environment where humans are present. An example research questions for this area of HRI is: "How to develop a sensor motor control that regulates interactions with the (social) environment?"

2. Human-centered HRI, highlights the view of the human as user of the robot. Humans should be able to fulfill tasks together with the robot in a comfortable and acceptable way. Thus, the focus is laid on how people react to the robot, how they interpret its behavior independently from its architecture and the cognitive processes, that might happen inside the robot. Research interests in this area lie in the design of the robotic appearance in accordance to its functionalities, in analyzing user requirements which a robot can address, evaluation studies of HRI interaction in terms of usability or acceptance, and in developing new methods and methodologies for HRI.

3. Robot cognition-centered HRI highlights the robot as an intelligent system. Here the robot is seen in a traditional sense of artificial intelligence as a machine able to make decision on its own and solve problems of a particular application domain. Specific research questions in this domain include the development of cognitive robot architectures, machine learning, and problem solving.

HRI is concerned with various usage areas of robots which can be clustered in accordance with Green et al (2008) into (1) robots as tools, (2) robots as guides and assistants, and (3) humanoid robots. Robots as tools should, above all, support the completion of physical tasks, like robots that increase the success rate in harvesting, robots used for the placement of radioactive waste in centralized intermediate storage, or human-robot teams in urban search and rescue. Green et al (2008) highlight that for effective human-robot collaboration adjustable autonomy and situational awareness of the robotic agent are of essential necessity.

A guiding robot could be used in a museum as an autonomous mobile robot, which communicates and interacts with humans as a partner or guide. Studies on these robots showed that this type of robot should portray human-like behavior to be accepted by the people. Furthermore, there is a need for effective natural speech and multi-modal communication approaches that go beyond speech, such as gestures (Green et al, 2008).

Humanoid robots are anthropomorphic designed robots, having the shape of humans. Research on this kind of robot focuses on making robots appear human in their behavior and communication abilities. In terms of collaboration, the interest lies in building socially intelligent cooperative robots that can work and learn in partnership with people. Thus the goal is to have robots learn as quickly and easily and in a similar way as humans (Green et al, 2008).

Current developments in HRI tend to a human-centered design approach, aiming to design the interaction with a robot in a way that is perceived as usable and natural. User-centered design[1] in HRI includes many different aspects, like requirement analysis, modeling the interaction, and evaluating it with potential users (see Chapter 2).

1.2 Sociological Aspects of HRI

To what extent is the research discipline of sociology concerned with research interests of HRI? Traditional, sociology (Weberian stance) argues that non-humans should be excluded from the field. Also, the sociologist Luhmann would not consider verbal and gestural exchanges of humans with technological devices as communication, as artifacts have no consciousness and thus not the ability of understanding. However, the research field of robotics headed towards a direction where engineers develop robots after anthropomorphic images. Some researchers even go beyond anthropomorphic images, but work on artificial life by developing cognitive systems, which already passed the false

[1]The terms human-centered design and user-centered design are used interchangeably in this PhD thesis

belief test (Leonardo) and the mirror test (Nico). It is hoped (and feared) that humanoid robots will in future act like humans and be an integral part of society. Thus, social science definitely should consider the interaction between humans and robots as research subject.

Already sociologist of technology, like Hans Linde and Werner Rammert, have discussed the classical Weberian model, where the social world remains for human entities and argued for a rethinking of this model towards machines and technological devices. However, reviewing the state-of-the-art literature on HRI research (see Chapter 2) and looking on the data gathered during the case studies presented in this thesis (see Chapter 4), it becomes obvious that people believe robots to be social actors. Furthermore, studies in the area of sociology of knowledge point into the direction that humans tend to express similar behaviors towards artifacts and objects as towards humans, such as talking to a robot or expressing emotions. Knorr Cetina is speaking in this context of a "post-social" world where non human entities enter the social domain.

This trend is also obvious, as robots more and more enter the direct interaction space of humans, not only in the working context, but also in the service sector and the domestic context. Research focuses more and more on robots as "social actors" and how long-term relationships between humans and robots can be enabled. Toy robots have already entered the rooms of children. According to the World Robotics study 2008 (conducted by the IFR Statistical Department, which is hosted by the VDMA Robotics + Automation association) there are more than 3000 service robots for domestic use (vacuum cleaners, lawn mowers) and more than 2000 for personnel use (entertainment and leisure) by the end of 2007. However, it is expected to increase in numbers within the next four years (approximately 30% in the service sector and 70%).This raises several questions about the future role of humans in society in comparison to robots. What will be the main characteristics distinguishing humans from robots? Will robots have to follow the same social norms as humans? Will the increased interaction with robotic agents cause a change in the quality of human relationships?

While reviewing the relevant HRI related literature for this PhD thesis (see Chapter 2), it became clear that research fields like (cognitive) psychology, anthropology, linguistics, and even philosophy already have entered the research field of HRI, but there is still little attention paid to sociological work. However, robotics are currently entering the "next level", as it is considered by engineers that robotic cognition needs to gain information from the socio-cultural context in which robots are integrated. Based on social theory and human-human interaction studies insights can be gained how humans make sense of the world, communicate with each others and plan and conduct action sequences. During my work on this PhD thesis together with engineers and computer scientists it became obvious that these disciplines can be complementary for sociology. On the one hand for me as a sociologist it was interesting to get an understanding about social behavior towards machines and social interaction during joint tasks with robots. On the other hand for engineers and computer scientists it was valuable to get to know humans' communication and problem solving strategies in the interaction with robotic systems to improve them.

1.3 Research Objectives

The integration of humanoid robots into collaborative work settings is a challenging endeavor. It needs to be considered that naive human users will be confronted with severe problems when interacting with humanoid robots. Autonomous robots are perceived differently than other computing technologies as Kiesler and Hinds (2005) show. Users tend to have a far more anthropomorphic mental model for autonomous robotic agents than for other interface technologies. Similarly, as mobile autonomous robots have to adapt to their physical environment, they also have to adapt to the humans they are working with. In other words: The interaction between robots and humans always has to be negotiated between those two actors. Furthermore, robots "learn" about themselves and the environment in which they are situated through means of their interaction history world, which heavily distinguishes them from traditional computing technology (Kiesler and Hinds, 2005). All of these factors influence the interaction between humans and humanoid robots in collaborative work spaces.

However, these facts still need to be explored in how they affect the evaluation of interaction between humans and robots. An interesting question remains, how to consider the basic different perception of humanoid robots in the evaluation of human-robot interaction scenarios. This PhD thesis presents a novel approach for evaluating the interaction experiences of humans with humanoid robots. The approach is based on four main evaluation factors, namely usability, user experience, social acceptance, and societal impact. It is assumed that these are the relevant core factors, enabling future working scenarios in which robots are included to increase productivity and safety. Based on these core factors, a theoretical and methodological evaluation framework - the Usability, Social Acceptance, User Experience, and Societal Impact (USUS) evaluation framework - is developed from a human-centered HRI perspective. The evaluation framework is built around a theoretical model (derived from literature) and a methodological framework, defining potential suitable methods to address the theoretical framework. By means of eleven empirical case studies and a broad online survey, the evaluation framework will be assessed in terms of feasibility and a revised empirically validated model will be proposed (see Chapter 5). The revised USUS evaluation framework should then serve as tool to help other evaluators to explicitly target various evaluation factors with humanoids robots and to consider a broader evaluation context than pure performance measures.

The main objectives of this thesis are:

1. Proposing a theoretical and methodological framework for evaluating usability, social acceptance, user experience, and societal impact for working scenarios with humanoid robots.

2. Assessing the feasibility of the proposed evaluation framework by means of empirical case studies.

3. Developing a reference model for USUS evaluation framework to reveal interrelations between the factors and indicators of the evaluation framework.

4. Providing guidelines for the application of the proposed framework to support other evaluators.

1.4 Organization of the Thesis

This PhD thesis is written from the perspective of the human-centered HRI research. It is concerned with the usability, social acceptance, user experience, and societal impact of HRI in collaborative work settings with humanoid robots. The main objective is the validation of a theoretical and methodological framework for evaluating HRI by means of empirical case studies (the USUS evaluation framework). This framework is based on a multi level indicator approach defining the relevant factors for evaluation and reasonable methods to address the indicators during the development process of human-centered HRI.

Chapter 2 This thesis begins with an introduction on evaluation approaches in HCI and HRI in general. State-of-the-art research in HRI on measures, methods, and evaluation approaches addressing the factors usability, social acceptance, user experience, and societal impact will be discussed. Next, the theoretical and methodological frameworks for in human-robot interaction will be reviewed. The chapter will conclude with a review on existing guidelines in HRI with a focus on how these guideline were derived from theoretical and empirical work.

Chapter 3 The theoretical and methodological evaluation framework proposed for human-robot collaboration with humanoid robots will then be explained in detail in this chapter. A definition of all four evaluation factors will be presented. Next, all indicators determining a factor will be defined and explained by means of a literature review. This chapter concludes with the description of the method mix proposed to address the indicators during the evaluation approach.

Chapter 4 Here, the proposed evaluation framework will be validated by eleven case studies conducted within the FP6 EU-Project "Robot@cwe: - Advanced robotic systems in future collaborative working environments". Studies using expert evaluation techniques, as well as user studies conducted in the field and in the laboratory are presented with a focus on how the results support the proposed evaluation framework, in particular the methodological framework.

Chapter 5 This chapter will present a reference model explaining the synergies and influences between the evaluation indicators of the user experience and social acceptance indicators. The reference model is based on a cross-analysis of the quantitative data gained in a broad online survey. The reliability and validity of the user experience and social acceptance indicators is tested by a factor analysis and Cronbach's alpha. Based on the data analysis a revised theoretical evaluation framework will be proposed.

Chapter 6 This is where the application of the proposed framework will be discussed. In a first step, the methodological challenges which had to be faced during the case studies will be reviewed and methodological considerations for future application of the framework will be concluded. Furthermore, a revised methodological framework will be proposed. Then, the process of applying the proposed evaluation framework based on guidelines and research questions derived from the case studies will be described and its relation and integration in the user-centered design approach for HRI.

Chapter 7 The final chapter will lead back to the research objectives and discuss the proposed USUS evaluation framework and its validation. Some ideas for promising future work in the research area of human-centered HRI will conclude this PhD thesis.

1.5 Contributions of the Thesis

In this thesis, human-robot interaction is viewed from a user-centered point of view taking into account individual and societal levels of robot perception (see Figure 1.1).

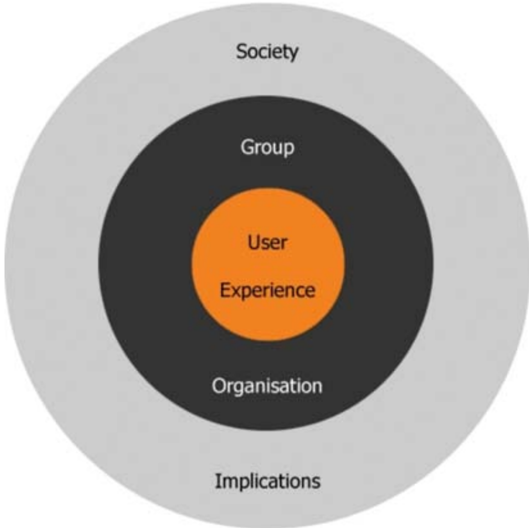


Figure 1.1: Individual and Societal Levels of Robot Perception

There is a focus on two high-level questions: "how-to-evaluate human-humanoid interaction beyond performance issues" and "how to produce comparable results in human-humanoid interaction studies"? The indicators, factors, and methods proposed and used in this thesis are not new and have already been employed in other research areas, as HRI, HCI, and social sciences. However, the concern can be raised that these methods need to be applied with care to HRI, due to the unique perception of autonomous systems and the inherent limitations of all methods. The USUS evaluation framework should support targeting a holistic interaction evaluation. Moreover, the validation of the framework should help to gain more insights on the interplay and interdependencies between the different indicators by means of a reference model. The contributions of this thesis are threefold:

1. It contributes to the research field of HRI by proposing a theoretical and methodological framework to deal with usability, social acceptance, user experience, and societal impact evaluation issues. More specifically, it contributes to different evaluation settings to gain comparable results.

2. It contributes to methods of empirical social research and HCI by propos-

ing adaptations of existing evaluation methods to novel evaluation settings.

3. Last but not least, it contributes to the sociology of technology through adapting traditional sociological theory for the research field of HRI and by gaining insights on future society and individuals' interactions with robotic systems.

The feasibility of the proposed evaluation framework is also demonstrated through the eleven case studies conducted to validate it (see Chapter 4) by means of application guidelines (see Chapter 6) for its future usage. By letting other researchers access the questionnaires and interview guidelines (see Appendix A) and sharing the methodological challenges and considerations (see Chapter 6), this work has an additional benefit: The evaluation studies and results are reproducible by other researchers and new evaluation studies can be conducted (e.g. in another context, like domestic settings) on this basis.

Chapter 2

State of the Art

2.1 Outline of the Chapter

This chapter presents a literature review on the concept of interface evaluation in general, on existing evaluation frameworks in HRI, as well as on state-of-the-art evaluation experiments and methodologies addressing usability, social acceptance, user experience, and societal impact in HRI. This chapter will conclude with a section on evaluation and design guidelines derived from HRI as this is the final goal of the evaluation framework validation in this PhD thesis.

The literature review investigated the goals and purposes of evaluation approaches and how user-centered evaluation can be realized in HRI. It reports on different types of existing theoretical and empirical frameworks. The advantages and limitations of these frameworks will be discussed and the differences to the proposed framework in this thesis will be outlined. Additionally, an overview of existing parameters and measures for usability, social acceptance, user experience, and societal impact of successfully conducted evaluation studies in HRI will be given. Differences between qualitative and quantitative approaches and lab-based and field-based studies will be discussed and their pros and cons will be summarized. Finally, design and evaluation guidelines will be presented with a focus on how researchers derived them from theoretical considerations or experimental studies to improve the design and development of robotic agents.

2.2 Evaluation: Purpose and Goals

Evaluation approaches in HCI and HRI primarily determine if a system achieves its predetermined goals and purposes. In this PhD thesis, evaluation is understood according to the definition of Preece et al (2002) as “a process of systematically collecting data that informs designers about how to improve the use of a system for a particular user or group of users for a particular activity in a certain type of environment”. The goal of evaluation is to assess how well a design fulfills users’ needs and whether or not users like a system. Dix et al (2004) describes three main goals of evaluation:

1. To assess the extent and accessibility of the system functionalities.

2. To assess users’ experience of the interaction.

3. To identify any specific problems with the system.

This does not imply that evaluation is a one time conducted event. Evaluation is an ongoing activity which is used to monitor the whole system development and which is embedded into the so called user-centered design approach.

2.2.1 User-centered Design

User-Centered Design (UCD) is a design philosophy taking into account the process orientation of system development. The major focus of UCD lies in the user centricity of design; meaning that the design of the system adapts to the needs and wants of the potential user group so the system gets optimized in a way that the user does not need to change his/her habits when working with it. An international standard exists as basis for many UCD methodologies, the ISO 13407: Human-centered design process. This defines a general process (see Figure 2.1) for including human-centered activities throughout a development life-cycle, without specifying the methods to achieve this (ISO 13407, 1999). Four activities form the main cycle of user-centered system development. The cycle (process) ends, when the system meets all specified requirements. The four activities of the system are as follows:

1. Specify the context of use: At this stage, the potential user group of the system should be identified; specifically the purposes and conditions under which they will use the system.

2. Specify requirements: Business, interface, and user requirements that must be met have to be identified here, so the system will be used successfully.

3. Create design solutions: At this point of the cycle the system is developed (which is done in stages). The system is built from a rough concept to a complete design.

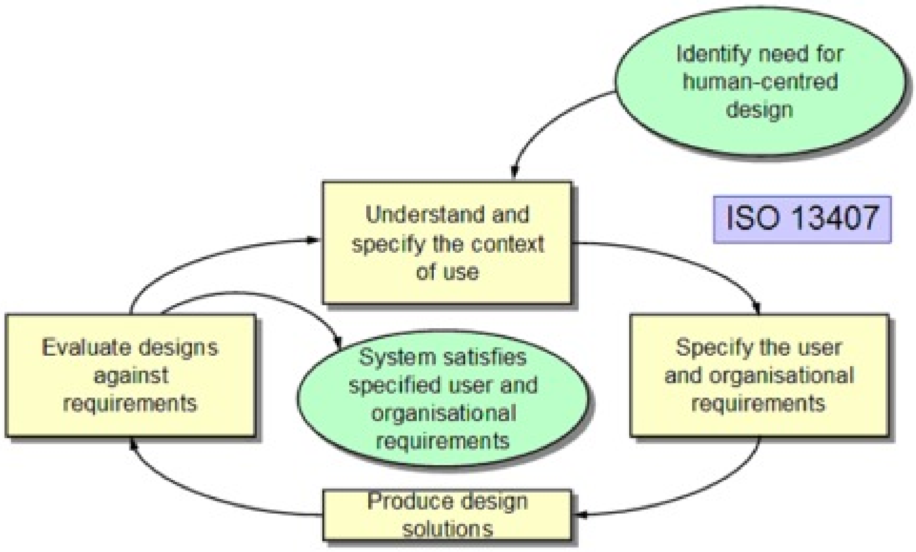


Figure 2.1: ISO13407: Human-centered Design

4. Evaluate designs: This activity can be seen as the most important one, as it is the repeated "quality testing" of the system. Depending on the stage of the system design, it will be evaluated with two different methods (expert-based or user-based evaluation).

User-centered design already found its way into HRI. Hüttenrauch (2007) e.g. showed in his thesis how user-centered design can inform the design of a mobile services robot (fetch-and-carry robot) in an office environment. Therefore, he adapted user-centered design methodologies and approaches from HCI for HRI. Based on a longitudinal study, he investigated the user roles andthe spatial management when face-to-face interaction with the robot occurred (Hüttenrauch, 2007). Similarly, the first results of a long-term HRI development project based on a user-centered design approach for a mobile service robot can be found in Lee et al (2009). Riek and Robinson (2009) specifically propose an affective-centered design cycle for HRI to support the affective quality in robotic design to increase the acceptance of robotic agents.

User evaluation ideally incorporates user-centered design principles: (1) early focus on users, (2) empirical measurement of usage, and (3) iterative design (Courage and Baxter, 2004). The first principle describes the systematic and structured collection of user requirements in order to gain an understanding of what users really want and need, how they currently use similar systems, and/or how they would like to use it. Based on this gathered information, a reasonable evaluation strategy can be formulated and further improved in the user-centered design process.

The second principle focuses on the usage of the system. This can be assessed early in the design-cycle to determine if the usage is user friendly, e.g. by means of usability testing of early prototypes (Courage and Baxter, 2004). Users want systems that are easy to learn and use as, well as effective, efficient, safe, and satisfying. Being entertaining, attractive, challenging, etc. is also essential for some products. Thus, knowing what to evaluate, why it is important, and when it is reasonable to evaluate it, has to be defined in an evaluation plan (Preece et al, 2002).

Repeated evaluation of the system is suggested in the third principle. It stresses that evaluation should be an iterative process, as not all information can be gathered the first time in a single user study. After a first evaluation, it is common to find that some user or system requirements cannot be met as stated in the first specification. In this case, the requirements will need to be revised and the evaluation strategy has to be reformed to reflect the changes of the requirements (Stone et al, 2005).

2.2.2 Methodological Considerations and Challenges

Two general methodological evaluation approaches can be distinguished: *formative and summative*. An evaluation is *formative* if its results are a direct feedback for optimizing the design process. The formative evaluation contains the idea of quality assurance: what and how to design. At various points of the development process, the system is evaluated. Therefore, it is important to design the evaluation process in a way that the immediate results can be feedback for the development of the program. A typical way to evaluate in the field of HCI is to let a user fulfill realistic tasks with a prototype while asking him to think aloud, obtaining direct information of evolving problems (Rosson and Carroll, 2002).

A *summative* evaluation is conducted in the end when the system is already fully developed to asses the design result. The goal is a summarizing assessment of direct or subsequent effects and advantages of the system: does the system meet its specified goals, is it better than its competitors? It is sought to know if the program will be as successful as estimated or even more. Therefore, the summative evaluation tries to measure quality and contains the idea of quality control: How well did we do? (Rosson and Carroll, 2002).

Similarly, it can be distinguished between *analytic and empirical* evaluation methods. Empirical evaluation integrates actual users in the evaluation process and therefore gathers solid facts which can be done via different methods such as observation or questionnaires. Nevertheless, the gathered material needs a good interpretation based on a deep understanding of the tested system. The *empirical* evaluation then can be conducted in two styles: laboratory studies and field studies (Dix et al, 2004).

1. Laboratory studies are controlled tests. Normally, a usability laboratory is well equipped with audio/visual recording possibilities and analysis facilities. One advantage is that the participant works in an interruption free environment. On the other hand, the testing situation misses its context and can not reflect the real life. In summary, laboratory studies are appropriate when a controlled context is needed and only constrained single-user tasks are tested.

2. In a field study, the evaluator goes into the user's environment to observe the system in its context. The natural situation in which the evaluation takes place can show interactions which could be missed in a laboratory. However, participants in field studies are interrupted by the evaluator or the evaluation material. It is more difficult to interpret the results of field studies as the influence factors for human behavior cannot be controlled.

Analytic evaluations identify the characteristics that influence the performance of the user with the system. Thus, analytic evaluations are based on interpretation but do not produce solid facts. The mixed approach to combine the strengths of an empirical and an analytical evaluation is called mediated evaluation. In this case, the analytic evaluation is conducted during the design process and the results are used as basis for an empirical evaluation.

According to Clarkson and Arkin (2007) "evaluation on HRI systems has not received its due attention". Only few experimental evaluation studies have been conducted so far. Moreover, most of them only assure that HRI display and interaction controls are intuitive for their users. However, more researchers have recently come to recognize the need for evaluation approaches, common metrics, and frameworks in HRI.

2.3 State of the Art on Evaluation Frameworks for HRI

Human-Robot Interaction is a young, but growing research field. "If we are to make progress in this field then we must be able to compare the results from different studies." (Bartneck et al, 2009). There is an increasing need of structured evaluation and measurement approaches in HRI to allow a systematic prototype development. The theoretical and methodological frameworks proposed in this thesis follow Bartneck's claim and try to give HRI researchers and developers a toolkit for assessing their work. For the development of the evaluation framework, other already existing frameworks for design, development, and evaluation of HRI were taken into account and presented below.

One of the first theoretical frameworks for HRI was presented by Thrun (2004). He distinguishes three types of robots in this framework: (1) industrial robots, (2) professional service robots, and (3) personal service robots. Thrun's goal with this framework was "to introduce the reader to the rich and vibrant field of robotics in hope of laying out an agenda for future research on human robot interaction" (Thrun, 2004). Therefore, he describes in great detail different HRI interface capabilities, potential user groups and usage contexts for the three robot types. This laid a basis for future evaluation approaches of robotic systems.

Yanco et al (2004) went into a similar direction when they published their updated version of taxonomy for HRI to provide a basis for researchers. This taxonomy takes into account research from different neighborhood disciplines like HCI, Computer Supported Cooperative Work (CSCW), and social sciences to inform research in HRI. In their taxonomy, they propose eleven categories which should be considered when investigating the interplay between humans and robots: (1) task type, (2) task critically, (3) robot morphology, (4) ratio of people to robots, (5) composition of robot teams, (6) level of shared interaction among teams, (7) interaction roles, (8) type of human-robot physical proximity, (9) decision support for operators, (10) time and space taxonomy, and (11) autonomy level/level of intervention. This taxonomy set the first standards towards comparable HRI research approaches and generalizable evaluation study results. To increase the generalizability in HRI evaluation studies, the need for standardized measures, metrics, and methodological approaches in HRI increases as well. This fact is represented in and by the inaugural workshop on "Metrics in HRI" held at the 3rd ACM/IEEE conference on HRI (HRI2008). The aim of this workshop was to "propose guidelines for the analysis of human-robot experiments and forward a handbook of metrics that would be acceptable to the HRI community and allow researchers both to evaluate their own work and to better assess the progress of others".

One of the first attempts in providing standardized metrics for HRI research was proposed by Steinfeld et al (2006). The basis of their evaluation framework for task-oriented human-robot interaction with mobile robots consists of five metrics: (1) navigation, (2) perception, (3) management, (4) manipulation, and (5) social They also take into account "biasing effects" of their metrics, such as communication factors (e.g. delay, jitter,bandwidth), robot response (e.g. system lag and update rate), and the user (e.g. training, motivation, stress). In general, the five metrics proposed by Steinfeld et al (2006) are still rather generic. The authors wrote in the conclusion of their article their proposed evaluation plan "is to provide a living, comprehensive document that future research and development efforts can utilize as a HRI metric toolkit and reference source".

Similarly first approaches exist to provide measures for user experience in HRI. Bartneck et al (2009) developed a standardized validated questionnaire based on an extensive literature review of existing questionnaires in HRI. It addresses: anthropomorphism, animacy, likability, perceived intelligence, and perceived safety. All items of these five questionnaires are based on a semantic differential. The aim and hope of the authors is to provide robot developers with a tool that allows them to easily monitor their progress from a user-centered point of view.

Moreover, structured work is already done to investigate the concept of acceptance in HRI. The Dutch research group surrounding Vanessa Evers in Human-Computer Studies (University of Amsterdam) focuses on the exploration of the feasibility of the Unified Theory of Acceptance & Use of Technology (UTAUT) model (see Venkatesh et al (2003)) for HRI. They investigate, above all, the acceptance of robotic system in elder-care institutions and if the UTAUT model is a suitable explanation model for acceptance and rejection of robotic technology by elderly users (Heerink et al, 2006a). Influence factors such as enjoyment (Heerink et al, 2008b), social presence (Heerink et al, 2008a), and social abilities (Heerink et al, 2006b) are also investigated. Based on these studies, the research group of Vanessa Evers wants to set-up an acceptance model for robotic usage by elderly users.

The evaluation framework proposed in this PhD thesis tries to integrate four main evaluation concepts in the theoretical framework for human-humanoid interaction: usability, social acceptance, user experience, and societal impact (USUS evaluation framework). In applying the methodological framework, the evaluation framework intends to answer the generic research question of whether or not people experience humanoid robots as support for cooperative work and accept them as part of future society. The methodological framework consists of a method mix proposing suitable methods to address several factors/indicators in parallel in one study. An overview will be given on methodological approaches already successfully used in HRI to evaluate, usability, social acceptance, user experience, and societal impact.

2.4 State of the Art on Evaluation Methodologies for HRI

Most of the studies described below concentrated their evaluations on one specific evaluation factor, but only rarely on several factors at the same time. The methods described in these studies represent qualitative as well as quantitative approaches and were conducted in lab-based as well as field-based settings.

Several evaluation studies on usability, social acceptance, user experience, and societal impact of robots in working environments have been conducted. Hinds et al (2004) e.g. researched people's perception of robots as working partners. She investigated differences in reliability and feeling of responsibility between the human and robot. Results showed that participants relied more on humans as working partners but results were different regarding the concept of responsibility. Interestingly, there was a slight to no difference between participants' feelings of responsibility towards a human and a robotic working colleague. The comparison between human-like and machine-like robots in terms of the willingness to rely on also was not significant.

Takayama et al (2008) investigated how people accept and feel about a robot in their workplace. They looked into which aspects of work are perceived as being appropriate for robot workers, and how people feel about robotic workers compared to human workers. Participants preferred robots to do jobs that require memorization, perceptual abilities, and service-orientation. Furthermore, they could show that participants have a more positive attitude towards robots collaborating with people instead of replacing human workers and skills by robots. Another interesting aspect of the integration of robots into working environments was revealed by Mutlu and Forlizzi (2008). Based on a long-term ethnographic study, where a deliverable robot was introduced into the everyday working environment of a hospital, they could show that perception of the robot depends on the user group. Their results showed significant differences in how people responded to the robots between two different organization units in the hospital. The delivery robot was differently integrated into the work-flow of the two user groups and the way of interaction was also different.

For reasons of creating a more holistic picture on how the integration of humanoid robots influences working environments, the evaluation framework proposed in this thesis takes into account several evaluation factors, split into multiple indicators. The goal is to address more than one indicator during a user study by means of a methodological mix. However, the evaluation framework naturally takes into account already successfully conducted user studies and existing measures in the field of HRI and takes advantage of established theoretical work on defining the indicators (see Chapter 3). State-of-the-art studies will present and address the factors usability, social acceptance, user experience, and societal impact in the following sections.

2.4.1 State of the Art on Evaluating Usability

Knowledge gained in HCI research served as a starting point for usability evaluation in HRI. Yanco et al (2004) e.g claimed to transfer traditional factors of HCI and CSCW to the assessment of usability in HRI. In HRI, the evaluation of usability is addressed by expert evaluations using guidelines and heuristics (Scholtz, 2002; Clarkson and Arkin, 2007; Lohse and Hahnheide, 2008; Marble et al, 2004) as well as experimental studies (Breazeal et al, 2004; Fong et al, 2001; Goetz and Kiesler, 2002). HRI researchers address several concepts under the term usability: performance, effectiveness, efficiency etc. Studies in several different contexts have been conducted; single- or multirobot configurations (Olsen and Wood, 2004), rescue scenarios (Drury et al, 2005), and robots in elderly care (Heerink et al, 2006a), aiming to inform the design of the robot and improve the interaction with it.

The reflections on assessing usability issues in HRI reported in the aforementioned studies indicate that it is relevant to address the factor usability with expert- as well as user-based methods. Furthermore, it seems to be necessary in HRI usability studies to see usability as a holistic concept. Usability evaluation in HRI has to take into account aspects like utility, flexibility, learnability, and robustness of the robotic system.

2.4.2 State of the Art on Evaluating Social Acceptance

One of the first extensive empirical analyses on the acceptance of a robotic system has been conducted by Apostolos (1985). She assessed the acceptance of a robotic arm as a personal assistant in a rehabilitative setting by conducting several empirical surveys. Her studies could not reveal conclusive evidence to the hypothesis that aesthetic movement of a robotic arm can affect an individual's attitude towards the acceptance and use of robots, but it was a first try to investigate acceptance directly in the context of use.

Until today, several different methods and measures have been used to investigate social acceptance of robots. Mainly it is possible to distinguish between laboratory-oriented and field-oriented methods. Laboratory-based trials are in general more often based on behavioral measures, field-based trials on methods of empirical social research, like surveys, interviews etc.

Kulic and Croft focus their research on multi-modal capturing of human physiology (heart rate, perspiration rate, and facial muscle contraction). The aim was to increase the robustness of robot perception of a human's state (like anxiety, valence, attention) and thus increase the acceptance of the robotic agent (Kulic and Croft, 2005, 2007).

Shiwa et al (2008) focus on the System Response Time (SRT) of a communication robot with human-like social features. A preliminary experimental study revealed that participants preferred one-second delayed responses from the communication robot rather than immediate responses. Furthermore, they found out that participants favored a human-oriented communication style, as conversational fillers were preferred when the response time was longer than two seconds. However, their experimental results give only an insight on the influence of SRT on social acceptance and did not take into account natural realistic disturbance moments, as the study was conducted in a lab-based setting.

An interesting approach on how physiological measurements can be used to evaluate social acceptance can be found in Koay et al (2005). They developed a hand held device with which participants could indicate their subjective level of comfort towards a robotic agent. The measured distances were analyzed in combination with recorded video footage, indicating situations in the human-robot interaction where participants felt uncomfortable. It could be demonstrated that the device revealed certain uncomfortable states that would otherwise have remained visually hidden. These results support approaches using a methodological mix, as these states would have been missed without combining physiological data with behavioral analysis.

Moreover, more qualitative approaches have been used to study human-robot interaction in terms of social acceptance. Forlizzi (2007) e.g. conducted an ethnography in the home context, studying how people interact with a robotic vacuum cleaner. A methodological mix of semi-structured interviews and visual story diaries were used to gather behavioral data. When comparing the data material, it could be revealed that a robotic vacuum cleaner influences the social family life differently than a normal one. Furthermore, the author argued that the study indicates that simple social cues can positively influence the acceptance and adoption rate of robotic devices. Unfortunately, which cues could not be exactly specified through this study.

Stubbs et al (2007) conducted a two-year common ground analysis based on a behavior observation of a collaborative human-robot system to reveal factors disturbing the development of human-robot collaboration. Their data set consisted of the observational field notes and artifact documents, such as emails and robot plans generated by the human collaborators. Based on these results, they could show that mental models change depending on the autonomy level of the robot. However, this study was very specialized on a specific target group and how they experience their work with robots.

Another longitudinal study was conducted in the context of robot assisted therapy by Robins et al (2005) on interaction profile analysis of autistic children. They conducted qualitative and quantitative analysis of video material, which revealed a change in social interaction patterns of autistic children. Their study proved the value of long term studies on the social acceptance of therapeutic robots for children with autism.

The assessment of the factor acceptance thus seems to be relevant for HRI in different contexts (work, home, therapy) and above all, to support long-term relationships between the user and the robotic systems.

2.4.3 State of the Art on Evaluating User Experience

User Experience is a well known concept in HCI, which can be understood in three different ways. (1) UX is a concept going beyond traditional usability aspects, e.g. effectiveness and efficiency, by taking into account various aspects such as fun in the holistic experience of a system. (2) UX is related to positive emotions and affects during the interaction with an interactive system. (3) UX is related to the usage of the interactive system and thus embedded into specific usage situations (Hassenzahl and Tractinsky, 2006).

The perception of robotic systems in terms of a positive experience is of increasing interest in HRI. Fong et al (2003) remarked in their survey that a positive experience is of importance for robots in domestic settings, but also in the working context robots need to be experienced positively to guarantee a successful collaboration. According to Fong et al (2003) four different criteria for characterizing robots exist. Robots may have an (1) anthropomorphic appearance, which may result in projecting attributes to objects that cannot be fulfilled by a system. Thus, it seems reasonable to guarantee a balance of illusion and functionality. Another possibility of how robots can be designed is to design a (2) zoomorphic look that imitates living creatures to establish a human creature relationship. (3) Caricatured robots are designed on a fairly abstract visual level, which can be used to manipulate user's attention. (4) Finally, functionally designed robots clearly reflect their function by their design (e.g. service robots).

Considering the different visual styles of robots there are two different design theory approaches regarding embodiment (Fong et al, 2003). The first biologically inspired view postulates that robots should mimic the social intelligence of humans. The visual design should be "human-like" or at least "creature-like". This approach utilizes models of different disciplines to define the appearance of the robot, behavior, and technical design. It is grounded on the media equitation theory - humans are experts in interacting socially (Reeves and Nass, 1996). Thus, it seems reasonable to design systems which mimic human interaction habits. The other benefit can be seen in the possibility to test theories on which the initial design is based on (ethnology, structure of interaction, theory of mind, and developmental psychology). The second functionally designed approach implies that robots should be socially designed. Social behavior mechanisms are only described to convey specific social behavior, which is intended to fulfill a specific purpose.

The first thoughts on appearance of humanoid robots were made by Mori (1970). The well known "Uncanny Valley" proposed that if robots become too human-like they will become eerie (also see MacDorman and Ishiguro (2006)). Since then several studies on the appearance of robots in general (e.g. see Kaplan (2005); Kozima et al (2007)) and humanoid robots in specific (e.g. see Riek et al (2009); Bartneck (2008)) have been conducted.

The perception of humanoid robots is often closely related to human-oriented perception and the so-called "joint-intention theory". Several user studies have been conducted recently exploring how gesture and eye gaze direction can be intentionally used in collaborative settings with humanoid robots (e.g. see Mutlu et al (2009); Staudte and Crocker (2009)).

2.4.4 State of the Art on Evaluating Societal Impact

The societal impact of robotic agents in future societies can only be roughly predicted. The relevance of societal impact studies, however, lies in the identification of possible critical scenarios which can be avoided before the technological development of robots is mature enough to make them possible. Initial methodological approaches to assess the influence of robots for future societies will be presented.

One of the first studies on the societal impact of computing technology was conducted by Turn (1984), in which he considered the following impact areas as relevant: value systems (individual, group, national), individual rights and freedoms, environmental problems (pollution), demographic consideration (migration, density), economic aspects (employment, productivity), social issues (health, education, welfare) and institutional impacts (political, legal). In his study, Turn stressed that societal impact studies should be conducted with the purpose to identify potential problems of future societies (and recommend corrective actions) and not to analyze the current societal state.

Similar impact areas were identified as relevant in 1981 by the office of Technology Assessment in an exploratory workshop on "Social Impacts on Robotics". The aim of this workshop was "to examine the state of robotics technology and possible public policy issues of interest". Participants of this workshop proposed the following four areas of "social issues" as relevant for future societies in terms of the integration of robotic technology: (1) Productivity and Capital Formation, (2) Labor, (3) Education and Training, and (4) International Impact.

Additional considerations on a future society where robots are an integral part of everyday life have been made by Epstein (1994a,b). He asked the question of whose responsibility it is if a robot violates a human. A new perspective on this question can be found in Mudry et al (2008), who claim that the traditional responsibility ascription is biased by the fact that a role is assigned to autonomous robots that should be devoted to humans. Thus, a new ethical framework is needed which considers future scenarios from a new perspective. A purely theoretical approach to predict the future "robotic society" can be found in Kaplan (2005). He defined value profiles to visualize that robots are not everyday objects. Furthermore he defined necessary attributes for robots to be accepted as everyday objects in future society.

Mutlu and Forlizzi (2008) conducted an ethnographic study to analyze the possible impact of robots in the working environment. They investigated how an autonomous delivery robot influences the organizational structures of a hospital. They could show that the level of integration of the robot into the physical working environment and the work flow differed between working units. However, the authors themselves stated, that these are preliminary results that cannot be generalized for other kinds of situations, organizations, robot designs, and tasks performed by robots.

Sparrow and Sparrow (2006) conducted a survey to investigate "everyday claims" about future robotic technology usage. They concluded with a rather negative view on the future usage of robotic technology in elderly care, claiming that the usage of robots would reduce the number of social relationships. Moreover, they concluded that robots evoking an emotional attachment are unethical. However, they saw a positive tendency in scenarios where humans are not replaced by robots, but where robots support humans in everyday situations instead.

Similarly questionnaire based surveys are conducted to get insights on people's expectations towards the future society and the societal impact of robots. Arras and Cerqui (2005) conducted an impressive paper and pencil based survey with 2042 participants. The results testified a positive attitude towards potential robotic co-workers, flat mates or robotic body parts. The authors point out that approaches like this, with a sufficient number of participants, are valuable to gain insights in societal trends. They admit also that it is necessary to conduct similar surveys in different cultures and to compare the results. Furthermore, pure survey data actually gives an insight on people's attitude towards robots and not the actual ways interaction with them.

Another aspect of societal impact which is often addressed in HRI research are cultural differences in the perception of robots (e.g see Bartneck (2008); Bartneck et al (2006); Kaplan (2004)). This research is mainly concerned with the question if and why robots are differently experienced by people with Asians (particularly Japanese) cultural background compared to people with Western cultural background.

2.5 State of the Art on Design Guidelines for HRI

The fourth objective of this PhD thesis is to provide evaluation guidelines (see Chapter 6) derived from the case studies (see Chapter 4) which will trigger the future application of the proposed USUS evaluation framework. In the following, an overview on existing evaluation and design guidelines will be given with the main focus on how they were derived from empirical and theoretical work in HRI research.

One of the first set of guidelines generated for the evaluation of HRI was based on research experience in HCI and developed by Scholtz (2002). Her five guidelines should remind researchers to be aware of the points when evaluating the design of HRI. Experts in HRI (usability) should walk through a prototype of the user interface and examine its usability in accordance to the following guidelines:

1. Is sufficient status and robot location information available so that the operator knows the robot is operating correctly and avoiding obstacles?

2. Is the information coming from the robots presented in a manner that minimizes operator memory load, including the amount of information fusion that needs to be performed in the operators' heads?

3. Are the means of interaction provided by the interface efficient and effective for the human and the robot (e.g., are shortcuts provided for the human)?

4. Does the interface support the operator directing the actions of more than one robot simultaneously?

5. Will the interface design allow for adding more sensors and more autonomy?

Yanco et al (2004) also claimed the transfer of HCI and CSCW knowledge to the assessment of usability of HRI. They further developed the guidelines of Scholtz (2002) to tailored heuristics for usability evaluation and proposed an HRI taxonomy to characterize robotic interaction.

In the tradition of Yanco et al (2004), Marble et al (2004) also developed guidelines to increase the usability of human-robot interaction:

> "Like Yanco et al., we firmly believe that robotic systems must be designed with as much environmental, and task realism as possible. Furthermore, we also believe that formal, iterative system testing is the only route that ensures a system the supports, the capabilities and needs of the users."

However, the proposed guidelines only address usability (and only the indicators effectiveness and efficiency) and are only suitable for the testing of a single interface or architecture. The following nine points represent these guidelines:

1. Simplification of the environment to allow problem solution can corrupt the ability of the human-robot system to achieve its goals; therefore, the system must be tested in real world conditions to determine if it accurately meets these real world needs (Effectiveness).

2. Robotic systems will be effective only if the behaviors they use to achieve task goals are comprehensible for and predictable by the human team members; therefore, system design must assess how the human will work with the system (Effectiveness).

3. The test environment must reflect the complexities of the real-world environment in which it will be used (Effectiveness).

4. The test environment must incorporate uncertainty regarding the environment or the goal that will be seen in the true task (Effectiveness).

5. The task cannot be designed to exploit the capabilities of the robot; rather the robot's capabilities must be designed to exploit aspects of the environment and the task should emphasize the complexities encountered in the real world.

6. To accurately reflect the complexity of the task, testing must involve users who are similar to those who will put the system to actual use, not only those operators who are most familiar with the control architecture (Effectiveness).

7. Testing must incorporate the need for an operator to maintain a level of awareness in more than one environment (Efficiency).

8. Issues of teaming and the ability of the human to trust the robot enough for effective teaming must be addressed and assessed in the testing. (Effectiveness).

9. Task constraints may dynamically change with the incorporation of human-robot teams. However, these constraints may still shape how the human expects the system to behave (Effectiveness).

Additionally the following design guidelines for improved human-robot interaction developed by Drury et al (2004) mainly focus on aspects of usability evaluation, in specific awareness of human-robot interaction. Based on the first three guidelines, presented in the following, the authors even designed an interface prototype:

1. Enhance Awareness. Provide a map of where the robot has been. Provide more spatial information about the robot in the environment to make operators more aware of their robots' immediate surroundings.

2. Lower Cognitive Load. Provide fused sensor information to avoid making the user fuse the data mentally.

3. Increase efficiency. Provide user interfaces that support multiple robots in a single window, if possible. In general, minimize the use of multiple windows.

4. Provide Help in Choosing Robot Modality. Provide the operator assistance in determining the most appropriate level of robotic autonomy at any given time.

Bartneck and Forlizzi (2004) developed guidelines based on a framework of what makes a robot social, to improve the design of social robots. This design guidelines initially went a step further from pure usability guidelines to guidelines also taking into account aspects like user experience or social acceptance. They proposed the following three design guidelines for social robots:

1. The form of the social robot should match its abilities.

2. The social robot should mimic human-human dialogue in human-robot dialogue and be able to manage communication failures.

3. The robot should mimic human social norms and be able to provide a consistent set of behaviors.

The authors state that their guidelines are very broad, but "that more specific guidelines will be generated as robots are designed and built to address specific application areas". However, all the guidelines presented in this section, provided a good starting point for the development of the USUS evaluation framework. The evaluation and design guidelines which are derived by the validation of the proposed USUS evaluation framework will be shaped in a similar way. This will provide guidance for other researchers to apply the framework in future.

Chapter 3

The USUS Evaluation Framework

3.1 Outline of the Chapter

The main purpose of this chapter is to present the first proposed USUS evaluation framework based on the factors usability, social acceptance, user experience, and societal impact for Human-Robot Collaboration with humanoid robots. The USUS evaluation framework consists of a (1) Theoretical factor-indicator framework and a (2) methodological framework.

The description of the USUS evaluation framework will start with that of the theoretical factor-indicator framework. Each of the four evaluation factors will be defined and explained in detail. Next, the factors will be split up into several indicators, which are extracted and justified by literature review. Then the methodological framework will be described which is based on a method mix derived from various disciplines (HRI, HCI, psychology, and sociology). These various disciplines are proposed to address the factors and indicators during a human-robot interaction evaluation. The chapter concludes with a reflection on the development process of the theoretical and methodological framework and the limitations of this first proposed USUS evaluation framework.

3.2 The Theoretical USUS Framework

Evaluation activities of human-robot interaction scenarios often aim to explore interaction experiences more than conducting theoretically grounded studies, whose aim is to explicitly measure evaluation factors. Thus, the goal of the USUS evaluation framework was to develop a theoretical basis for evaluation actives combined with a proposed method mix allowing the systematic assessment of human-robot interaction scenarios.

The USUS evaluation framework for HRI with humanoid robots in collaborative settings consists of a multi level indicator model. In other words each factor is operationalized by several indicators to enable a holistic evaluation of collaborative work scenarios with humanoid robots. The theoretical factor-indicator framework was arranged based on an extensive literature review taking into account state-of-the-art HRI research on evaluation frameworks and methodological approaches (see Section 2.3, SubSection 2.2.2). However, it cannot be guaranteed that the theoretical as well as the methodological framework are exhaustive (see Section 3.8).

Figure 3.1 visualizes the combination of the theoretical factor-indicator framework with the methodological framework (see Section 3.7) to holistically assess HRI in a collaborative working environment. Chapter 4 of this thesis will present eleven case studies which were conducted in the framework of the EU-funded FP6 project "Robot@cwe: Advanced robotic systems in future collaborative working environments". By means of the case studies the chosen indicators should be justified and the suitability of the proposed method mix validated. The description of the theoretical and the methodological framework can also be found in Weiss et al (2009a).

3.3 Usability as Evaluation Factor

3.3.1 Definition of Usability

Usability is understood as the ease of use of an interactive system. However, usability is a concept consisting of several aspects which constitute a usable system. According to the ISO 9241-11 (1998), usability is "the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use." Preliminary research on usability in HRI focused mainly on performance issues such as, intuitiveness of control devices for robots, situation awareness of operators, feedback reaction time, and effectiveness and efficiency of the interaction in general (Drury et al, 2003; Scholtz, 2002; Marble et al, 2004). However, other studies indicate that further aspects should be taken into account when assessing the usability of a humanoid robot, e.g. learnability (Breazeal et al, 2004), flexibility (Song et al, 2007), robustness (Stubbs et al, 2007), and utility (Stubbs et al, 2007).

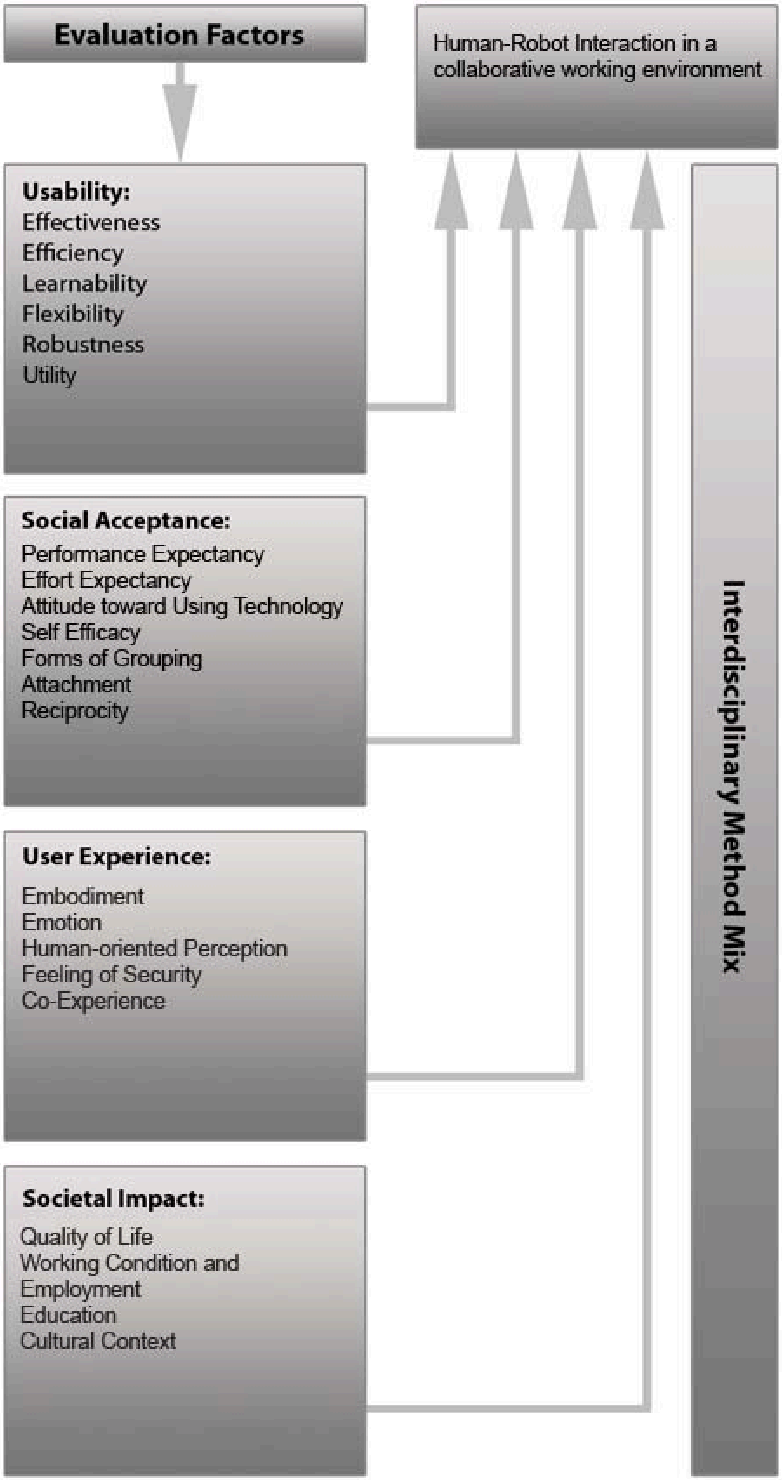


Figure 3.1: The Proposed USUS Evaluation Framework

3.3.2 Indicators for Usability

Effectiveness The ISO 9241-11 (1998) defines effectiveness as "the accuracy and completeness with which users achieve specified tasks". Thus, effectiveness describes how well a human can accomplish a task together with a robot. Effectiveness refers to the "success rate" with which a user carries out a task. "Error rate" and "task completion rate" can also be used to address this indicator.

Efficiency In the ISO 9241-11 (1998), efficiency is defined as "the resources expended in relation to the accuracy and completeness with which users achieve goals". Efficiency is the rate or speed with which a user can successfully carry out a task with a robot and is often measured by benchmarks like "task duration" or "required clicks".

Learnability The indicator learnability, as part of usability, originally evolved in conjunction with software engineering. The main focus is "ease of learning" meaning how quick and easy the usage of a system can be learned by a naive user. This seems to be a key aspect for successful HRI, as users normally have no previous interaction experiences with humanoid robots. The indicator learnability consists of several principles such as intuitiveness, familiarity, consistency, generalizability, predictability, and simplicity.

Flexibility Flexibility is another core-indicator for usability in HRI with humanoids, as these robots are task independently designed and thus various user models have to be considered in the interaction design. Flexibility is understood as the variety of options the user has to carry out a task together with an interactive system. Furthermore, flexibility describes the ease with which the user can modify the system to his/her personal requirements.

Robustness The indicator robustness addresses the fact that various errors can occur when a naive user interacts with a robotic system. To guarantee a good usability the user must have the possibility to correct some of his/her faults on his/her own. However, robustness also includes that the system is error preventing, stable, and responsive. A robust system ensures by a high level of support that the user successfully reaches his/her interaction goals with the robot without help.

Utility Utility describes the number of tasks the user can carry out with an interactive system. Furthermore, this indicator addresses the number of functionalities the system incorporates. Thus, utility refers to which options a system offers to solve specific tasks. As humanoid robots offer a huge variety of interaction possibilities utility was chosen as final usability indicator[1].

3.4 Social Acceptance as Evaluation Factor

3.4.1 Definition of Social Acceptance

User acceptance of interactive technology is defined by Dillon (2001) as "the demonstrable willingness within a user group to employ technology for the tasks it is designed to support". However, acceptance of autonomous systems is influenced by additional factors. As the literature shows, a general retention of robots can be observed for Western cultures (Hornyak, 2006; Kaplan, 2004). A broader view on acceptance, taking into account "social aspects", seems to become relevant for socially situated robots Fong et al (2003). For this PhD thesis the factor social acceptance is defined as "an 'individuals' willingness based on interaction experiences, to integrate a robot into an everyday social environment". In other words social acceptance in this PhD thesis can be seen as user acceptance of technology as a social actor. In order to avoid potential misunderstandings, social acceptance in this PhD thesis is not understood as the acceptance of a robot on the societal, but individual level.

For the evaluation of technology acceptance, several frameworks and models already exist (an overview can be found in Venkatesh and Davis (2000)). All models are based on a three step concept: (1) the individual's reaction on using information technology, (2) the intentions to use information technology based on the first reactions, (3) the actual usage of the technology.

3.4.2 Indicators for Evaluating Social Acceptance

The indicators for the factor social acceptance were set-up on the basis of two models: (1) the UTAUT model (Venkatesh et al, 2003) to address user acceptance aspects (indicators 1 to 4) and (2) the theory of "Object-centered Sociality"(Knorr-Cetina, 1997) (indicators 5 to 8) to address the issue that humanoid robots are perceived as social actors. This combination should assure the required broader view on social acceptance of robots, which are socially situated in working environments.

[1]The indicator utility was indirectly addressed in the case studies, as it is part of the study settings to define the number of relevant tasks in a Human-Robot Interaction scenario and the required functionalities to carry out the tasks.

Performance Expectancy According to the UTAUT model, performance expectancy is the strongest predictor for the usage intention. Venkatesh et al (2003) defines this indicator as“ the degree to which an individual believes that using the system will help him or her to attain gains in job performance.” However, for user studies addressing this indicator, it needs to be considered that gender and age could influence the expectancy of performance. For example, younger men are more task-oriented and therefore possibly indicate higher performance expectancy values than older men and women (Venkatesh et al, 2003).

Effort Expectancy This indicator addresses to which extent the user perceives a system's ease of use. In the UTAUT model, Venkatesh et al (2003) defines this indicator as “the degree of ease associated with the use of the system.” In other words, effort expectancy takes into account the users' believes and thoughts of the degree of effort, difficulties, complexity, and usage understanding. Similarly to performance expectancy, effort expectancy is affected by age and gender, along with experience. Younger women will more often expect an effort, especially at early stages of experience (Venkatesh et al, 2003).

Attitude toward Using Technology In the UTAUT model, Venkatesh et al (2003) defines attitude toward using technology as “an individual's overall affective reaction to using a system”. For the USUS evaluation framework, attitude toward using technology is defined as “the sum of all positive or negative feelings and attitudes about solving working tasks supported by a humanoid robot”(Weiss et al, 2009a). A cross cultural study conducted by Bartneck et al (2006) with Dutch, Chinese, and Japanese participants could already show differences in the attitude towards robots. They used theNegative Attitude towards Robots Scale (NARS) to investigate on people's attitude towards social, emotional, and general interaction with robots. Interestingly, the Japanese participants did not have a more positive attitude towards robots, which was contrary to the authors expectations.

Self Efficacy The indicator self efficacy is not included as a direct determinate in the UTAUT model (Venkatesh et al, 2003), but is included to the USUS evaluation framework due to its estimated relevance for successful human-robot interaction. Self efficacy describes the personal perception of a users' ability to reach a goal with an interactive system. Perceived self efficacy can thus be defined “as people's beliefs about their capabilities to produce designated levels of performance that exercise influence over events that affect their lives.

Self efficacy beliefs determine how people feel, think, motivate themselves, and behave. Such beliefs produce these diverse effects through four major processes. They include cognitive, motivational, affective and selection processes" (Bandura, 1997).

Forms of Grouping To form a group is a core element of human social behavior. The indicator forms of grouping describes the fact that humans define their own identity by sharing common characteristics with others and by distinguishing themselves from other groups. Knorr-Cetina (1997) could show that humans ascribe personal characteristics also to objects with which they often closely interact and somehow form a group with an object. The arising question is, will humans also show behaviors like that when cooperating with a robot?

Attachment Attachment as a psychological concept explains the bond a human infant develops to its caregiver (Bowlby, 1958). However, Knorr-Cetina (1997) could show that humans also develop an emotional bond to objects ("my favorite book"). In general, the idea of emotional attachment towards technology already found its way into HCI and HRI research (Kaplan, 2001, 2000; Wehmeyer, 2007). For the USUS evaluation framework, attachment is defined as an affection-tie that one person forms between him/herself and a robot. One main aspect of emotional attachment is that it endures over time. Thus, Norman (2004) explains emotional attachment as the sum of cumulated emotional episodes of users' experiences with a device in various context areas. These experiences are categorized into three dimensions: a visceral level (first impression), a behavioral level (usage of the device), and a reflective level (interpretation of the device).

Reciprocity In accordance to Gouldner (1960), reciprocity can be seen as a pattern of social exchange, meaning mutual exchange of performance and counter-performance and as a general moral principle of give-and-take in a relationship. In Human-Robot Interaction research, the indicator of reciprocity became relevant with the increase of social robot design (Kahn et al, 2004). Humans can quickly respond socially to robotic agents (see also Reeves and Nass (1996)), but the question arising is if humans also experience "robotic feedback" as adequate reciprocal behavior. Thus, in human-robot relationships reciprocity can be defined as "the perception of give-and take in a human-robot interaction scenario".

3.5 User Experience as Evaluation Factor

3.5.1 Definition of User Experience

User Experience (UX) is a multifaceted term from the research field of HCI where the community is still searching for a common definition (Law et al, 2008). An HCI definition of user experience which can be suitably adapted for HRI was coined by Alben (1996) who defines UX as

> “all aspects of how people use an interactive product: the way it feels like in their hands, how well they understand how it works, how they feel about it while they’re using it, how well it serves their purposes, and how well it fits into the entire context in which they are using it.”

This definition visualizes that the users’ experiences with an interactive system, are on one hand, related to the system itself and on the other hand, dependent on the usage context. According to Hassenzahl (2003) user experience is significantly influenced by the users’ interaction goals, intra-psychological dispositions, the environment, involved people, and the system itself.

3.5.2 Indicators for User Experience

In the research field of HRI, an increased interest in studying user experience indicators can be observed (see SubSection 2.4.3). For the context of cooperative working environments, a positive user experience is desirable as it will improve the overall usability of the human-robot interaction. Below, the user experience indicators for the USUS evaluation framework will be presented. The indicators 1 (embodiment), 2 (emotion) and 3 (human-oriented perception) derived from the survey of Fong et al (2003) to classify socially interactive robots and were chosen to particularly address the uniqueness of users’ perception of humanoid robots. However, the indicator 2 (emotion) is also a traditional HCI user experience indicator just as indicator 4 (feeling of security) and 5 (co-experience). Other classical user experience indicators from HCI research were intentionally not integrated, as they were substituted in bigger concepts (e.g. fun as part of emotion) or integral part of the social acceptance indicators (e.g. sociability vs. social acceptance, motivation vs. attitude towards robots, user involvement vs. reciprocity).

Embodiment The indicator embodiment takes into account the physical morphology of the robotic agent and its impact on the beliefs about and expectations of the robot (Fong et al, 2003). An assumption often found in HRI literature is that an anthropomorphic design will support the successful human-robot interaction, as expectations of social interaction will be invoked which are intuitive to naive users (see Breazeal and Scassellati (1999); Brooks (2002)). Other researchers are of the opinion that humanoids generate unrealistic expectations (or even fear) as they are no longer perceived as machines, due to their human features Dautenhahn (1999). Turkle (1984) as one of the first researchers that claimed that a clear distinction between human and machine is necessary for successful interaction.

Emotion Humans tend to interact emotionally with technology (Reeves and Nass, 1996) and even perceive, e.g. computers, as social actors. Due to that fact, the emotions robots invoke in humans are estimated as an essential evaluation indicator for HRI. Hassenzahl (2001) was one of the first HCI researchers who stressed the importance of emotions in interactive system design. He argues that relevant emotional episodes that are aroused during the interaction with a product are a key element of user experience. A broad range of emotions can be observed when an interactive system fulfills the users' expectations, like joy, pride, and surprise ad attraction. However, negative emotions can also arise when interacting with robots, like fear, boredom, and insecurity.

Human-Orientated Perception The indicator human-orientated perception takes into account the design approach in HRI where human-like perception is simulated to ease the interaction. Typical human-oriented perception features incorporated into the design of social robots are: tracking of the human face, interpreting human speech, and recognizing facial expressions (Fong et al, 2003).

Feeling of Security The feeling of security indicator takes into account the fact that when humans and robots directly cooperate in the same environment, safety and security issues arise (Dautenhahn et al, 2006). More importantly, the perceived safety of a robot influences the user experience. Studies have been conducted to investigate how to eliminate the risk of hazards in Human-Robot Collaboration (HRC) (Kulic and Croft, 2003). However, the main focus for the indicator feeling of security in the USUS evaluation framework is to assess the design of the HRI so that humans experience them as safe.

Co-Experience with Robots The term co-experienced was coined in HCI by Battarbee (2003) to describe experiences with interactive systems which come-up when individual experiences change as they become part of social interaction. Humans construct "situations through an interpretive process in which they take into account the non-symbolic gestures and interpretations of others" (Battarbee, 2003). For human-robot interaction this indicator comes into play when several humans discuss their interaction experiences with others or if the user experiences the robot as social actor, developing a social communication paradigm. Furthermore, co-experience with robots is of importance when the robot acts as mediator within a group of people (Severinson-Eklundh et al, 2003).

3.6 Societal Impact as Evaluation Factor

3.6.1 Definition of Societal Impact

Societal impact can be understood as every effect an activity has on social life. For the USUS evaluation framework societal impact is defined as follows:

> "Societal impact describes all effects the introduction of robotic agents causes for the social life of a specific community (taking into account cultural differences) in terms of quality of life, working conditions and employment, and education" (Weiss et al, 2009a).

In order to avoid possible misunderstandings, it is better to state at this point that the primary concern of the USUS evaluation framework is not technology assessment in its traditional notion, but to gain an insight on imaginaries about the "future robotic society". As technology development is not an independent factor, it impacts society regarding the beliefs, expectations, and attitudes of people. The imaginaries of potential users of robotic technology can help to shape robotic systems in a way to avoid negative assumptions. At the time when the societal impact can be retrospectively "measured", robotic technology already entered our society and its shaping becomes more difficult.

To define the indicators of societal impact, theoretical assumptions on future robotic societies in the cyborg and post-humanism literature were reviewed (e.g. Badmington (2000); Gray (2001)). The relevant areas of societal impact are, according to this literature: quality of life, health and security, working conditions and employment, and education.

3.6.2 Indicators for Societal Impact

Quality of Life, Health and Security Quality of human life is in accordance to Gray (2001) determined by several freedoms, e.g. freedom of speech, free choice of religion, and the freedom of travel. However, stability in human relationships (family constellations) and mutual reliance heavily influence the quality of human life. He argues that the integration of artificial intelligence and robotic technology will change the very nature of our relationships with each other, the human health system, and security aspects (electronic privacy, the freedom of consciousness, and the freedom of information).

Examples for these possible changes already exist, like robotic surgery (Schräder, 2005), robots in a therapeutic contexts (e.g. the Robodoc: (Dobson, 2004) or Paro (Wada and Shibata, 2006)), and robots in elderly care (Forlizzi, 2005). Similarly, examples can be found for robotic technology influencing security aspects of life: robots for surveillance (Robowatch, 2007) and robots in urban search and rescue (Drury et al, 2003).

Working Conditions and Employment This indicator takes into account "all aspects affecting how people carry out their job and how employers take care of their employees, including things like working contracts, wages, working times, and work organization" (Weiss et al, 2009a). Technological developments have always affected working conditions (increasing productivity and efficiency). However, the introduction of robotic and automation technology into working environments may lead to the replacement of humans, as robots can do some tasks more precise and quicker, e.g. harvesting (Green et al, 2008). Otherwise, robotic technology could probably fill the gap of lacking service personal (Forlizzi, 2005).

Thus, the replacement of human labor by robotic technology is one of the most crucial points for this evaluation indicator, which would have immense consequences for future society. According to Decker (2007) replacement has to be split up into three areas: technological replacement (robots will only replace technology if they can carry out tasks more efficient), economic replacement (cost - benefit calculations), and legal replacement (necessary new legal structures, regarding reliability aspects and consumer protection).

Education The integration of new technology into society is always accompanied by new types of education. Lifelong learning is a key term in this context. On one hand, the usage of robots needs to be taught. On the other hand, new skills need to be taught to reduce the amount of unemployed people due to robotic replacement. In other words, sufficient education and training must be offered to prepare people for the "robotic" labor market.

As autonomous interactive systems will be the next technological development entering our homes and working environments, it will be necessary to prepare people for the utilization in a physical manner, but also potentially in a psychological manner.

Cultural Context The term culture can be understood in accordance to (Badmington, 2000) as the whole range of practices, customs and representations of a society. However, culture does not exist in the abstract. On the contrary, it is in the broadest sense of the term textual, inscribed in the paintings, operas, sculptures, furnishings, fashions, bus tickets, and shopping lists which are the currency of both aesthetic and everyday exchange (Badmington, 2000). Cultural differences play an important role in human-robot interaction, as a difference can be observed in the acceptance of robotic systems in Western and Eastern countries (e.g. see Hornyak (2006)).

3.7 The Methodological Framework

The design of quality evaluation studies in HRI which produce reliable and reproducible results is a major challenge. For the evaluation of interactive systems a variety of different methods are available from various disciplines. In general it can be distinguished between five types of methods (1) task performance metrics, (2) behavioral/ observational measures (3) physiological measures (4) self-assessments, and (5) interviews. It depends on the required data, the available resources (time, money, prototype fidelity), and the research questions to choose, which method is best. A basic principle of user-centered design is to conduct expert evaluations, taking into account the users' point of view or user-based evaluations, where potential users assess a system regarding usability, acceptability, etc.

As the research field of HRI is still developing, there is a growing need of suitable evaluation methods. A primary issue in HRI research that could be derived from Chapter 2 is a lack of combining multiple methods of assessment to obtain convergent validity in HRI studies. Some evaluation methods already have been adopted and/or modified from other research fields as psychology and HCI, but their accuracy is still unclear.

It is important that the selected method is adapted to the evaluation context and the system capabilities. To evaluate the proposed theoretical indicators for usability (see Section 3.3), social acceptance (see Section 3.4), user experience (see Section 3.5), and societal impact (see Section 3.6); a combination of methods is proposed, where qualitative research is combined with quantitative measures for the evaluation approach.

Table 3.1 presents the methodological framework as the proposed method mix that was used in eleven case studies (see Chapter 4). The main goal of the methodological framework is to propose a suitable method mix to address several indicators in one study setting. Thus, it gives a general overview of combinations and synergies between the methods and indicators. In the following sections, the general application of the proposed methods are described. Each method has advantages and disadvantages which can be overcome by combining several methods. [2].

[2] A more detailed description of the concrete methodological set-ups will be given for each case study in Chapter 4.

Factor	Indicator	Expert Eval.	User Studies	Questionnaires	Physio. Measures	Focus Groups	Interviews
Usability	Effectiveness	X	X				
	Efficiency	X	X				
	Learnability	X	X				
	Flexibility	X	X				
	Robustness	X	X				
	Utility						X
Social Acceptance	Performance Expectancy			X			
	Effort Expectancy			X			
	Attitude toward using Technology			X		X	
	Self Efficacy			X		X	
	Forms of Grouping			X		X	
	Attachment			X		X	
	Reciprocity			X			
User Experience	Embodiment			X		X	
	Emotion			X	X	X	
	Human-oriented Perception			X			
	Feeling of Security			X		X	
	Co-Experience			X		X	
Societal Impact	Quality of Life						X
	Working Condition and Employment						X
	Education						X
	Cultural Context						X

Table 3.1: The Proposed Methodological Framework

3.7.1 Expert Evaluation

Application

In traditional HCI research, expert evaluations are used to assess a system in terms of its usability and detect as many usability problems as possible in a way that is less cost- and effort-intensive than user testing. A small team of experts in interface design examine the system and judge it in compliance with general usability principles. Therefore, expert evaluations are also often called "usability inspection methods".

Heuristic Evaluation A heuristic evaluation is intended to find and describe usability problems of a system on the basis of fundamental principles, so called heuristics (Nielsen, 1992). Heuristics are describing essential attributes that a system should feature to ensure that the user is able to perform a task within a specified context in an effective, efficient, and satisfying way. Such a heuristic evaluation is usually performed by a small team of interface experts, inspecting the system and comparing to what extent the principles have been adopted. All detected divergences and incongruities are defined asusability problems. Each evaluator compiles a list of usability problems and heuristic violations concerning the application.

Afterwards, the lists of all evaluators are compiled into one general list and the achievements are compared. A catalog with the total list of usability problems is generated. All experts then have to rank all problems according to their severity. The severity of a usability problem is a combination of three factors:

1. The frequency in which the problem occurs: Is it common or rare?
2. The impact of the problem if it occurs: Will it be easy or difficult for the users to overcome?
3. The persistence of the problem: Is it a one-time problem that users can overcome once they know about it or will users repeatedly be bothered by the problem?

Finally, the result of a heuristic evaluation is a complete list of all detected usability problems ranked according to their severity.

Cognitive Walktrough A cognitive walkthrough (Wharton et al, 1992) is conducted by at least two usability experts assessing the usability of a system based on predefined task structures. The expert evaluators try to imagine how a typical (potential) user would complete these tasks with the assumption of minimizing the cognitive load. Thus, the cognitive walktrough is primarily used to evaluate the usability of a system in terms of its learnability and how intuitive it can be used. During each task analysis, the experts ask themselves a set of questions for each subtask. Typically, four questions are asked. If one of these question, presented below, has been answered with "no", a usability problem is detected.

1. Will the user try to achieve the effect that the subtask has? Does the user understand that this subtask is needed to reach the user's goal?

2. Will the user notice that the correct action is available? For example, is the button visible?

3. Will the user understand that the wanted subtask can be achieved by the action? For example, the right button is visible but the user does not understand the text and will therefore not click on it.

4. Does the user get feedback? Will the user know that he has done the right thing after performing the action?

Advantages and Disadvantages

Expert evaluation methods to identify usability problems are commonly used in HCI research as they offer three main advantages: (1) they are cheaper in terms of resources than user studies, (2) they are faster than user studies, and (3) they can be used in a very early stage of system development. Thus, expert evaluation methods can provide the researcher with knowledge about minor and major usability issues and, moreover, recommendations of how to fix them. However, often problems can occur regarding the knowledge level of the expert evaluators or the difficulty in reaching domain experts. If expert evaluators are not part of the development team of a system, a further problem can occur where the expert is not aware of technical limitations. Finally, it is hard to assess the quality of the evaluation results the experts provide: are the relevant problems found or minor design issues.

3.7.2 User Studies

Application

Laboratory-based By means of empirical data user studies try to answer concrete research questions and hypotheses. In general user studies are based on representative tasks carried out by potential users. Participants are structurally observed during the task conduction and measures are recorded, such as task completion (effectiveness) and task duration (efficiency). The aim of user studies is to detect usability problems. The think aloud technique (i.e. participants are asked to comment on all steps they perform during the task) supports the data gathering in user studies and helps to better understand the reasons for the problem. Normally, user studies are audio and video taped to allow the researcher a structured post-analysis and interpretation of the study setting.

Field-based In general, the procedure of field-based studies is similar to lab-based user studies. However, they are not necessarily task-based and the main focus lies on behavior observation of users in a realistic environment. The methodological difficulty in field trials is, primarily, that disturbing influence factors, like ambient noise and bystanders etc., cannot be controlled and the observation often has to be conducted in a passive and unstructured manner.

Wizard of OZ Technique If during a user study, the robotic system is not fully autonomous, but a human "wizard" is acting behind-the scenes to generate the system response, it is called Wizard of Oz (WOZ) experiment. This technique was originally developed to test Natural-Langauge (NL), multimodal and (software-) agent-based applications with real users (Dahlbäck et al, 1993; Kelley, 1984). To allow user testing in very early stages of the prototype development, which can not be fully implemented at that stage, a human "wizard" enacts (or simulates) the system features in interaction:

> "Central to the methodology is an experimental which I call the OZ paradigm in which experimental participants are given the impression that they are interacting with a program that understands English as well as another human would. In fact, at least in the earlier stages of development, the program is limping along, only partly implemented. The experimenter, acting as 'Wizard', surreptitiously intercepts communications between participant and program, supplying answers and new inputs as needed." Kelley (1984)

This technique implies several advantages for user studies on human-robot interaction: (1) testing in early prototyping stages, (2) controlling of intrusive factors in the experimental setting, (3) controlling of safety and security issues during experiments, (4) simulation of relevant social cues or feedback modalities.

Advantages and Disadvantages

The use of user studies, in which task performance measures and behavioral measures are taken as evaluation data, are becoming more common in HRI studies. The pure measurement of task performance metrics only gives little insights on evaluation factors beyond usability, but can be well combined with questionnaires and interviews. Regarding field-based evaluation studies, the data analysis is an issue, as the data analysis and interpretation are more complex than in controlled lab-based settings. Field-based studies give more insights into the "real" natural usage of robotic systems than predetermined laboratory-based studies. WOZ studies allow a more comparable assessment of usability indicators and the simulation of specific context factors, however, when evaluating user experience and social acceptance a major disadvantage occurs: do we measure the user's perception of the robot or of the human wizard, who stands behind the scenes?

3.7.3 Standardized Questionnaires

Application

Questionnaires are self-reporting research instruments using a series of questions or items which have to be answered or rated by the participants of the survey. Standardized (or structured questionnaires) mainly consist of closed questions where participants can choose from a set of pre-defined answers. The aim is to gather statistically analyzable data by a large number of participants to enable statistically significant results. Later in this PhD thesis, (see Section 5.2) questionnaires will also be described, which were developed especially for validating the proposed framework.

NARS - Negative Attitude towards Robot Scale This questionnaire is based on a psychological scale to measure the negative attitudes of humans towards robots. It was originally developed by Nomura et al (2006), while in this PhD thesis the English translation of Bartneck et al (2005) is used and referenced. This questionnaire tries to visualize which factors are the reason preventing individuals from interacting with robots. The questionnaire consists of 14 questions, which have to be rated on a 5 point Likert scale ranging from "strongly disagree" to "strongly agree". The 14 questions build three sub scales: S1 = negative attitude toward situations of interaction with robots; S2 = negative attitude toward social influence of robots; S3 = negative attitude toward emotions in interaction with robots (the questionnaire can be found in the Appendix: Annex 1).

AttrakDiff This questionnaire was designed to measure beyond the usability (pragmatic quality) and the user experience (hedonic quality) of a system (Hassenzahl, 2003). The questionnaire is a semantic differential consisting of numerous antithetic word-pairs, e.g. "disagreeable - likable". All items have to be graduated by the participants on a scale ranging from the negative word pole to the positive word pole; the negative pole (in sense of the scale) is coded with "-3" and the positive with "+3". The results are presented in four different sub scales: "PQ = Pragmatic Quality of the System", "HQI = Hedonic Quality Identification", "HQS = Hedonic Quality Stimulation", and "ATT = Attractiveness". The pragmatic quality describes the usability of the system and indicates how successful users are in achieving their goals using the product. The Hedonic Quality Identification (HQI) scale describes to what extent the system allows the user to identify with it. Whereas, the Hedonic Quality Stimulation (HQS) scale is an indicator to what extent the system can support stimulating functions, contents, and interaction- and presentation-styles. Finally, attractiveness describes a global value of how the quality of the system is perceived by the participants. The hedonic and pragmatic qualities are independent from one another, and contribute equally to the rating of attractiveness (the questionnaire can be found in the Appendix: Annex 2).

SUS - System Usability Scale The System Usability Scale (SUS) was developed at Digital Equipment Corp. as a tool which allows a simple and quick standardized evaluation of system usability (Brooke, 1996). The questionnaire consists of ten statement-based items, giving a global view of subjective assessments of usability. The statements are rated by the participants on a 5 point Likert scale from "Strongly Disagree" to "Strongly Agree". The overall scale takes into account the effectiveness, efficiency, and satisfaction of participants with a system. Thus, the result of the questionnaire is a single number representing a composite measure of the overall usability of the system being studied, meaning that scores for individual items are not meaningful on their own. The range of this cumulative score is in between 0-100: "80-100: Users like the system"; "60-79: Users accept the system"; "0-59: Users do not like the system"(the questionnaire can be found in the Appendix: Annex 3).

The PANAS - Positive and Negative Affect Schedule The Positive and Negative Affect Schedule (PANAS) is a psychological questionnaire. It was developed by Watson et al (1988) to measure the affective state of participants. It consists of 10 positive affects (interested, excited, strong, enthusiastic, proud, alert, inspired, determined, attentive, and active) and 10 negative affects (distressed, upset, guilty, scared, hostile, irritable, ashamed, nervous, jittery, and afraid). Participants have to rate items on a 5 point Likert scale according to the strength of feeling the emotion, whereas 1 = "very slightly or not at all," and 5 = "extremely". The results of the questionnaire are two scale scores, one for positive emotions and one for negative emotions. These scores indicate how participants felt e.g. after a user study (the questionnaire can be found in the Appendix: Annex 4).

Advantages and Disadvantages

Self-assessment methods, like questionnaires are the most commonly used methods for evaluation in HRI studies. Self assessments can easily provide valuable information how participants perceived the interaction with a robotic system. However, the answers of the participants could be socially desired answers, e.g. how participants expect others to answer this question or what they expect that the experimenter wants to hear. Another issue is that the experimenter cannot immediately corroborate participants answers. In many cases, the questionnaires are used at the very end of a study. Moreover, the current mood can also influence participants response behavior. Thus, participants can only answer them on a reflective level. In other words, questionnaires give no direct insights to how participants feel exactly during the interaction on a behavioral level. This is why it is important to use different methods additionally, like task performance measures and physiological measures.

3.7.4 Physiological Measurements

Application

Physiological measurements are an alternative approach to self-assessment measures (e.g. questionnaires and interviews) to explore emotions. By means of measuring affect and arousal, emotions can be measured at the moment they evolve. Several methods can be used to measure the emotional state of a subject (Minge, 2005): electro-dermal activity (SCR- skin conductance response; SRR - skin resistance response); heart rate (ECG - electrocardiogram); size of pupils (eye tracking system); activity of specific regions of the face (observation of zygomaticus major and corrugator supercilii); etc. Thus, combining physiological measurement with questionnaires and interviews during a user study can support the assessment of user experience indicators in terms of hetronom and autonom identification of perception.

Advantages and Disadvantages

Physiological measures become more and more popular in the HRI research field. The main advantage of this methodological approach is that participants can hardly manipulate the activities of their nervous system. Furthermore, due to modern measuring equipment physiological measures offer a minimally-invasive method to determine arousal in-situ. However, some measures cannot be interpreted straightforward; arousal indicators need to be calculated. Moreover, confounds (health status, temperature in the room etc.) can heavily influence these measures. In general the interpretation of physiological data is difficult: Does the rise of arousal indicate cognitive load, stress, surprise etc.? Because of this, it is important to use additional qualitative data in the analysis process of physiological data.

3.7.5 Focus Groups

Application

Krueger and Casey (2000) define a focus group as a "carefully planned series of discussions designed to obtain perceptions on a defined area of interest in a permissive, non-threatening environment".

This method was originally designed as a marketing research tool and has been adapted for research in many fields, such as social sciences, but also in HCI and software development and evaluation settings. In general, focus groups are similar to group discussions. However, the difference is that focus groups are a more structured process to collect and interpret data. Participants trigger the topics during the discussion which is moderated by a trained leader. This enables the researcher to explore participants' beliefs, attitudes, thoughts, and feelings about topics and how they experience the usage of an interactive system. Focus groups only generate qualitative data, which can be used as valuable input for other instruments like questionnaires to produce quantitative generalizable results.

Advantages and Disadvantages

Focus groups can be relatively easy to assemble and allow a flexible and inexpensive way to investigate research questions. A main disadvantage is that focus groups only represent the opinion of a small group of people. It is highly important to select a representative sample for focus group discussions. Furthermore, good facilitation skills are needed from the leader to handle various types of people and ensure the contribution of all participants. In general it is the nature of the focus group method that it is open ended and cannot be entirely predetermined. However, focus group data can be used to inform the design of other methods, like user studies or questionnaire design.

3.7.6 In-depth Interviews

Application

In-depth interviews are "person-to person" discussions about a specific topic of interest. The method generates qualitative data about the interviewee's thoughts, feelings and beliefs. In the methodological framework, in-depth interviews are intended to be combined with user studies to gain insights about the societal impact indicators. Key characteristics of in-depth interviews are:

1. Open-ended Questions: Questions should be worded in a way excluding "yes" and "no" as answer possibility for the interviewee.

2. Semi-structured Format: The interview is based on interview guidelines consisting of predefined questions with up-coming questions based on information provided by the respondent that should be integrated into the flow of the interview as well. Therefore, a specific order should be avoided.

3. Seek Understanding and Interpretation: During the interview the interviewer has to seek clarity and understanding of the respondent's answers. The correct interpretation of the answers is the core challenge of successful in-depth interviews.

4. Conversational: In-depth interviews should be structured conversation, although, the role of the interviewer is more listener than talker oriented. The task of the interviewer is to lead smoothly and naturally from one topic (question) to the next.

For the validation of the framework, semi-structured interviews were conducted, where each guideline could directly be assigned to one of the indicators of the framework.

Expert Interviews Expert interviews are also a qualitative research technique aiming to discuss open ended questions, but with experts in a specific domain. Thus, the aim lies more in a knowledge transfer from interviewed experts (e.g. humanoid robot developers) to the researchers than in understanding beliefs and attitudes.

Delphi Studies A special way of conducting expert interviews is Delphi studies. In Delphi studies, several rounds of interviews are conducted with experts, whereas the interview guidelines are always based on the results of the pre-interview. The goal is to find a solution of a complex problem or to find a common understanding of a specific research topic.

Advantages and Disadvantages

Interviews can provide valuable in-depth qualitative data which may, for instance, not be gathered with questionnaires. However, using interviews as evaluation method also raises several issues. The response style of participants influences the gathered data. Participants can answer in an affirmative way, in a negative way, and in a socially desired way. However, well written interview guidelines and interviewer training can help to increase the amount of informative data.

3.8 Reflection on the Proposed USUS Evaluation Framework

The theoretical USUS evaluation framework is derived purely from a literature review. The baseline given as a starting point was that the evaluation framework should go beyond traditional usability studies and included the perception and acceptance of robotic systems and the impact of humanoid robots on a societal level. This pre-condition led me to the four main factors: usability, user experience, user acceptance and societal impact. Although, state-of-the-art research in the field of HRI was taken into account, it was relevant for me to include also psychological and sociological literature to develop a first theoretical basis. The indications coming from sociology, like social interaction with objects, includes aspects such as forms of grouping, reciprocity and attachment. This has led to my assumption that the traditional HCI concept of user acceptance has to be extended to "social acceptance". Starting from this assumption the next step was in finding indicators for user experience that go beyond HCI concepts as well, which led to the indicators embodiment and human-oriented perception. The review of the post-humanism literature finally enabled me to identify key aspects regarding the impact of robotic technology on a societal level.

In this regard the development of the theoretical USUS factor-indicator model was mostly about generating indicators that had to be solidified through arguments based on literature, without taking into account the interrelations between the indicators and the four main factors at this stage. I never assumed a strict separation between the indicators and not even between the factors. The proposed USUS evaluation framework as it is presented in this chapter was a starting point for exploring these evaluation criteria. To find out more about the proposed indicators and their interrelations, I performed eleven case studies (see Chapter 4) and a broader online survey (see Chapter 5).

However, to obtain information about the theoretical factors and indicators, a methodological framework needed to be proposed. My development of the methodological framework was based on my experiences regarding methods of social empirical research and HCI. My view on methods stems from the circumstance that studying, explaining, and understanding human social behavior is rather complex and a “wrongly” chosen method can easily lead to “artificial results”. Thus, I proposed a method mix and tried to limit this effect by counterbalancing the disadvantages of one method by the advantages of others, e.g. combining task performance measures with questionnaire data or questionnaires with physiological measures. Whereas, it seems that the methodological framework could be more easily validated than the theoretical one, I did not omit the issue that “classical” methods needed, in some cases, adaption and modification. Therefore, they can be reasonably used to assess human-robot interaction scenarios (which is described in the specific study settings in Chapter 4).

The evaluation framework presented in this PhD thesis and the evaluation studies conducted with it is explicitly explorative and can only be a starting point. It is in the nature of the USUS evaluation framework that it is limited in its application areas. The framework was developed specifically for humanoid robots in working contexts where direct contact interaction with humans is considered. If the usage of the USUS evaluation framework is intended by other evaluators, methodological considerations have to be taken into account, which are described in more detail in Chapter 6.

Chapter 4

Validation of the Framework: Conducted Case Studies

4.1 Outline of the Chapter

This chapter reports on the experimental setting, the methods used, the procedure, and the results of eleven case studies conducted in order to validate the proposed USUS evaluation framework (see Section 3.2). The case studies were conducted with the following three different humanoid robots (see Figure 4.1):

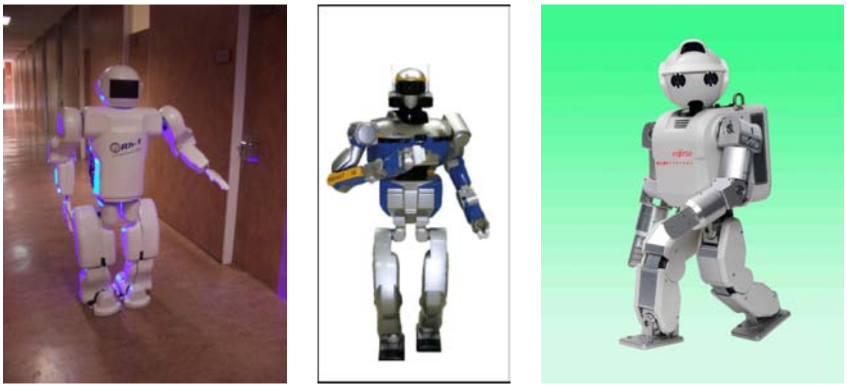
a RH-1: This robot was developed by the robotics lab of the Universidad Carlos III de Madrid (UC3M). It was used only in the first project year as testing platform. Later UC3M also worked with the HOAP-3 robot .

b HRP-2: This robot is the final version of the "Human Robotics Project" headed by the Manufacturing Science and Technology Center (MSTC). It is used by the Centre National de la Recherche Scientifique (CNRS) and the National Institute of Advanced Industrial Science and Technology (AIST).

c HOAP-3: This robot has been designed in order to provide portability. The height of the robot is 60 cm and weights only 8.8 kg. It was used by the Learning and Algorithms and System Laboratory (LASA) at Ecole Polytechnique Federale de Lausanne (EPFL). Since, the second project year, UC3M also used this robot.

All case studies presented in this thesis have two different view points: the content-based research goal (giving insights into the theoretical framework) and the validation of the methodological framework. Both sides will be presented for all case studies; however, the discussion of the results will mainly focus on the validation of the methodological framework.

Figure 4.1: The Three Robots



(a) RH-1 (b) HRP-2 (c) HOAP-3

4.2 Case Study 1: Focus Group about the HRP-2 Robot

This case study was carried out as two comparative focus groups stimulated with a video input and additional creative material ("Robot Building Set"). It was conducted in two focus group discussions moderated by discussion guidelines addressing user experience and social acceptance indicators. Furthermore, three standardized questionnaires were used to gather quantitative data as well. In each focus group, nine participants took part. It was decided to choose an uneven number of participants in order to get a clear majority in group decision processes. The case study was conducted at the ICT&S Center, University of Salzburg, February 2008. The content-based research question in this study was to explore: "Is the HRP-2 robot perceived differently depending on the fact if it collaborates autonomously or tele-operated with a human in terms of social acceptance and user experience?". The study can also be found in Weiss et al (2009h) and in Weiss et al (2009f).

Form the methodological viewpoint of validating the proposed evaluation framework, this case study can be seen as an initial activity to verify the indicators for the factors user experience and social acceptance by means of a qualitative approach and to close possible indicator gaps of the theoretical framework. Additionally, the qualitative results gathered in this study had been the starting point for the development of standardized questionnaires to address the mentioned factors in later case studies (see Section 4.4, Section 4.5, Section 4.6).

4.2.1 Study Setting

A 2 (group) x 2 (type of robot) between subject study was conducted, comparing the perception of the HRP-2 robot under two conditions (operation modes). Therefore, two groups (group I and II) with the same characteristics were recruited. Group I was presented a video showing the HRP-2 robot in tele-operated mode (condition I), whereas group II was presented the HRP-2 robot acting autonomously (condition II). The dependent variable therefore was "perception" and the independent variable was "operation mode" (tele-operated vs. autonomous).

The reason for choosing a focus group in this case study was the fact that it represents a fast and cheap qualitative method for gaining first-hand experiences of people's attitudes toward certain aspects of HRI. Moreover, attitudes towards working conditions are often influenced by group discussion processes in everyday life, meaning that when a humanoid robot would be introduced into a real working environment, employees would probably discuss acceptance and experience factors together.

Two focus stimuli were used: a video input and the "Robot Building Set". The first discussion stimulus was a video showing the HRP-2 robot collaborating with a human, which was presented to both groups. The only difference between the groups was that group I was shown that the robot is tele-operated, whereas group II thought it worked autonomously (see Figure 4.2). The duration of the videos was three minutes.

The second stimulus was the "Robot Building Set" which was used to get insight on the indicator embodiment. This creative stimulus consisted of a set of the most important parts of a robot feet, arms, head, torso) in three different variations (functional, human-like and animal-like design). Participants had the task to create a robot they would accept as working partner out of the completely mixed parts (see Figure 4.2). This was done after the focus group discussion.

4.2.2 Instruments

Discussion Guidelines The focus group discussion was moderated by a researcher using semi-structured discussion guidelines, where each question addressed one of the indicators which were intentionally investigated during the focus group (the complete guidelines can be found in Weiss et al (2009f) and in the Appendix: Annex 5). The whole discussion was protocoled.

Figure 4.2: Case Study 1: Focus Group Stimuli

(a) The Two Videos in Comparison

(b) The "Robot Building Set"

Questionnaires Two questionnaires where used during this case study to gather additional quantitative data: The NARS questionnaire and the AttrakDiff questionnaire (descriptions for all questionnaires can be found in Sub-Section 3.7.3). The Table 4.1 gives an overview of which methods were used to assess which indicator in this case study.

Table 4.1: Method Mix: Case Study 1

Factor	Addressed Indicator	Used Method
Social Acceptance		
	Attitude toward Using Technology	NARS
	Self Efficacy	Focus Group Discussion
	Forms of Grouping	Focus Group Discussion
	Reciprocity	Focus Group Discussion
User Experience		AttrakDiff
	Emotion	Focus Group Discussion
	Feeling of Security	Focus Group Discussion
	Co-Experience	Focus Group Discussion
	Embodiment	Focus Group Discussion
		Robot Building Set

4.2.3 Procedure

The study started with an introduction to the topic, then participants had to fill in the NARS. Next the video stimulus was shown to the participants, followed by filling out the AttrakDiff questionnaire to rate the user experience of the HRP-2 robot. Then, the group discussion about humans working together with robots was carried out, according to the semi-structured discussion guidelines. Participants ended the session by collaboratively designing a robot best suited for the task shown in the video by using the "Robot Building Set".

4.2.4 Results

Insights on the Theoretical Framework The results of this case study showed that the autonomous HRP-2 robot was differently perceived than the tele-operated HRP-2 robot regarding the indicator feeling of security. Participants of both conditions indicated to feel more secure with a human controlled HRP-2. Moreover, the questionnaire data (AttrakDiff) revealed that the tele-operated HRP-2 was perceived more positively with regards to its attractiveness and hedonic qualities. The mean values for the autonomous robot were: PQ: -0.19 ($SD: 1.09$), HQI: -0.50 ($SD: 1.40$), HQS: -0.36 ($SD: 1.14$), and ATT: -0.34 ($SD: 1.33$). The mean values for the tele-operated HRP-2 were: PQ: -0.01 ($SD: 0.09$), HQI: 0.70 ($SD: 0.53$), HQS: 0.35 ($SD: 1.00$), and ATT: 0.22 ($SD: 0.93$). The differences in the means of the scales, however, only were statistically significant for the HQI scale ($t(15) > 2.384; p < 0.05$).

Attitude towards technology: The data from the NARS questionnaire showed that participants had a very high interest in technology. Moreover, results of the NARS showed that participants were neither positive nor negative, but neutrally attuned to robots.

Self efficacy: Participants of both groups indicated that special education will be necessary for working with a robot. Thus, it seems that collaboration with robots does cause a lower self-efficacy in the personal abilities of handling it.

Forms of grouping: Comments made by the participants of both groups clearly show that robots are seen as tools, not as equal working colleagues. This tendency is also confirmed by the "Robot Building Set", as both groups created a functional designed robot.

Reciprocity: Corresponding with the fact that a robot is seen as a tool rather than an equal working colleague, participants stated that the robot as a technical device should not act socially, but that humans may be able to treat a robot in a social way.

Emotion: Participants of both groups were not afraid that a robot like HRP-2 could one day substitute human beings. However, they are afraid that future robotic technology could be used negatively, not only to support humans. Participants fear a future society lacking of interpersonal emotional relationships, if it will become possible to develop an emotional attachment to a robot.

Feeling of security: Both groups stated that they would feel more secure interacting with a tele-operated robot, but also mentioned that they could imagine getting used to an autonomous robot, as the fear of novelties is human-like.

Co-experience: Participants of both groups agreed that a robot cannot substitute a human working colleague. Thus, they stated they are unwilling to treat a robot like a human working partner.

Embodiment: All in all, participants are open to robots as working partners as long as there is a clear distinction between a human and a robotic working colleague. In other words, robots are designed functionally, that they are not treated as equal (human) working colleagues, and that they are not intended to substitute humans in a social (non-functional) way (see also Weiss et al (2009h)).

Insights on the Methodological Framework From a methodological view point, a video as stimulus for focus groups proved its value to gain first insights on the overall attitude of people regarding social acceptance and user experience for human-robot collaboration with a humanoid robot. It could be shown that a focus group, in general, is a reasonable approach to address the overall attitude of people regarding humanoid robots as working colleagues, particularly concerning different perceptions of the same robot depending on its operation mode. Similarly, the "Robot Building Set" proved its value to investigate the user experience factor embodiment in a creative way. Having both groups produce a similar result (both robots were functionally designed), indicates that the operation mode condition did not influence the decision process and this creative stimulus offers a simple way of visualizing embodiment preferences.

The main advantage of the method can be seen in the rich amount of qualitative data that can be gained regarding the indicators of user experience and social acceptance. This is of high value in the beginning of evaluation studies conducted by means of the USUS evaluation framework.

Two limitations of the proposed method are the type of robot and the task it conducts, which clearly influence the results of the study and that the sociocultural context. For other research topics (e.g. companion robots in elderly care), it needs to be further investigated if similar settings with different types of robots can lead to valuable results. To validate this methodological approach in terms of socio-cultural context, it will be necessary to conduct exactly the same focus group setting in different cultural environments (e.g. Western vs. Asian) to compare the results with a special focus, not only on the different perception of autonomous and tele-operated robots, but also on socio-cultural impacts.

This first qualitative case study revealed furthermore that even with a small sample size the AttrakDiff and the NARS questionnaire are reasonable instruments to address the factor user experience in general and the indicator attitude toward using technology in particular. Thus, the decision was made to use these two instruments in future case studies. The discussion material gained in the focus group should furthermore build the basis for the development of two pre-structured questionnaires: the UX-Questionnaire and the SoAc-Questionnaire (see Section 5.2), which were used in latter studies to directly address the proposed indicators of the theoretical framework. Moreover, as the participants often by themselves mentioned indicators of the societal impact factor during the discussion, it was decided to address these indicators in the following case studies by means of pre-structured interviews.

4.3 Case Study 2: Heuristic Evaluation of Three Robots

This case study was carried out by providing project partners of the FP6 EU-funded Project "Robot@cwe: Advanced robotic systems in future collaborative-working environments" via e-mail an instruction-manual and a problem list template to assess the usability of the HRP-2 and HOAP-3 robot and a simulation of the RH1 robot by means of a heuristic evaluation (more details on this method in general can be found in SubSection 3.7.1) according to the heuristics of Clarkson and Arkin (2007). For this case study, the heuristic evaluation method was adapted by means of video scenarios (more details can be found in Weiss et al (2009f) and Altmaninger (2008)). The instruction-manual can be found in the Appendix: Annex 6.

Nine expert evaluators identified usability problems in six different video scenarios (4 HRP-2 scenarios, 1 HOAP-3 scenario, 1 RH-1 scenario) and ten expert evaluators rated the problems in terms of severity. The content-related research question was "What are the major usability problems of HRP-2, HOAP-3, and RH-1 occurring in the conducted collaboration tasks?".

From a methodological view-point, the case study was conducted to find out which usability indicators can be reasonably addressed with an heuristic evaluation and to set-up a methodological variation. This methodological variation allows to conduct a heuristic evaluation with HRI experts, not usability evaluation experts, assessing a system that is not necessarily on-site for evaluation. The heuristic evaluation was organized by the ICT&S Center, University of Salzburg in April 2008.

4.3.1 Study Setting

The study was conducted via e-mail, meaning that the project partners received an instruction-manual (see Figure 4.3) how to conduct the heuristic evaluation and a template to fill in the problems found. A complete problem list of all usability problems for both robots was than compiled by the ICT&S Center, University of Salzburg. Afterwards, each expert evaluator conducted a severity ranking of all problems found.

Scenario	Collaboration: Raising Object
Problem Number	1
Problem Title	Missing Feedback
Detailed Problem Description	The human does not receive any understandable feedback when the robot notices the object, which should be carried together. Therefore the human does not start the task although the robot is ready to do so.
Heuristic violated by the problem	Visibility of system status
Video time stamp of the detected problem	

Figure 4.3: Case Study 2: HE Instruction-Manual

4.3.2 Instruments

The Heuristics The heuristics from Clarkson and Arkin (2007) were used. As these heuristics only proved their validity for on-site heuristic evaluations, it was decided to adapt the method with video material and scenario descriptions. Therefore, usability problems can be interpreted afterwards and ranked according to the severity by all expert evaluators.

The Video Stimulus The evaluators were asked to formulate and videotape a collaborative scenario in a narrative story including (1) the roles involved in the scenario (e.g. a human worker and a robot that should collaboratively raise an object), (2) the tasks and subtasks (e.g. task: raise an object together - subtasks: identify the object; giving feedback to the human, grab the object etc.), and (3) the setting (e.g. the robot is positioned in front of the human, the robot moves autonomously, the human needs to wear sensors). To visualizes the usability problems, evaluators additionally had to note the video time-code to the found usability problem. The Table 4.2 gives an overview of which methods were used to assess which indicator in this case study.

Table 4.2: Method Mix: Case Study 2

Factor	Addressed Indicator	Used Method
Usability		
	Effectiveness	Heuristic Evaluation
	Efficiency	Heuristic Evaluation
	Learnability	Heuristic Evaluation
	Flexibility	Heuristic Evaluation
	Robustness	Heuristic Evaluation

4.3.3 Procedure

When the expert evaluators received the instruction-manual and the template, they had two weeks time to conduct the heuristic evaluation. Afterwards they sent all material back to the ICT&S Center, University of Salzburg, where a complete problem list of all scenarios and found problems was compiled. Different from traditional heuristic evaluations, all problems had to be ranked in severity from all experts, which was possible due to the provided video material.

4.3.4 Results

Insights on the Theoretical Framework In total, six expert evaluators assessed six different human-robot collaboration scenarios. The most often violated heuristics were: “visibility of system status”(n=13), “appropriate information presentation”(n=9), and “flexibility of interaction architecture”(n=10). The first two heuristics combined covered 22 from all 59 heuristic violations. Interestingly, the heuristic “aesthetic and minimalistic design” was not violated according to the expert evaluators, indicating that HRI experts think usability problems of the assessed systems are not due to design flaws.

Regarding the severity of the usability problems, the evaluators rated one third of the problems as major usability problems or even usability catastrophe, and the remaining problems as minor usability problems.

Effectiveness: The lack of "visibility of system status" (this heuristic was violated 13 times out of the overall 59 violations according to the expert evaluators) indicates that an effective collaboration between a novice user and the evaluated humanoid robots can not be guaranteed. This assumption is also supported by the fact that the heuristic on "sufficient information design" was violated 6 times and the heuristic on "appropriate information presentation" 9 times.

Efficiency: According to the expert evaluators, the heuristic "synthesis of system and interface" was violated 7 times which leads to the assumption that an efficient collaboration between the human and the robot cannot be assured in the assessed scenarios.

Learnability: The heuristic on "use of natural cues" was violated 13 times indicating that novice users would have problems in collaboration, as the robotic systems evaluated are lacking on intuitiveness. However, the expert evaluators could not predict how this circumstance would actually affect the learnability of the robotic systems for novice users.

Flexibility: According to the expert evaluators, the heuristic on "flexibility of the system" was violated 10 times, indicating that there is still a lack in the task independent design for human-robot collaboration.

Robustness: The fact that the heuristic on "help users recognize, diagnose, and recover from errors" was violated 7 times, indicates that it would be difficult for novice users to learn the collaboration with the robotic system by trial and error.

Insights on the Methodological Framework From the methodological view point, this case study revealed that the proposed indicators for usability can be reasonably addressed with the described methodological variation of a video material supported heuristic evaluation. Additionally ,this case study proved that the heuristics of Clarkson and Arkin (2007), originally intended to be used for direct assessment of human-robot interaction, can be reasonably applied for video-based assessment of human-robot interaction. By presenting scenarios of human-robot interaction on video, an interdisciplinary evaluation team could successfully identify usability problems with varying severity in the interaction scenarios. As numerous severe problems could be found at an early stage in the development phase, this approach can be considered effective. Even a virtual interaction scenario was evaluated effectively concerning usability aspects.

The main advantage of evaluating video-based human-robot interaction scenarios is the fact that it is more economical than live interactions. The evaluation of video-taped interactions requires less effort compared to live interactions, as the scenario only has to be conducted once for the recording of the video material. Moreover, conducting an evaluation remotely and being independent from a location (i.e. evaluation is not limited to the research laboratory where the robot is developed) does not only save resources, but also does not predetermine the sample size of expert evaluators who can assess the scenario in parallel.

The main limitations in this methodological variation can be seen in the material which needs to be provided for the expert evaluators: (1) the instruction-manual and (2) the video scenarios. During this case study, a missing methodology understanding of the expert evaluators could be observed, which could not be supported by the instruction-manual. This had to be explained by individual e-mail answers. Thus, the instruction-manual needs to be written very detailed and precise to ensure that the expert evaluators, who have no previous knowledge about the methodology, perform the analysis in a comparable manner.

Another way of addressing this issue in future studies, would be to introduce a co-located moderator who supports the expert evaluators. Furthermore, the scenario visualization is a critical point: all action sequences need to be video-taped from the human's as well as from the robot's perspective to allow the identification of usability problems. For instance, two of the HRP-2 scenarios were only recorded from one camera perspective, which could be a reason why only minor heuristic violations were identified. Regarding the number of required evaluators, at least five or more evaluators should conduct the heuristic evaluation, as possible low return rates can be absorbed and the evaluator teams can be composed with a broader variety of research backgrounds.

However, only little insights on the indicator learnability could be gained by this methodology. Thus, it was decided to conduct an additional cognitive walkthrough based on the videos gathered during two of the laboratory-based user studies (see Section 4.10). This approach should enhance the methodological framework of the USUS evaluation framework to allow a quick assessment of the usability indicator learnability without conducting specialized user studies.

4.4 Case Study 3: First User Study with HRP-2

This case study was the first laboratory-based user study. It was based on two tasks, where participants had to control the HRP-2 robot via speech commands to pick up (task 1) and put down an object on another place (task 2). In this case study, indicators of all four evaluation factors were attempted, to have a holistic view on the theoretical framework. To measure usability, task completion rate and task duration were protocoled, as well as a "Retrospective Think Aloud" was conducted (as the user testing was based on speech commands, the participants should not be negatively influenced by the traditional "Think Aloud" technique). Furthermore, participants had to fill in the SUS questionnaire. User experience was addressed with the AttrakDiff questionnaire and an additionally developed pre-structured questionnaire on the user experience indicators. The social acceptance indicator attitude towards using technology was again addressed by the NARS questionnaire. The study concluded with a qualitative interview on participants' imagination of a future society with robots, to study the societal impact indicators. Four participants took part in this preliminary laboratory-based case study, which was conducted together with LAAS/CNRS (Laboratory for Analysis and Architecture of Systems), University Paul Sabatier in Toulouse France, July 2008. The study can also be found in Weiss et al (2009c) and in Weiss et al (2009f). The content based research questions of this user study were:

1. "How do novice user experience the collaboration with the humanoid robot HRP-2 when interacting via speech commands?"

2. "How do users perceive the system in terms of usability?"

3. "Does the general attitude towards robotics change because of the interaction with the robot?"

4. "How do people imagine the future society after interacting with the robot?"

From a methodological view point the focus of interest was to verify if the standardized questionnaires from other disciplines, as well as the self-developed UX-Questionnaire (see Section 5.2) indicators are reasonable to evaluate the proposed factors. Furthermore, it was necessary to investigate the workload of such a study setting for the participants and which methodological considerations have to be taken into account for future user studies.

4.4.1 Study Setting

This user study was based on two tasks (listed below), which participants had to conduct with the HRP-2 robot (see Figure 4.4). Four participants took part in the study, one female and three male. They were between 20 and 40 years old.The tasks were introduced by the following scenario:

> "Imagine you are working at a construction site and need a special tool. A humanoid robot, its name is fourteen, is supporting you by providing you with your needed tools. You can control the robot with predefined commands, therefore it picks-up and transports the tool."

Task 1 Tell the robot to pick up the orange ball. Therefore, you have to use the following commands. After each command, wait for the reaction (answer) of the robot, before you start with the next one.

1. "14" (the general command which activates the robot)
2. "Go to the green box"
3. "Look down"
4. "Take the orange ball"

Additional commands:

- "Turn to the left"
- "Turn to the right"

Task 2 Your task is to now tell the robot to put the orange ball on the table. The following commands are:

1. "14" (the general command which activates the robot)
2. "Turn to the left"
3. "Go to the yellow table"
4. "Look down"
5. "Put the orange ball on the yellow table"

Additional commands:

- "Turn to the left"
- "Turn to the right"

Figure 4.4: Case Study 3: Study Setting

4.4.2 Instruments

Usability Measurements The entire user study was video-taped. Task completion and task duration were recorded during the user study to address the usability indicators effectiveness and efficiency. Furthermore, a "Retrospective Think Aloud" was conducted and participants were asked about improvements and changes they would suggest to make the interaction easier in future.

Questionnaires The SUS questionnaire (for details see SubSection 3.7.3) was used to additionally address the perceived usability of the interaction. To get a general insight on the user experience of the robot, the AttrakDiff was used (see also SubSection 3.7.3 for explanation). Furthermore, a questionnaire was developed (subsequently called UX-Questionnaire) which addressed all five user experience indicators with five statements each. These had to be rated on a 7 point Likert scale (in style of the AttrakDiff) by the participants. The complete questionnaire and more details on its development can be found in Section 5.2.

For the final in-depth interview the following four questions were discussed with the participants, each addressing one of the indicators for societal impact.

1. In which way could robots be integrated into working life in the future? - (indicator: working condition and employment)

2. How could life change if robots are integrated into a construction site? - (indicator: quality of life, health, and security)

3. How will the usage of robots in the working context influence the future education system? - (indicator: education)

4. How could you imagine that society will support the use of robots in the future? - (indicator: cultural context)

Table 4.3 gives an overview of which methods were used to assess which indicator in this case study.

Table 4.3: Method Mix: Case Study 3

Factor	Addressed Indicator	Used Method
Usability		SUS Questionnaire Retrospective Think Aloud
	Effectiveness Efficiency	Task Completion Task Duration
Social Acceptance		
	Attitude toward Using Technology	NARS
User Experience		AttrakDiff
	Emotion Human-Oriented Perception Feeling of Security Co-Experience Embodiment	UX-Questionnaire UX-Questionnaire UX-Questionnaire UX-Questionnaire UX-Questionnaire
Societal Impact		
	Quality of Life, Health, and Security Working Conditions and Employment Education Cultural Context	In-depth Interview In-depth Interview In-depth Interview In-depth Interview

4.4.3 Procedure

The study started in a room separated from the test setting, where participants had to fill in the NARS for the first time. Next, the two tasks were conducted. After each task, participants were asked what they thought and felt during the interaction with the robot, if they had any problems during the interaction, and if they would like to change or improve something to make the task fulfillment easier. Participants then rated the difficulty of the task. After the two tasks were carried out, participants filled in the questionnaires in the following order: (1) SUS, (2) AttrakDiff, (3) UX-Questionnaire, (4) NARS (second round). After the questionnaires were filled in, the user study was concluded with the in-depth interview on participants' imagination about how future society could look like regarding the integration of robotic technology.

4.4.4 Results

Insights on the Theoretical Framework The results of this case study gave first insights how novice users experience the collaboration with the humanoid robot HRP-2 when interacting via speech commands. Besides gathering data on usability and user experience indicators, participants imagination of the integration of robots into the future society were collected and the attitude change towards robots after interacting with the HRP-2 robot could be revealed.

Usability: The SUS questionnaire equaled a result of 78.8, meaning that the participants accepted the HRP-2 robot in terms of its usability. However, the SUS is no absolute ratio, it depends on the experimental context. In this case study, participants evaluated the speech interaction with the robot. This positive perception of the usability could not be found in the data of the "Retrospective Think Aloud", as 18 negative statements could be extracted and only 11 positive ones.

Effectiveness: Task 1 was solved 4 times, but 2 of those times with help. Task 2 was solved 4 times, but 1 time with help. However, this task completion rate indicates a high effectiveness of HRP-2 for this specific task.

Efficiency: Task 1 was solved with a task duration mean of 03:53 minutes (SD : 104.42 sec; range: $02:23-5:36$) and task 2 revealed a task duration mean of 4:55 minutes (SD : 198.00 sec; range: $02:57-08:32$). As the ideal task duration time was set-up 3 minutes for task 1 and 5 minutes for task 2, this indicates a rather good efficiency.

User Experience: The general assessment of the user experience of HRP-2 by means of the AttrakDiff questionnaire revealed that participants experienced the robot rather heterogeneously, even though the mean values were all around 0 (above all the scale HQS shows a board SD). The ratings for the scales (reaching from -3 to +3) revealed the following mean values for the four scales: PQ: $0.00(SD : 0.38)$, HQI: $-0.54, (SD : 0.44)$, HQS: $0.04(SD : 0.62)$, and ATT: $0.07(SD : 0.25)$. The UX-Questionnaire (scale 1 to 7) revealed more detailed insights on the indicators for user experience: emotion ($mean : 5.85, SD : 0.3$), embodiment ($mean : 5.33, SD : 0.47$), feeling of security ($mean : 4.93, SD : 0.47$), human-oriented perception ($mean : 5.2, SD : 0.38$), and co-experience ($mean : 3.7, SD : 0.25$). These results are closely related to the results of the focus group.

Attitude towards Using Technology: The NARS questionnaire revealed significant decreases. When comparing the results of the two rounds, the "Negative attitude toward Situations of the Interactions with Robots" ($t(3) = 5.20, p < 0.05$) and the "Negative Attitude toward Social Influence of Robots" were both rated significantly lower ($t(3) = 3.43, p < 0.05$).

Societal Impact: When discussing robots in the future working life, three out of four participants used the term "replacement". Also, positive aspects were mentioned by participants, such as robots can be used in dangerous settings, but a negative future was generally predicted by participants: "What will be the value of human work in future?"(indicator: working condition and employment).

A similar tendency was seen by participants regarding the second question, how life will change because of robots. Two participants mentioned unemployment, the other two argued that work performed by robots will be more accurate (indicator: quality of life, health, and security).

Regarding the third question on the future education system, participants had two different concepts: (1) people have to learn the usage of robots in school,(2) children will be able to learn outside school how to use robots (indicator: education).

The last question participants agreed that in the Western world, the introduction of robots will be a cost-benefit issue, and people will have to get used to them (indicator: cultural context).

Insights on the Methodological Framework From the methodological view point, this user study proved the feasibility to combine qualitative and quantitative methods to gain insights on several USUS factors and indicators in one setting. Linking the data from the pre-structured questionnaires with the "Retrospective Think Aloud" and the societal impact interviews, facilitated the clustering and interpretation of the gathered data. Furthermore, it could be shown that the proposed SUS and AttrakDiff questionnaire can be fruitfully used to assess human-robot interaction scenarios with a humanoid robot.

A main advantage of this case study was that the NARS questionnaire can be reasonably used to investigate a change in attitude before and after interacting with a robot, even with a small sample size. Regarding the procedure to interview participants twice on their attitude with the NARS questionnaire, it is worth mentioning that participants experienced it as positive that they had the opportunity to fill in the NARS questionnaire twice: "That's good; I already changed my mind about robots". Similarly, the study duration and workload was experienced as acceptable for the study participants. In general, confronting novice users the first time with a robot in this type of study setting allowed to observe first time user reactions on humanoid robots and to find out if these reactions influence participants' attitude towards robots.

However, a major limitation of this case study was that participants experienced human-robot interaction, which was only based on speech commands, not as collaboration. All four participants saw the interaction with the robot more as working with a sophisticated tool than collaborating with a robotic system. Another limitation noticed in this case study was the problem of assessing the indicator feeling of security in a laboratory-based setting, since a human operator accompanied the study to take care of the robot. In settings like this, participants expect their safety as granted and have difficulties in answering questions regarding feeling of security.

A further, issue, which came up in this user study was that more efforts need to be invested into the recruitment of participants, primarily regarding the aspect of working with "novice" users who have no previous knowledge in robotics. In the "Retrospective Think Aloud" and the societal impact interviews, it became obvious that people heavily relate to their education and profession: "As I study computer science, I have a mental model how the vision of the robot works." "As I am working in education I can tell you that robots should not teach children!" However, the additionally developed UX-Questionnaire could not be validated according to its scales in this preliminary user study. As this could be due to the small sample size, it was decided to keep the questionnaire in its current form and validate it by means of a broader online survey (see Section 5.3).

Furthermore, the indicator emotion was detected as hard to assess only by the UX-Questionnaire, as this questionnaire only addresses how heavily the emotional level of participants was affected during the interaction with the robot. To distinguish between positive and negative feelings, it was decided to add the psychological PANAS questionnaire in the next user studies (see

SubSection 3.7.3 for details). As the SUS questionnaire revealed a rather different result than the “Retrospective Think Aloud” it was assumed that this questionnaire is highly context dependent and reflects only a few statements of the participants (“I never thought it is so easy, you just need to talk to the robot!”), thus it seemed reasonably to keep the think aloud technique to get a more holistic picture of occurring usability problems.

For one participant, the study lasted one hour in total, thus it was decided to add an additionally developed questionnaire on social acceptance, the SoAc-Questionnaire (see Section 5.2), to the following studies, as a study duration of 1.5 hours is experienced as capable for participants.

4.5 Case Study 4: First User Study with HOAP-3

The second laboratory based case study was a task based user study, which was conducted with the HOAP-3 robot. Participants had to conduct two “learning by demonstration” tasks with the robot. For the first task they had to teach the arm of the robot to (1) push a box, and for the second one (2) to close a box. Again it was tried to address all factors of the evaluation framework, even including more indicators than in the first laboratory based study. Usability was again measured by the task completion rate and the task duration, plus a “Retrospective Think Aloud” (also in this study the interaction with the robot was partly based on voice commands) and the SUS questionnaire (see SubSection 3.7.3).

Furthermore, participants had to fill in the AttrakDiff questionnaire (see SubSection 3.7.3), the UX-Questionnaire and the SoAc-Questionnaire (see Section 5.2), and give an interview on open ended questions regarding their imagination of a future society with robots as working partners. Eleven participants took part in this study conducted together with LASA/EPFL at Lausanne, Switzerland, August 2008. The study can also be found in Weiss et al (2009e) and in Weiss et al (2009f).The content based research questions of this study were the following:

1. “How do novice user experience the collaboration with the humanoid robot HOAP-3 when the interaction is based on learning by demonstration?”

2. “How do users perceive the robot in terms of usability?”

3. “Does the general attitude towards robotics change because of the interaction with the robot?”

4. “How do people imagine the future society after interacting with the robot?”

From the methodological view point of validating the evaluation framework, the focus was on producing comparable data to the HRP-2 user study and to fill methodological gaps, witnessed during the previous test. Furthermore, the case study should give insights if the instruments measure the same indicators for a different HRC scenario.

4.5.1 Study Setting

This user study was based on two tasks which participants had to conduct with the HOAP-3 robot (see Figure 4.5; Figure 4.6). The tasks were introduced by the following scenario:

> “Imagine you are working at an assembly line in a big fabrication plant. A new robot is introduced, which should support you in completing tasks. You can teach the robot specific motions by demonstrating them (meaning move the robot’s arm like you expect it to move it later on its own); the robot will repeat the learned motion. You can repeat this demonstration-repetition cycle as long until you are pleased with the result.”

Task 1 This task is to teach the robot to push this box from its working space into yours on its own. The task is split up into the following action sequences:

1. Show the robot the specific task card by putting it on the table in front of the robot (move it around until the robot recognizes it).
2. Demonstrate the robot to push the box with its right arm, by putting the box very close in front of the robot and moving its arm.
3. Let the robot repeat what it learned.
4. (If necessary) repeat sequences 2 and 3 until you are pleased with the way the robot pushes the box.

- The interaction with the robot is based on speech commands. Just follow the commands of the robot and answer to it with yes or no (or any other answer proposed by the robot).
- You only need to teach the right hand arm of the robot by moving its elbow.

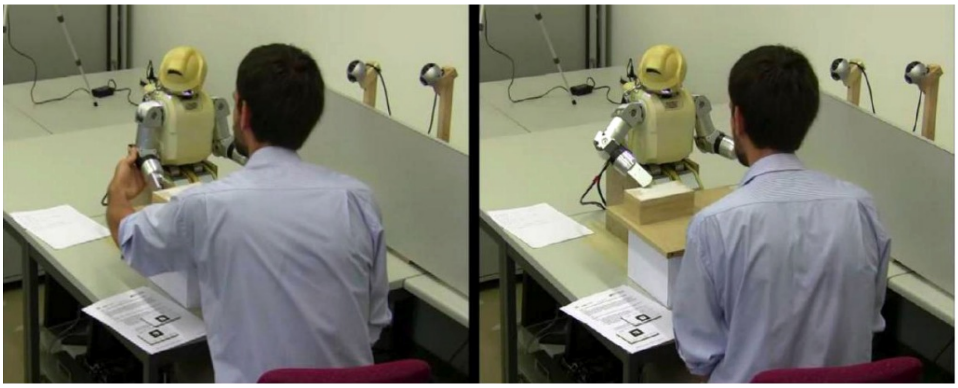


Figure 4.5: Case Study 4: Study Setting Task 1

Task 2 This task is to teach the robot to close this box on its own. The task is split up into the following action sequences:

1. Show the robot the specific task card by putting it on the table in front of the robot (move it around until the robot recognizes it).

2. Demonstrate the robot to close the box, by putting the box very close in front of the robot and moving its arm.

3. Let the robot repeat what it learned.

4. (If necessary) repeat sequences 2 and 3 until you are pleased with the way the robot closes the box.

- The interaction with the robot is based on speech commands. Just follow the commands of the robot and answer to it with yes or no (or any other answer proposed by the robot).

- You only need to teach the right hand arm of the robot by moving its elbow.

As the pretest of the user study showed that the tasks are experienced as different in its level of difficulty (task 1 was estimated more difficult than task 2) the order of the tasks was counterbalanced between the participants to reduce a potential learning effect. Twelve participants (8 females, 4 males) took part in this case study. The age range was from 16 to 40 years with an average of 26.58 years.

Figure 4.6: Case Study 4: Study Setting Task 2

4.5.2 Instruments

Usability Measurements To measure usability, task completion and task duration were recorded during the user study. Furthermore, a "Retrospective Think Aloud" was conducted.

Questionnaires The same questionnaires were used, as described for the preliminary HRP-2 test (see SubSection 4.4.2). Additionally, one validated questionnaire from psychology, the PANAS, was added to address the user experience indicator emotion with respect to positive and negative emotions (for details see SubSection 3.7.3). A questionnaire was also developed based on the indicators defined for social acceptance, subsequently called SoAc-Questionnaire. It was based on 32 items which had to be rated on a 5 point Likert scale by the participants. The complete questionnaire and more details on its development can be found in Section 5.2. Table 4.4 gives an overview of which methods were used to assess which indicator in this case study.

4.5.3 Procedure

The user study started with the NARS questionnaire, which had to be filled in by the participants before they saw the robot. Afterwards, participants conducted the two tasks with the robot. Following each task, they were interviewed for their thoughts and feelings during the interaction with the robot, suggestions for improvements to make future interaction easier, and they had to rate the difficulty of the task.

Next, the participants were asked to fill in the standardized questionnaires in the following oder: (1) PANAS, (2) SUS, (3) AttrakDiff, (4) UX-Questionnaire, (5) SoAc-Questionnaire, and (6) NARS (second round). After

Table 4.4: Method Mix: Case Study 4

Factor	Addressed Indicator	Used Method
Usability		SUS Questionnaire Retrospective Think Aloud
	Effectiveness Efficiency	Task Completion Task Duration
Social Acceptance		
	Attitude toward Using Technology	NARS SoAc-Questionnaire
	Performance Expectancy Effort Expectancy Self Efficacy Forms of Grouping Attachment Reciprocity	SoAc-Questionnaire SoAc-Questionnaire SoAc-Questionnaire SoAc-Questionnaire SoAc-Questionnaire SoAc-Questionnaire
User Experience		AttrakDiff
	Emotion Human-Oriented Perception Feeling of Security Co-Experience Embodiment	UX-Questionnaire PANAS UX-Questionnaire UX-Questionnaire UX-Questionnaire UX-Questionnaire
Societal Impact		
	Quality of Life, Health, and Security Working Conditions and Employment Education Cultural Context	In-depth Interview In-depth Interview In-depth Interview In-depth Interview

the questionnaires were filled in, the user study was concluded with the in-depth interview on participant's imagination about how future society could look like regarding the integration of robotic technology (the questions guiding the interview can be found in SubSection 4.4.2).

4.5.4 Results

Insights on the Theoretical Framework This case study could show that learning by demonstration based on direct contact interaction and speech commands is a valuable approach to bring novice users closer to humanoid robots. All participants, except two, could successfully carry out the teaching task with the HOAP-3 robot and all participants (except to other ones) were satisfied with the final repetition of the robot. Furthermore, the study showed that participants were willing to show the HOAP-3 robot several demos until the robot is able to repeat the learned motion in a satisfying way. Moreover, this user study could also reveal insights on the theoretical USUS factors and indicators, which can be found in the following paragraphs.

Usability: The SUS revealed a score of 65.83 for HOAP-3. This means that users accept the system. However, in this case study, the SUS addressed the demonstration of specific motions and interaction via speech commands, which has to be taken into account. This almost neutral perception of the usability of the HOAP-3 robot was also reflected in the think aloud data. A total of 66 statements (35 negative, 31 positive) on usability could be extracted and 17 suggestions for improvements.

Effectiveness: Task 1 and 2 were completed 11 out of 12 times. This task completion rate indicates a high effectiveness of HOAP-3 for these two tasks.

Efficiency: Task 1 was solved with a task duration mean of 04:43 minutes (SD : 176.82 sec; range: $01:59-10:45$) and task 2 revealed a task duration mean of 3:38 minutes (SD : 90.14 sec; range: $01:38-06:35$). As the ideal task duration time was set-up with 5 minutes for both tasks, results indicate a rather good efficiency.

User Experience: The general assessment of the user experience of HOAP-3 by means of the AttrakDiff questionnaire revealed that participants experienced the robot rather positively. The mean values for PQ ($mean: 0.8, SD: 0.71$), HQI ($mean: 0.75, SD: 0.63$), HQS ($mean: 1.22, SD: 0.91$), and ATT ($mean: 1.5, SD: 0.75$) were all around 1 (scale -3 to +3).The UX-Questionnaire (scale 1 to 7) revealed more detailed insights on the indicators for user experience: emotion ($mean: 5.45, SD: 0.74$), embodiment ($mean: 5.6, SD: 0.85$), feeling of security ($mean: 5.16, SD: 0.77$), human-oriented perception ($mean: 4.77, SD: 1.49$), and co-experience ($mean: 4.05, SD: 1.4'$). The PANAS questionnaire additionally revealed for the indicator emotion that 3 of the 7 female participants (one did not fill in the questionnaire) showed a more positive affect after interacting with HOAP-3 ($TP05 = 36, TP07 = 37, TP08 = 35$) than normal average (30.62). The other 3 female participants showed average values around 30. Regarding male participants, all except 1 stated a positive affect ($TP03 = 22, TP01 = 25, TP02 = 22, TP10 = 28$) lower than the normative average (32.06). The negative affect was lower than average for all participants except 1 woman ($TP09 = 23$; average: $16, 68$).

Social Acceptance: The SoAc-Questionnaire (scale 1: absolutely disagree to 5: absolutely agree) revealed more detailed insights on the indicators for social acceptance: performance expectancy ($mean : 3.34, SD : 0.50$), effort expectancy ($mean : 4.35, SD : 0.35$), attitude towards using technology ($mean : 3.76, SD : 0.61$), self efficacy ($mean : 3.67, SD : 0.38$), attachment ($mean : 4.22, SD : 0.52$), forms of grouping ($mean : 3.43, SD : 0.62$), and feeling of reciprocity ($mean : 4.09, SD : 0.39$). The results indicate that the user study communicated to the participants that the collaboration with humanoid robots would not cost much extra effort. It also communicated that they could develop an emotional attachment to the robot and feel like a team with HOAP-3. The NARS questionnaire revealed a significant decrease for the scale "Negative Attitude toward Social Influence of Robots" ($t(11) = 3.17, p < 0.05$). Participants rated this scale significantly lower after interacting with HOAP-3.

Societal Impact: When discussing the integration of humanoid robots into the future working life, participants were very positively attuned, mentioning that robots will conduct boring, repetitive tasks more precisely. Similarly, a future society with households robots for cooking, cleaning, and service robots for elderly care were predicted (indicator: working condition and employment).

A similar tendency was seen by participants regarding the second question, how life will change because of robots. Participants thought of increasing productivity and the possibility of humans to work in creative jobs if robots do the repetitive and dangerous tasks. Only some participants mentioned the concern of an increase of unemployment (indicator: quality of life, health, and security).

Regarding the third question on the future education system, participants had very heterogeneous ideas e.g.: (1) no extra education needed (2) people who have to work with robots will have extra trainings, (3) all children should learn about dangers and ethics of robots in school (indicator: education).

In the last question participants agreed that humanoid robots will be seen very positively in the future, if it was proven first that they will support the well-being of people. Furthermore, a better introduction for higher developed countries was predicted (indicator: cultural context).

Insights on the Methodological Framework From a methodological view point, this user study proved that the methodological framework applied to the preliminary HRP-2 user study was also feasible for a completely different interaction scenario, namely a learning by demonstration scenario based on interactive tutelage. Furthermore, this user study showed that the integration of the SoAc-Questionnaire to address all social acceptance indicators in one user study was a reasonable approach. None of the participants complained in the final interview about the study duration. The bigger sample size in this user study was of high value to identify more usability issues and gain a more divers response spectrum in the societal impact interview and the "Retrospective Think Aloud". However,the sample size was again not big enough to validate the additionally developed SoAc-Questionnaire according to its scales. It was decided to keep the questionnaire in its current form and validate it by means of a broader online survey together with the UX-Questionnaire (see Section 5.3).

A main limitation of this case study was that many of the participants were overcautious in moving the arm of the robot, as they did not want to break the robot. Several participants stated that they would have liked to get some basic technical background information about vision system and the degrees of freedom of the robot to better understand how to solve the task. It became obvious that participants did not like if things happened (e.g the robot could not see the box in front of it), which they cannot explain to themselves. This could be avoided if the robots capabilities are introduced in educational lessons beforehand. Thus, it was decided to give participants in the second HOAP-3 user study a short training phase, where they could observe how the robot processes commands given by the computer interface (see Section 4.6).

4.6 Case Study 5: Second User Study with HOAP-3

The third laboratory based case study was also conducted with the HOAP-3 robot. However, the main difference was that the interaction with the robot was remote-controlled via a computer interface, so this scenario provided no direct contact interaction with the robot. Participants had to conduct two tasks via the computer interface: (1) move the robot through a maze and find the exit, and (2) let the robot check all antennas and detect the broken one. This study is considered the same in relation to the instruments used and the procedure as the previous one. The only difference was the SUS questionnaire, which addressed the computer interface instead of the robot.

Twelve participants took part in this study conducted together with UC3M at Madrid, Spain, September 2008. The study can also be found in Weiss et al (2009f). The content based research questions of this study were the following:

1. "How do novice user experience the collaboration with the humanoid robot HOAP-3 when interacting via remote control?"

2. "How do users perceive the robot and the computer interface in terms of usability?"

3. "Does the general attitude towards robotics change because of the interaction with the robot?"

4. "How do people imagine the future society after interacting with the robot?"

From the methodological view point of validating the evaluation framework, the focus was on re-testing the methodological set-up of the previous user study with HOAP-3. Additionally, the focus was to investigate if the theoretical and methodological set-up is also reasonable for a remote-controlled HRC scenario.

4.6.1 Study Setting

This user study was based on two tasks which participants had to conduct via a computer interface with the HOAP-3 robot (see Figure 4.7). A total of 12 participants(6 females, 6 males) took part in this case study. The age range was from 16 to 26 years with an average of 23.75 years. The first task was introduced by the following scenario:

> "Your space shuttle has been hit by an asteroid and you were forced to an emergency landing. Your communication and internal ship monitoring system does not work, probably due to a damage caused by the crash. The good news is that you have the necessary material to replace the broken antenna for your communication system to send for help. As you are the only human survivor of the ship and the environment could possibly be dangerous for human beings, you decide to let this dangerous work be done by the ship's robot HOAP-3. At first you have to navigate HOAP-3 to the exit of the shuttle."

Task 1 Help the robot to find its way through the corridor and find the door to the outside. The task is to move the robot by means of the computer interface. It is completed if you see the door through the interface and say "door found". The interaction with the robot is based on a computer interface, with which you can control the robot. The second task was introduced by the following scenario:

> "After you have accomplished the first task to get the robot HOAP-3 out of the shuttle, you now have to help your HOAP-3 to find the broken antenna. The problem is that your shuttle has several antennas of different shape and color and you cannot distinguish the defected one from the others by sight, but HOAP-3 can. Inside the robot there is a mechanism which enables the robot to detect the malfunctioning parts."

Task 2 Your task is to control the movements of HOAP-3 again, while it is processing and checking the different antennas. In this task, you have to help the robot to recognize the broken antenna. If HOAP-3 has recognized a malfunctioning device it will put a square around it on the interface. The task is finished if you recognize the broken antenna through the interface and say "broken antenna recognized".

Figure 4.7: Case Study 5: Study Setting

(a) Study Setting: Participant (b) Study Setting: Maze (c) Computer Interface

4.6.2 Instruments

Exactly the same instruments were used as in the previous laboratory case study with HOAP-3 (see SubSection 4.5.2) However, in this case study participants had to fill in the SUS questionnaire with respect to the computer interface and not to the robot. Table 4.5 gives an overview of which methods were used to assess which indicator in this case study.

Table 4.5: Method Mix: Case Study 5

Factor	Addressed Indicator	Used Method
Usability		SUS Questionnaire Retrospective Think Aloud
	Effectiveness Efficiency	Task Completion Task Duration
Social Acceptance		
	Attitude toward Using Technology	NARS SoAc-Questionnaire
	Performance Expectancy Effort Expectancy Self Efficacy Forms of Grouping Attachment Reciprocity	SoAc-Questionnaire SoAc-Questionnaire SoAc-Questionnaire SoAc-Questionnaire SoAc-Questionnaire SoAc-Questionnaire
User Experience		AttrakDiff
	Emotion Human-Oriented Perception Feeling of Security Co-Experience Embodiment	UX-Questionnaire PANAS UX-Questionnaire UX-Questionnaire UX-Questionnaire UX-Questionnaire
Societal Impact		
	Quality of Life, Health, and Security Working Conditions and Employment Education Cultural Context	In-depth Interview In-depth Interview In-depth Interview In-depth Interview

4.6.3 Procedure

The user study started with the NARS questionnaire, which had to be filled in by the participants before they saw the robot. Then participants got to know the HOAP-3 robot, the movements it is capable of, as well as vision and navigation through the interface in a learning phase. Afterwards, the two tasks were performed by the means of a computer interface which enabled participants to see through the eyes of HOAP-3 and control its movements. After the tasks, participants were asked to fill in the standardized questionnaires in the same order than in the previously described case study (see SubSection 4.5.3).

This case study closed with a final interview on the societal impact indicators (the questions guiding the interview can be found in SubSection 4.4.2).

4.6.4 Results

Insights on the Theoretical Framework This case study could show that the participants perceived the interaction with the HOAP-3 robot via the computer interface as very professional, intuitive, and easy to handle. The participants considered "seeing through the eyes of the robot" as close cooperation. The main advantage of this human-robot interaction scenario was that it gave insights on both, the computer interface and the HOAP-3 robot.

Usability: The usability evaluation in this study mainly addressed the computer interface with which the participants had to control HOAP-3. The SUS revealed a score of 83.13 for the computer interface with which participants controlled HOAP-3. This means that users liked the interface for controlling the robot. This positive perception of the usability of the interaction with HOAP-3 robot via the computer interface was also reflected in the think aloud data. A total of 18 positive statements and 14 negative statements on usability could be extracted. Participants also gave 21 suggestions for improvements which mainly concerned additional functionalities for the computer interface, like including a map, 360 degrees of vision, including acoustic feedback etc.

Effectiveness: Task 1 was completed by 10 out of 12 participants (83.33%) and was rated with an avarage difficult level of 3 by the participants (on a scale from 1 = very easy to 5 = very difficult). The task completion rate indicates a good effectiveness of the computer intreface for solving this task with HOAP-3. Task 2 was also successfully carried out by all 12 participants and rated with an avarage difficulty of 1.58.

Efficiency: Task 1 was solved with a task duration mean of $07:12$ minutes (SD : 142.64 sec; range: $04:41-11:07$) and task 2 revealed a task duration mean of $0:55$ minutes (SD : 51.41 sec; range: $0:17-3:02$). As the ideal task duration time was set-up with 7 minutes for task 1 and 50 seconds for task 2. Results indicate an average efficiency.

User Experience: The general assessment of the user experience of HOAP-3 (seeing through the eyes of HOAP-3 and controlling it via the computer interface) by means of the AttrakDiff questionnaire revealed that participants experienced the robot positively. The mean values for PQ ($mean : 1.03, SD : 0.65$), HQI ($mean : 1.38, SD : 0.55$), HQS ($mean :$

$1.29, SD : 0.91$), and ATT ($mean : 2.13, SD : 0.76$) were all above 1 (scale -3: negative word pole to +3: positive word pole). Attractiveness was rated best and the pragmatic quality lowest, indicating that shortcomings of the computer interface did not negatively influence the experience of seeing through the eyes of HOAP-3.

The UX-Questionnaire (scale 1: absolutely disagree to 7: absolutely agree) revealed more detailed insights on the indicators for user experience: emotion ($mean : 5.68, SD : 0.73$), embodiment ($mean : 6.29, SD : 0.71$), feeling of security ($mean : 4.51, SD : 0.69$), human-oriented perception ($mean : 5.40, SD : 0.83$), and co-experience ($mean : 4.08, SD : 0.97$). Together with the results from the previous case study, it can be estimated that the embodiment of HOAP-3 is a primary relevant indicator for a positive user experience.

The PANAS questionnaire additionally revealed for the indicator emotion that three of the six female participants had a considerably more positive affect after interacting with HOAP-3 ($TP10 = 38, TP11 = 42, TP12 = 42$) than $normalaverage(30.62)$. The other three female participants showed average values around 30. Regarding male participants four stated a positive affect ($TP3 = 36, TP4 = 40, TP5 = 44, TP6 = 45$) higher than the normative average (32.06). The negative affect was lower than average for all male (15.2) and female (16.68) participants except two ($TP12 = 27, TP6 = 21$). These results indicate that HOAP-3 raises a positive affect.

Social Acceptance: The SoAc-Questionnaire (scale 1: absolutely disagree to 5: absolutely agree) revealed more detailed insights on the indicators for social acceptance: performance expectancy ($mean : 4.01, SD : 0.50$), effort expectancy ($mean : 4.08, SD : 0.65$), attitude towards using technology ($mean : 4.31, SD : 0.41$), self efficacy ($mean : 3.66, SD : 0.37$), attachment ($mean : 3.31, SD : 0.52$), forms of grouping ($mean : 3.62, SD : 0.48$), and feeling of reciprocity ($mean : 3.15, SD : 0.80$). The results indicate that the user study communicated to the participants that the collaboration with humanoid robots would not cost much extra effort. It also communicated that they could develop an emotional attachment to the robot and feel like a team with HOAP-3. Regarding attitude towards using technology, the NARS questionnaire revealed a decrease for all three scales, but only statistical significant for the scale "Negative Attitude toward Emotions in Interaction with Robots"($t(11) = 2.25, p < 0.05$). Participants rated this scale significantly lower after interacting with HOAP-3.

Societal Impact: When discussing the integration of humanoid robots into the future working life, participants thought of robots working at construction site or in dangerous and risky situations (space, ocean). Also, trivial tasks were suggested for robots like cooking, cleaning, and service robots. One participants also mentioned that robots could assist in medical settings (indicator: working condition and employment).

Regarding the second question, how life will change because of robots, 75% of the participants answered work will be safer and more precise. Participants thought work will be faster and the performance will be increased with robots. However, they also mentioned the fear of loosing their jobs. To minimize this risk, a suggested solution was to have human supervisors for the robots, aswell as having more humans work in the office in the future (indicator: quality of life, health, and security).

Regarding the third question on the future education system, participants agreed that young people will more easily learn how to interact with robots and thus they should be trained as soon as possible. Some participants thought of having "robotics" as a school subject. Specific expert knowledge on robotics could be taught at universities and trained at work places. One participant cautioned against using robots as teachers as "we do not have enough knowledge if robots are qualified teachers" (indicator: education).

The last question was answered very heterogeneously. Participants stated that robots will be accepted if they make everyday life easier, but that Hollywood movies deliver a negative picture of robots, meaning that marketing will be necessary to push robotic usage (indicator: cultural context).

Insights on the Methodological Framework From a methodological view point, this case study showed that the evaluation framework can also be reasonably used for HRI studies without direct contact interaction. Interestingly, the fact of seeing through the robot's eyes was experienced as close collaboration. Even though participants saw the robot only in the learning phase, its embodiment was rated highly positive, indicating that user experience factors can also be addressed in this kind of setting. Also, the usability results show that the usability indicators can be addressed with the same instruments by targeting the computer interface and not the robot itself. However, one has to be aware that the usability issues mentioned in this case study were related to the computer interface as interaction medium. For instance participants had problems controlling the robot through the maze because of the lack of the missing feedback provided by the interface. The participants generally liked that the interface provided little functionalities, but nevertheless allowed to solve the tasks in an individual and flexible way. This was a major advantage of this study setting.

A main limitation of this user study was the realization of the interaction scenario. Several participants complained that the interaction scenarios and the tasks were not realistic enough, as it was a space scenario and the environment did not simulate it enough. Similarly, the “antenna”, which had to be found in the second task, was represented by a ball, which confused some participants. Furthermore, some participants complained that the second task was too easy and that they only needed two clicks to solve it. The lesson learned for the HRC HRP-2 scenario (see Section 4.8) was that the environment and the objects in the scenario should match with the real world. An additional methodological lesson learned was that if a time line interaction analysis is needed, (click at the interface, movement of the robot) an analysis tool is needed that allows to synchronize and annotate two videos from different perspectives in parallel. Thus, for the HRC HRP-2 scenario (see Section 4.8), the analysis tool ELAN was used to allow a time line analysis of the video and heart rate data.

4.7 Case Study 6: Mixed-Reality Simulation of HRC

This case study was carried out as a preliminary Wizard of Oz user study based on a mixed-reality simulation. The simulation was based on a 3D model of the HRP-2 robot and implemented with the Crysis game engine (more details on the technical implementation can be found in Weiss et al (2009b)). The human-robot collaboration was based on the task of carrying and mounting an object together, whereas the object (a board) existed in the virtual as well as in the real world and built the contact point for interaction.

To evaluate the user study exactly the same instruments were used as for the two user studies with the HOAP-3 robot (see SubSection 4.5.2; SubSection 4.6.2). As the content-related research question was “How do differently simulated feedback modalities influence the human-robot collaboration”four different experimental conditions.

The four experimental conditions were: Con0, interaction without feedback, Con1, interaction with visual feedback (blinking light showing that the robot understood the command), Con2, interaction with haptic feedback, and Con3: interaction with visual and haptic feedback in combination were tested with 24 participating subjects (counterbalanced in age and gender). The study can also be found in Altmaninger (2008), Weiss et al (2009g), and Weiss et al (2009b). From the methodological view point, the question asked if the proposed theoretical and methodological framework is also suitable for user studies, which are not based on the interaction with an actual humanoid robot, but completely simulated. Furthermore, the interest laid in getting insights on methodological challenges for the final integrated scenario of human-robot collaboration (see Section 4.8). The study was conducted in Salzburg in the TV-studio of the University of Applied Sciences in August 2008.

4.7.1 Study Setting

This user study was based on one task which had to be conducted with the simulated robot via a mobile board as "input modality" (see Figure 4.8). The task was introduced by the following scenario:

> "Imagine you are working at construction area and get the task from your principal constructor to mount a gypsum plaster board together with a humanoid robot. You can control the robot with predefined voice commands, to carry out the following action sequences"

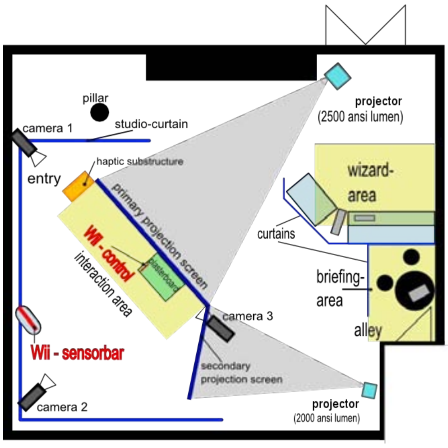
Task Lift, move, and mount a gypsum plaster board together with a humanoid robot. This task consists of the following action sequences:

1. Start the interaction by calling the robot
2. Lift the board together with the robot.
3. Move the board together with the robot to the right spot.
4. Tilt the board forward to the column together with the robot.
5. Tell the robot to screw the board.

4.7.2 Instruments

Exactly the same instruments were used as in the laboratory case studies with HOAP-3 (see SubSection 4.5.2 and SubSection 4.6.2). Table 4.6 gives an overview of which methods were used to assess which indicator in this case study.

(d) Study Setting: Participant



(e) Study Setting: Implementation

Figure 4.8: Case Study 6: Study Setting

4.7.3 Procedure

The user study started with the NARS questionnaire, which had to be filled in by the participants before they saw the robot. Participants then conducted the task with the robot. After the task, they were interviewed on their thoughts and feelings during the interaction with the robot, on suggestions for improvements to make future interaction easier, and they had to rate the difficulty of the task. Then participants were asked to fill in the standardized questionnaires in the following oder: (1) PANAS, (2) SUS, (3) AttrakDiff, (4) UX-Questionnaire, (5) SoAc-Questionnaire, and (6) NARS (second round).

4.7.4 Results

Insights on the Theoretical Framework This case study showed that different feedback modalities influence usability, user experience, and social acceptance indicators. It gave valuable insights, how the visibility of system status impacts the participants performance with the robot. It is noteworthy that after the study, participants of all experimental conditions stated that the prototype offered enough details, to reasonably fill in the questionnaires and answer the interview questions.

Usability: The usability evaluation in this study mainly addressed the differences regarding the four feedback conditions. The SUS revealed a score of 74.90 for the simulation, in general. With respect to the conditions, has the SUS the following values: $Con0 : 66.25; Con1 : 71.25; Con2 :$

Table 4.6: Method Mix: Case Study 6

Factor	Addressed Indicator	Used Method
Usability		SUS Questionnaire Retrospective Think Aloud
	Effectiveness Efficiency	Task Completion Task Duration
Social Acceptance		
	Attitude toward Using Technology	NARS SoAc-Questionnaire
	Performance Expectancy Effort Expectancy Self Efficacy Forms of Grouping Attachment Reciprocity	SoAc-Questionnaire SoAc-Questionnaire SoAc-Questionnaire SoAc-Questionnaire SoAc-Questionnaire SoAc-Questionnaire
User Experience		AttrakDiff
	Emotion Human-Oriented Perception Feeling of Security Co-Experience Embodiment	UX-Questionnaire PANAS UX-Questionnaire UX-Questionnaire UX-Questionnaire UX-Questionnaire
Societal Impact		
	Quality of Life, Health, and Security Working Conditions and Employment Education Cultural Context	In-depth Interview In-depth Interview In-depth Interview In-depth Interview

$83.93, Con3 : 79.17$. These values indicate that the second condition was experienced best by the participants in terms of usability, which is interesting, as this indicates that haptic feedback is perceived as most usable and no more improved by visual feedback. However, the differences were not statistically significant ($F(3, 20) = 2.30, p > 0.05$).

Effectiveness: Of 24 participants, 11 carried out the task successfully according to the ideal way of solution. The task was completed successfully by 10 participants, but with errors during single action sequences. Only 2 participants needed a hint on how to complete the task and only 1 participant aborted the task. The task completion rate indicates a good effectiveness of the mixed-reality simulation, in general. The number of errors ranged from 0 to 9, but no difference with respect to the feedback modalities could be identified ($F(3, 19) = 0.63, p > 0.05$).

Efficiency: The single task could be carried out by the participants in a task duration mean of $02 : 14$ minutes (SD : 62.86 sec; range: $01 : 05 - 05 : 20$). As the ideal task duration time was defined with 2 minutes for the task, an average efficiency is indicated. No statistically significant differences could be found for the task duration with respect to the feedback modalities ($F(3, 20) = 0.72, p > 0.05$).

User Experience: The general assessment of the user experience of the mixed-reality simulation, by means of the AttrakDiff questionnaire, revealed that participants experienced the construction site scenario positive. The mean values for PQ ($mean : 0.85, SD : 0.92$), HQI ($mean : 0.60, SD : 0.88$), HQS ($mean : 1.13, SD : 1.09$), and ATT ($mean : 0.71, SD : 0.99$) were all around 1 (scale -3: negative word pole to +3: positive word pole). Hedonic quality stimulation was rated best, indicating that the interaction with the mixed-reality encouraged the participants.

Significant differences could be revealed for the HQ-S scale ($F(3, 20) = 3.20, p < 0.05$) and the ATT scale ($F(3, 20) = 3.43, p < 0.01$). A post-hoc test (LSD) showed that Con3 (interaction with visual and haptic feedback in combination) was perceived significantly better in the hedonic quality of stimulation than all other conditions. Furthermore, the attractiveness was significantly better rated in Con3 than in Con0 (interaction without feedback) and Con1 (interaction with visual feedback).

The UX-Questionnaire (scale 1: absolutely disagree to 7: absolutely agree) revealed more detailed insights on the indicators for user experience: emotion ($mean : 5.03, SD : 0.62$), embodiment ($mean : 4.29, SD : 0.91$), feeling of security ($mean : 4.72, SD : 0.93$), human-oriented perception ($mean : 3.77, SD : 1.16$), and co-experience ($mean : 3.08, SD : 1.43$). The results indicate that the mixed-reality simulation, in general, caused a positive experience, above all in terms of emotions and feeling of security (clearly this indicator was positively influenced by the fact that the robot was not physically embodied). Significant differences could be revealed only for the indicator embodiment ($F(3, 19) = 3.35, p < 0.05$). A post-hoc test (LSD) showed that Con0 (interaction without feedback) was perceived significantly better in terms of embodiment than in Con1 (interaction with visual feedback). This indicates that the blinking yellow light on the robot's body negatively influenced its general embodiment.

The PANAS questionnaire additionally revealed for the indicator emotion that all participants except three male ($TP04 = 33, TP11 = 33$, $TP16 = 33$) and four female participants ($TP15 = 37, TP19 = 31$, $TP21 = 33, TP23 = 36$) had a lower positive affect than normative average (male: 32.06, female: 30.62), after the interaction with the mixed-reality scenario. However, none of the participants stated a worse negative affect after the interaction. These results indicate that the mixed-reality scenario caused a neutral affect.

Social Acceptance: The SoAc-Questionnaire (scale 1: absolutely disagree to 5: absolutely agree) revealed more detailed insights on the indicators for social acceptance: performance expectancy ($mean : 3.55, SD : 0.85$), effort expectancy ($mean : 3.75, SD : 0.64$), attitude towards using technology ($mean : 3.70, SD : 0.83$), self efficacy ($mean : 3.60, SD : 0.78$), attachment ($mean : 2.21, SD : 0.92$), forms of grouping ($mean : 3.10, SD : 0.91$), and feeling of reciprocity ($mean : 2.92, SD : 0.98$). The results indicate that a mixed-reality set-up is feasible to investigate general questions on the social acceptability of humanoid robots. Moreover, an ANOVA revealed that the four different experimental conditions influenced the results of three indicators: performance expectancy ($F(3, 20) = 5.84, p < 0.05$), forms of grouping ($F(3, 20)6.26, p < 0.05$), and attitude towards technology ($F(3, 20) = 6.10, p < 0.05$). A post-hoc test (LSD) showed that in Con1 (interaction with visible feedback) performance expectancy was rated significantly lower than in all other conditions. Forms of grouping was also rated significantly lower in Con1 compared to all other conditions. The same effect could be observed for the indicator attitude towards technology. The NARS questionnaire did not reveal any significant changes due to the interaction with the mixed-reality scenario in any of the three attitude scales.

Societal Impact: Participants mentioned numerous positive effects robots could have on the future quality of life, if they are integrated into the working life, e.g less physical harms, higher efficiency and productivity, and fewer mistakes in routine tasks. However, negative effects were discussed such as: higher unemployment rates, less companionship at work and ecological aspects like high energy consumption (indicator: quality of life, health and security).

Participants mainly imagined 3D works (dull, dirty, dangerous) for humanoid robots in the future. Other areas mentioned where robots could be integrated were elderly care, surgery, and assistance (e.g for handicapped people). Above all, robots were assumed to take over monotonous tasks. Humanoid robots were estimated as more trustful than other working technology. It is worth to mention that one participant had no idea at all how a robot like HRP-2 could be integrated in the future working life (indicator: working condition and employment).

Regarding education, participants assumed that a higher technology knowledge will be relevant in the future regardless to human-robot coworking. However, they mentioned that children will learn the handling of robots easily; therefore, teaching technology at school becomes more prominent. In a work setting, technology usage should be taught, according to the participants. Only one participant mentioned that it is too early to think about an education system change because of robots (indicator: education).

Participants stated that robotic usage will be supported by mass media in future and that it will be the task of mass media to define and distribute a positive image of robots. Furthermore, participants thought that robots will be accepted when robots are used as sophisticated tool and not if they replace human workforce. Also science and politics were mentioned as important influence factors for the integration of robots in the future society (indicator: cultural context).

Insights on the Methodological Framework From a methodological view point, it could be shown that a mixed-reality prototype Wizard of Oz user study can be reasonably assessed with the USUS evaluation framework. A mixed-reality prototype gave the participants the feeling of really interacting with a robot and thus allowed the assessment of the interaction by means of the UX- and SoAc-Questionnaire. Furthermore, the study setting offered a good discussion basis for the societal impact interviews. It should be mentioned that the development of this study setting was cost and time intensive compared to other prototyping methods.

Several issues were identified as crucial in order to successfully evaluate a mixed-reality a Wizard-of-Oz user study by means of the USUS evaluation framework. The mixed-reality prototype must be of a high fidelity and allow a high degree of realism for a meaningful evaluation study. One has to be aware that the findings might not be generalizable/identical for the final robot, but they allow to argue for one of the interaction techniques in the further design of the interaction scenario (if the goal is to improve user experience). From the technical perspective it was found that participants wearing glasses had problems to focus on details in the projections. A projector with 1600 x 1200 pixel and a light intensity of 3000 ANSI lumen could probably solve this issue.

For the HRC scenario with the HRP-2 robot (see SubSection 4.8.1), the mixed-reality prototype study furthermore informed the study setting in three aspects: First, one feedback cue should only consist of one information layer (e.g. a blinking light should not indicate: correct command processed, but only command processed, independently if it is correct or incorrect). Second, the amount of verbal commands the participants can use during the interaction with the robot should not exceed three. Third, the users should be offered support of the an experimenter next to them during the interaction to encourage the user to continue and not immediately abort if the first trial does not succeed.

4.8 Case Study 7: Second User Study with HRP-2

This case study was the final laboratory-based user study conducted in the framework of the EU-funded FP6 project "Robot@cwe: Advanced robotic systems in future collaborative working environments". Similar as the previousley described case study on the simulation of HRP-2 (seeSection 4.7), the task in this case study was to carry an object together with the robot, but the focus was to (1) investigate differences in the human-robot collaboration, when the robot was tele-operated or autonomous and (2) cultural differences in perception between Western and Asian participants. To evaluate the user study, exactly the same instruments were used as for the two user studies with the HOAP-3 robot ((see SubSection 4.5.2 and SubSection 4.6.2). Additionally, Heart Rate Variability (HRV) was introduced as physiological measurement. The content-related research questions were "How does the participant experience the relationship towards (1) the autonomous HRP-2 robot, (2) the tele-operated HRP-2 robot?" and "Is there a difference in terms of perception of usability, user experience, and social acceptance if the participants have (1) Asian nationality, (2) Western nationality?". From the methodological view point, this final study should prove that all factors and indicators can be addressed in one user study, by means of a broad method mix. A total of 12 participants (6 Asian, 6 Western) took part in this final case study, which was conducted together with CNRS/AIST at Tsukuba University, Japan and the Technical University Munich, Germany in October 2009. The study can also be found in Weiss et al (2009g).

4.8.1 Study Setting

This final case study was based on direct human-robot interaction, where the HRP-2 robot acted partly autonomously and partly remote controlled by an human operator located in Germany (OG). Participants (acting as the human operator in Japan, OJ) had to lift, carry, and put down a table collaboratively with the HRP-2 robot. During lifting and putting down the table, HRP-2 was tele-operated. During carrying the table, HRP-2 was walking autonomously. The study was based on the following scenario:

> "Imagine you are working at a construction site and you receive a task from your principal constructor: carrying an object from one place to another together with a humanoid robot which is partly acting autonomously and partly operated by a human expert operator. The principal instructor tells you the task via a computer interface."

Task The task is to carry a table from Place A to Place B together with the humanoid robot HRP-2. This task is split into 4 action sequences:

- Action Sequence 0: The robot is sent to the table by the principal instructor
- Action Sequence 1: Lift the table together with HRP-2
- Action Sequence 2: Walk together with HRP-2 from Place A to Place B
- Action Sequence 3: Put down the table together with HRP-2

The interaction between the human and the robot took place in three sequences 1-3 (see Figure 4.9). Sequence 0 does not require any interaction between the human and the robot.

(a) Action Sequence 1 (b) Action Sequence 2 (c) Action Sequence 3

Figure 4.9: Case Study 7: Study Setting

4.8.2 Instruments

Exactly the same instruments were used as in the laboratory case studies with HOAP-3 (see SubSection 4.5.2 and SubSection 4.6.2). Additionally, HRV was introduced as physiological measurement to get more insights about the factor emotional attachment.

Heart Rate Variability To gain more objective insights on the "emotional state" of the user during the interaction (not only insights on the affective state after the interaction by means of self reporting measures, see SubSection 3.7.3), the heart rate of the user (OJ) and the human operator (OG) was measured. Based on this data, an arousal feature was calculated based on HRV as a continuous metric (sympatho-vagal balance ratio) and synchronized it with the video data (see Figure 4.10). Table 4.7 gives an overview of which methods were used to assess which indicator in this case study.

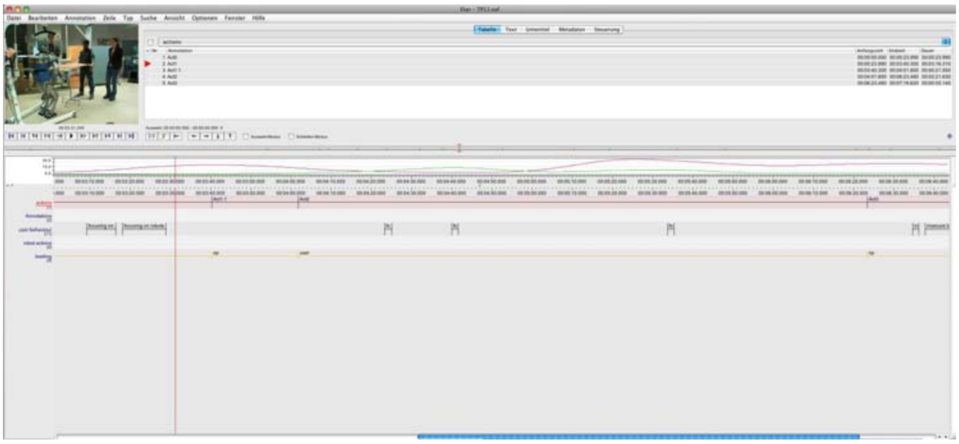


Figure 4.10: Case Study 7: HRV Data Analysis

4.8.3 Procedure

The final case study was conducted in the following manner: First, the test leader welcomed the participant. The participant (OJ) then received the "Suunto Memory Belt" (to record the heart rate) and was asked to put it on. At the same time, the human operator in Munich (OG) also put on such a belt. Next, the goals of the study as well as the test procedure were explained. After signing the video permission, collecting demographic data of the participant, and filling in the NARS questionnaire on the participant's general attitude towards robots, the participant was brought to the robot.

Table 4.7: Method Mix: Case Study 7

Factor	Addressed Indicator	Used Method
Usability		SUS Questionnaire Retrospective Think Aloud
	Effectiveness Efficiency	Task Completion Task Duration
Social Acceptance		
	Attitude toward Using Technology	NARS SoAc-Questionnaire
	Performance Expectancy Effort Expectancy Self Efficacy Forms of Grouping Attachment Reciprocity	SoAc-Questionnaire SoAc-Questionnaire SoAc-Questionnaire SoAc-Questionnaire SoAc-Questionnaire SoAc-Questionnaire
User Experience		AttrakDiff
	Emotion Human-Oriented Perception Feeling of Security Co-Experience Embodiment	UX-Questionnaire PANAS Heart Rate Variability UX-Questionnaire UX-Questionnaire UX-Questionnaire UX-Questionnaire
Societal Impact		
	Quality of Life, Health, and Security Working Conditions and Employment Education Cultural Context	In-depth Interview In-depth Interview In-depth Interview In-depth Interview

The participants then completed the tasks with the robot, and were interviewed regarding their thoughts and feelings during the interaction with the robot and their opinion towards the robot, depending if it acted autonomously or remote controlled. After the task conduction (all four sequences completed), both the participant (OJ) and the tele-operator (OG) put off the HRV belt.

The participants in Japan were asked to fill in the questionnaires handed out in the following order: PANAS, SUS, AttrakDiff, UX-Questionnaire, SoAc-Questionnaire, NARS (second round). Finally, the participants were interviewed regarding societal impact of robots in future societies.

4.8.4 Results

Insights on the Theoretical Framework This case study showed the high value of the theoretical guidance of the USUS evaluation framework as the research questions on differences in experience and acceptance due to cultural background were derived from it. During this case study it became obvious that Asian and Western participants have different problem solving strategies: Asian participants were very focused on the explanations of the researcher and tried to successfully carry out the task together with the robot at the first trial. Therefore they asked to repeat the instruction several times and tried to memorize them. Furthermore, Asian participants were more willing to adapt their behavior to the robot (e.g. walking speed) than Western participants. Western participants in comparison interacted more on a trial and error base with the robot and thus needed more trials. Often they also asked the research during the study to tell them the next step they have to perform.

Usability: The SUS questionnaire revealed a cumulative score of 47.08 for the HRP-2 robot. In general, this result indicates that users do not like the autonomous HRP-2 robot in terms of usability. In other words, participants perceived the interaction capabilities of the autonomous HRP-2 as improvable. This perception could also be found in the "Retrospective Think Aloud" data, as 14 positive and 25 negative statements concerning the scenario were identified. Furthermore, 9 suggestions for improvement of the scenario were given.

Effectiveness: Of 12 participants, 11 carried out the task successfully together with the HRP-2 robot. Only the trial with one participant needed to be aborted, as the interaction style of the participant was safety critical for the robot. The strong pull of the participant during sequence 2 could have broken the arm joint of the robot. The task completion rate indicates a high effectiveness of human-robot interaction scenario in general. The overall task difficulty was rated rather low by the participants ($mean : 2.50, SD : 0.29$), however, Western participants rated the task significantly more difficult than Asian participants ($t(10) = 3.06, p < 0.05$).

Efficiency: To assess the efficiency of the interaction with the HRP-2, the number of attempts and their duration in seconds was recorded for the action sequences 1 to 3 (as in action sequence 0 no interaction between the user and the robot happened). Action sequence 1 was furthermore split into sequence 1: human operator grasp table and sequence 1.1 human operator and user lift table. Table 4.8 gives an overview of the results for each participant. Regarding count and duration, significant differences could be revealed between Western and Asian participants. Western participants needed significantly more trials and subsequently equaled significantly higher task durations than Asian participants in sequence 1 ($t(10) = 3.35, p < 0.05$ and $t(10) = 3.33, p < 0.05$), sequence 1.1 ($t(10) = 4.02, p < 0.05$ and $t(10) = 2.86, p < 0.05$), and sequence 2 ($t(10) = 2.45, p < 0.05$ and $t(10) = 2.84, p < 0.05$).

Table 4.8: Case Study 7: Task Efficiency

Sequence 1		Sequence 1.1		Sequence 2		Sequence 3	
count	dur in sec	count	dur in sec	count	dur in sec	count	dur in sec
3	175,186,181	3	12,24,23	2	18,252	1	26
3	180,269,198	3	24,23,24	3	28,11,168	1	50
5	348,229,194,66,190	4	26,26,25,25	4	10,81,56,141,51	1	113
1	229	1	25	1	105	1	55
2	246,190	2	27,23	2	10,105	1	63
1	175	1	26	1	155	1	65
5	55,224,197,185,17	3	26,24,24	3	102,83,77		
2	274,177	2	24,24	2	10,200	1	61
2	197,204	2	26,27	2	16,293	1	52
2	193,186	2	13,23	1	295	1	71
1	196	1	22	1	142	1	55
1	194	1	25	1	121	1	64

User Experience: According to the AttrakDiff questionnaire, participants perceived the general user experience of the HRP-2 robot rather positive (scale -3: negative word pole to +3: positive word pole). HQS was rated best ($mean : 1.20, SD : 0.74$), which means that the participants were encouraged to interact with the HRP-2 robot. ATT ($mean : 1.18, SD : 0.93$) and HQI ($mean : 0.79, SD : 0.80$) were also rated positive, meaning that the participants perceived the robot as attractive and identified with it. The pragmatic quality (PQ) was rated lowest ($mean : 0.07, SD : 0.93$), meaning that the usability of the HRP-2 robot was experienced as rather neutral (neither very positive nor very negative), which goes in line with the results of the SUS questionnaire.

The UX-Questionnaire gave more detailed insights for the indicators of user experience (scale 1: absolutely disagree to 7: absolutely agree). Participants experienced the embodiment of the robot most positively ($mean : 4.22, SD : 2.03$) followed by emotion ($mean : 4.21, SD : 1.79$)

and feeling of security ($mean : 4.20, SD : 1.23$). The indicators rated lowest are co-experience ($mean : 3.88, SD : 1.23$) and human-oriented perception ($mean : 4.04, SD : 1.62$). Significant differences were found between Western and Asian participants, as Western participants rated the indicator emotion significantly higher than Asians ($t(7.67) = 2.53, p < 0.05$).

The PANAS questionnaire, additionally, revealed for the indicator emotion that only two of the six female participants had a more positive affect after interacting with HRP-2 ($TP01 = 41, TP5 = 34$) than normal average (30.62) The other female participants showed average values lower than 30. Regarding the male participants all except one stated a positive affect ($TP04 = 36, TP07 = 44, TP08 = 42, TP10 = 36, TP11 = 35$) higher than the normative average (32.06). The negative affect was rated higher than normative average (15.20) by three male participants ($TP04 = 16, TP07 = 17, TP11 = 23$) and one female participant ($TP = 20$, average: $16, 68$). Total, the male participants were more positively affected by the interaction with HRP-2 than the female participants. Female participants were less negatively affected than the male participants. The interaction with HRP-2 resulted in very little negative affect for the female participants.

Regarding the HRV, peaks on the user side are mostly highest during action sequence 2, second highest in action sequence 1 as soon as the user is involved in the lifting (the arousal increases during the user watches the robot while grasping the table). On the human operator side, the highest peaks could be observed during action sequence 1 when the operator had to grasp the table. Also a high, but mostly constant level of arousal could be observed on the operator side during the collaborative walking in action sequence 2.

Social Acceptance: The SoAc-Questionnaire (scale 1: absolutely disagree to 5: absolutely agree) revealed more detailed insights on the indicators of social acceptance. The indicator attitude towards technology ($mean : 3.99, SD : 0.37$) is rated highest, followed by the indicators facilitating conditions ($mean : 3.71, SD : 0.34$), effort expectancy ($mean : 3.70, SD : 0.59$) and forms of grouping ($mean : 3.64, SD : 0.52$). The indicator feeling of reciprocity was rated lowest ($mean : 3.18, SD : 0.59$), followed by attachment ($mean : 3.23, SD : 0.69$), self efficacy ($mean : 3.40, SD : 0.60$), and performance expectancy ($mean : 3.41, SD : 0.66$). The analysis of the NARS questionnaire revealed a decrease in all three scales through the interaction with the HRP-2 robot. However, only the scale “Negative Attitude toward Social Influence of Robots” decreased significantly ($t(11) = 2.88, p < 0.05$).

Societal Impact: Regarding the question how to integrate robots into the future working life, most of the participants again mentioned repetitive, boring, and exhausting tasks as well as dangerous and risky jobs. As working areas for robots mainly the household (cleaning tasks) and care settings (children, perts, elderly). However, robots were also mentioned as workers in factories, where they can work 24/7 and lift heavy objects. The areas where human labor will still be needed in future mentioned by the participants are creative and complex work. However, the integration of robots into working life will increase productivity, efficiency and safety according to the participants (indicator: quality of life, health and security).

When discussing possible changes in life due to the integration of robots into a construction area, most participants mentioned safety aspects like lower accident rates, working under extreme conditions etc. In other words, a higher level of safety at the workplace was seen as the major effect of robots which are integrated into a construction area. Additionally, the increase of productivity, effectiveness and efficiency was mentioned again. A decrease of costs was mentioned (as robots need no wage) and that life will be more convenient due to robots. However, risks were also mentioned by the participants, like e.g. that a human could be hurt during the collaboration with a robot and that unemployment will increase. A main factor, only mentioned by Asian participants, was that working tasks will shift from humans to robots and that robots will need to understand humans’ feelings in order to take care of human co-workers (indicator: working condition and employment).

Regarding the question if changes in the future education system are necessary, most of the participants affirmed this question. Most participants mentioned that it will be necessary to learn how to control and use robots. Therefore, different education measures will be required in school (robotic lessons in primary school) as well as at work (training courses). Mainly Western participants stated that it will not be necessary to teach children how to use robots as they will grow up with them regardless. Asian participants claimed that the goal should be to have intuitive robots where no collaboration training is needed at all. Some participants pointed out that rules (ethics) regarding how to interact with robots or robot usage conventions are necessary for the future (indicator: education).

When discussing the question with participants how society could support the usage of robots, Western people argued that there is already preparation through media and literature, but that they also assume society will rarely support robots. According to one Japanese participant, developed countries like Japan are early adopters of robots, whereas central Asia and Africa will adopt robots later. Furthermore, Western and Asian participants mentioned that a clear distinction between a human and a robot will be important and that guidelines on how to use robots and changes in law are necessary in a future robotic society. The fact that people will not easily accept robots in society is also mentioned often by Western participants. Japanese participants mainly suggested advertising the benefits of robots in order to increase people's acceptance, as helpful robots doing dangerous tasks or supporting disabled people are accepted more easily. According to Western participants, the costs of a robot, especially the cost-benefit proportion, will influence the acquisition of robots (indicator: cultural context).

Insights on the Methodological Framework From a methodological view point, it could be shown that even with a small sample size interesting aspects regarding in human-robot interaction between Asian and Western participants could be observed. Furthermore, HRV could prove its feasibility to gain more insights about the arousal of the user during the interaction with the robot. This information could be used in future as direct input for the robot to inform the interaction, e.g. if the arousal reaches a specific indicator value the robot stops approaching the human. Also the comparison of the user experience and social acceptance in the interaction depending if the user was interacting with the autonomous system or the human operator gave valuable insights in the "Retrospective Think Aloud" data and influenced the data of the societal impact interviews.

4.9 Case Study 8 and 9: The ACE Field Trials

Two field trials were also evaluated in the framework of the FP6 EU-Project "Robot@cwe: Advanced robotic systems in future collaborative working environments". These field trials were conducted with the mobile path finding robot Autonomous City Explorer (ACE). In both field trials only the indicators of the social acceptance factors were addressed and the societal impact of the robot was observed. These field trials were conducted bearing in mind the concept of a sociological breaching experiment (see also Weiss et al (2008, 2010)), where the first reactions of pedestrians were analyzed by means of passive unstructured observation. Furthermore, the experiment was combined with a street-survey addressing the social acceptance indicators. Both studies were conducted together with the Technical University of Munich in public spaces in Munich, Germany. More details on the first field trial can be found in Weiss et al (2008) and Weiss et al (2009f).

The participants for the street survey were randomly chosen to answer the questionnaire on a voluntary basis. A total of 48 participants took part in the first study conducted in July 2007 and 52 participants in the second one in September 2008. The content related research questions for both studies were: (1) "Does the introduction of a robot into a public place violate human norms of social behavior?" and (2) "Does the ACE robot offer the necessary navigation, interaction, and communication features to stimulate a feeling of social acceptance among the pedestrians". Form a methodological view point the main question was if a field trial setting is suitable to assess the indicators proposed for evaluating Social Acceptance. More details on the second field trial can be found in Weiss et al (2008), Weiss et al (2010), and Weiss et al (2009f).

4.9.1 Study Setting

In the first study (see Figure 4.11) the ACE robot moved remote-controlled via the Karlsplatz, a highly frequented public place in Munich, which is situated at the end of a shopping street, with access to local transportations (metro). Although the robot was remote-controlled for security reasons, the illusion of an autonomous system was preserved as the operator was hidden from pedestrians. Three researchers accompanied the experiment: one conducted the unstructured observation and two the interviews. The study lasted two hours.

In the second study (see Figure 4.11) the ACE robot had to move autonomously from Odeonsplatz to Marienplatz and back by asking pedestrians for the way. The pedestrians could tell ACE where to go by first showing in the right direction with an arm and then show ACE the right way (e.g. how far from here) on the map on its touch-screen. The development team of the robot stayed near it because of security reasons, but stayed invisible from pedestrians because of the well frequented environment. The study lasted five hours and was accompanied by four researchers, of whom one did the unstructured passive observation and the other three did the interviews.

In both settings, participants had the possibility to interact with the robot via its touchscreen (in the first study to get more information about the robot itself and on the Karlsplatz, in the second study to show the robot the way on the map). In the first study, 48 participants took part. In the second study 52 participants took part.

Figure 4.11: Case Study 8 and 10: ACE Field Trials

(a) Case Study 8 (b) Case Study 9

4.9.2 Instruments

Observation To asses the first reactions of pedestrians noticing or interacting with the ACE robot, passive unstructured observation was used. In both studies, a researcher took observational notes of the setting in general and the interaction with the robot in specific. In the first study, all researchers, additionally, wore a "SenseCam", which is a wearable digital camera that passively takes photographs or videos without user intervention while worn, to recored contextual information. In the second study, instead of observing the context it was decided to mount an additional camera on the "head" of the robot to directly record the interaction of participants with the robots touch screen

Extra Questionnaires For both studies, special questionnaires on the indicators for the "social acceptance" factor were developed, which were adapted to the field trial setting, meaning that the interviewing situation did not last longer than five minutes and the questions were related to the interaction possibilities with the robot (both questionnaires can be found in the Appendix: Annex 7, they are both in German). In the first study, two questionnaires

Table 4.9: Method Mix: Case Study 8 and 9

Factor	Addressed Indicator	Used Method
Social Acceptance		
	Attitude toward Using Technology	Extra Questionnaire
	Performance Expectancy	Extra Questionnaire
	Effort Expectancy	Extra Questionnaire
	Self Efficacy	Extra Questionnaire
	Forms of Grouping	Extra Questionnaire
	Attachment	Extra Questionnaire
	Reciprocity	Extra Questionnaire
Societal Impact		Observation

were used, one for people who did not interact with the robot via its touch screen and an extended one for those who did. In the second study, only one long questionnaire was used to gather more data material which was answerable for all participants and allowed comparative data analysis depending if participants interacted or not. Table 4.9 gives an overview of which methods were used to assess which indicator in this case study.

4.9.3 Procedure

The first field trial lasted two hours and was based on a passive unstructured participant observation and pre-structured interviews with participants selected by chance. The ACE robot was remote controlled, the human operator used three different navigation strategies repeatedly during the case study:

1. The ACE robot approached pedestrians in the distance of the intimate space. The robot introduced itself with a female voice and then asked the pedestrians if he/she wanted further information about: (1) the ACE robot, (2) the Karlsplatz, (3) the Karlstor (important monument), (4) news, (5) Technical University Munich, and (6) the weather. The pedestrians could select one of those options on the robot's touch screen.

2. The ACE robot navigated on the Karlsplatz without a direct goal in the distance of the personal and social space.

3. The ACE robot stood still in the middle of the Karlsplatz and started approaching pedestrians who showed interest towards it.

In the second field trial, the ACE robot was to moving autonomously from Odeonsplatz via Theatinerstraße to Marienplatz and back. During this navigation task, the ACE robot asked pedestrians for the way to either Marienplatz or Odeonsplatz. The pedestrians could show the ACE robot the way by first pointing in the right direction and then providing further information over a touch screen (e.g. how distant the goal was). The second field trial lasted three hours and was also based on passive unstructured observation and pre-structured interviews. Additionally, a camera directly installed on the head of the ACE robot recorded the participant's interactions on the touch screen.

4.9.4 Results

Insights on the Theoretical Framework In the following, the observational results and the questionnaire data of both field trials will be shortly described. The social acceptance indicators were formed by summative joining of the scores for the questions that represented it (see Appendix: Annex 7 for the items and the corresponding indicators). It needs to be mentioned that not all social acceptance indicators proved their consistency, as some indicators did not reach a Cronbach's alpha above 0.5 (first field trial: forms of grouping and cultural context, second field trial: forms of grouping and self efficacy). A low Cronbach's alpha value around 0.5 can be seen as sufficient, according to Pedhazur and Schmelkin (1991), if a high number of factors are tried to be assessed with a small number of participants.

Attitude toward using technology: In the first field trial, participants rated the indicator attitude rather positive ($mean : 3.88, SD : 1.37$). The observational data revealed the behavior pattern of "investigating the engineering of ACE" six times, which could be observed only for male pedestrians. In the second field trial, participants rated the indicator attitude toward using technology, in the street survey very positive ($mean : 3.96, SD : 0.13$). This indicator was rated significantly better by participants who actually interacted with ACE ($t(41.49) = 2.54, p < 0.05$). The analysis of the observational material showed that people were very curious towards new technologies and many people stated surprise in a positive way "It is able to go around me. I would not have thought that". Unfortunately, some people also seemed scared. However, most of the time curiosity prevailed over anxiety, which is probably due to the so called "novelty effect".

Performance expectancy: In the first field trial, participants rated the indicator performance expectancy very positive ($mean : 3.85, SD : 1.40$). In the second field trial participants rated the indicator of performance expectancy in the street survey the highest ($mean : 3.97, SD : 0.14$).

Furthermore, it could be observed that people tried to understand what the robot is doing by adopting its perspective. Thereby, they saw the robot accepting a subordinate role and taking care of the humans and the environment ("he is broadly going around us", "he is reacting to obstacles"). However, the image of robots seems to be characterized by Hollywood-films ("I wonder if this is like Wall-E"). Some people interpreted the robot as a tourist attraction.

Effort expectancy: In the first field trial as well as in the second one, participants rated the indicator effort expectancy very positive (first study - $mean : 3.97, SD : 0.70$; second study - $mean : 3.89, SD : 0.11$). In the second field trial, this indicator was furthermore significantly better rated by participants who actually interacted with ACE ($t(28.68) = 3.15, p < 0.05$).

Self Efficacy: In the first field trial, participants rated the indicator self efficacy the highest ($mean : 4.42, SD : 0.68$). In the second field trial, participants rated the indicator of forms of grouping in the street survey least positive ($mean : 2.85, SD : 0.14$). This indicator was significantly better rated by participants who actually interacted with ACE ($t(38.80) = 3.08, p < 0.05$). The observational data of the second study showed a high belief of the pedestrians in their capabilities to use the robot which corresponds to their rather positive attitude towards novel technology. People liked to prove to themselves in public, that they can manage the interaction with the robot. Children were especially self confident and actively approached the robot similar to the observations in the first field trial. This observation could also be made in the first field trial where the behavioral pattern of "controlling ACE with gestures or commands (waving, "follow me")" could be observed 19 times and the pattern "sending children first for pointing on the touch screen" four times.

Forms of Grouping: In the first field trial, participants rated the indicator forms of grouping rather positive ($mean : 3.54, SD : 1.54$). The observational data revealed the interesting finding that pedestrians built "interaction groups", meaning that a group of 10-15 strangers stood in front of ACE and each member of this "coincidence" group stepped forward to interact with ACE, while the rest of the group waited and watched the interaction. In the second field trial, participants rated the indicator of forms of grouping in the street survey rather positive ($mean : 3.45, SD : 0.12$). This indicator was rated significantly better by participants who actually interacted with ACE ($t(35.96) = 4.09, p < 0.05$).

Attachment: In the first field trial, participants rated the indicator attachment rather positive ($mean : 3.63, SD : 0.89$). However, in the second field trial, participants rated the indicator of attachment even lower ($mean : 2.85, SD : 0.14$). Interestingly, participants younger than 50 years rated this indicator significantly lower than older participants ($t(43.80) = -2.36, p < 0.05$). This effect could not be found in the questionnaire data of the first field trial. In the second field trial, some people showed companion like behavior towards the robot (e.g. "Let's have a look ... Oh yes ... come on ... take off"). The robot was directly addressed like a social actor. However, not everyone addressed the robot in 2nd person, even if they were standing right in front of it; a behavior which would be considered extremely impolite when interacting with a human and an indicator that the robot is not perceived as a partner by everyone.

Reciprocity: In the first field trial participants rated the indicator reciprocity the lowest ($mean : 2.25, SD : 0.81$). A reason for this could be that only 54% of the participants experienced the robot as interactive due to its limited interaction possibilites in the first study set-up- In the second field trial participants rated the indicator of reciprocity in the street survey rather positively ($mean : 3.50, SD : 0.16$). This indicator was significantly better rated by participants who actually interacted with ACE ($t(24.85) = 2.44, p < 0.05$). Furthermore, reciprocity was the only factor that was significantly better rated by men than women ($t(19.99) = -2.42, p < 0.05$). Moreover, participants younger than 50 years rated this indicator significantly lower than than older participants ($t(37) = -3.27, p < 0.05$).

Societal impact: During the street survey of the first field trial, the three questions regarding cultural context were asked. These questions were often discussed more intensively with the researchers than others. The question whether robots are pleasant working colleagues was answered quite heterogeneously. About 40% of the participants chose the middle answer category, arguing that they cannot answer the question, because they have not worked with robots. A total of 8 participants answered "strongly agree", arguing that robots could solve boring tasks or perform dangerous work for them. Similarly, 8 participants answered "strongly disagree", arguing that they are afraid not to be able to handle the robot.

The follow-up question, if the use of robots causes anxiety for the working place, was answered with "totally disagree" by 30 of the participants. Additionally, many participants commented this question explaining that their working tasks were too complex for a robot to solve. Only three participants answered that they were very afraid to be substituted by a robot. A total of 31 of the participants answered that they would strongly appreciate to own a robot. It is worth to mention that cultural context was the only indicator that revealed significant differences regarding the independent variable age in the first field trial. Older participants rated the questions regarding cultural context significantley higher, than younger ones($Kendall'sTau - B : 0.422, p = 0.036$).

The analysis of the video and audio material recorded by the additional camera mounted at the head of the ACE robot in the second field trial, additionally gave insights that pedestrians experienced the ACE robot as social actor in the public place. However, pedestrians expect that their safety is guaranteed in a public place and that the robot will not harm them. Many pedestrians showed the willingness to interact with ACE, however, the first impression prevailed. If the robot did not react accordingly to the expectation of the pedestrian in the first minute of interaction, participants aborted.

Insights on the Methodological Framework Conducting a field user study with a socially situated robot to evaluate social acceptance and societal impact of human-robot interaction in public places is helpful to gain additional insights for social acceptance beyond laboratory-based user studies. From the methodological view point, these two field trials gave the first indication that the USUS evaluation framework can also be applied in field trials, at least for the indicators of social acceptance and societal impact. However, advantages as well as disadvantages in applying this kind of methodology in the field could be identified.

When conducting a pre-structured interview in a public place it has to be considered that street-surveys are always based on non-probabilistic participant sampling, meaning that the researcher himself chooses the sample of investigation. This way of "rational choice" sampling entails several advantages as well as disadvantages. Non-probabilistic sampling is valuable if the target population is hard to define; it is adaptable during the research process, by explicitly selecting "required" participants. It allows to quickly gain first insight to the results.

Otherwise, non-probabilistic sampling does not guarantee to address all relevant participants to answer the research question, there is always the possibility to generate a distorted sample, and generaliazability of the data is problematic as the target population is not predefined. It could be shown that a field user study with an accompanying survey, can provide valuable insights on social acceptance and societal impact in terms of behavior patterns at the same time. Linking the methodological results can be helpful. The interaction analysis conducted in the second field trial in accordance to the participants autonom (questionnaire data) and the heteronom (observational data) was also highly feasible. However, this resulted in a more complicated methodological set-up with an increased number of required observers and cameras.

A main limitation of this methodological approach was that not all indicators might be well addressed in a public setting as socially desired response behavior is more dominant in a public setting and reflecting the answers in a "passing-by" situation is not as easy for the participants compared to a laboratory-based setting (where the participant is on his own with the experimenter). Furthermore, generalizability is a major issue in a field study like this, as using a "special" robot like the ACE robot does not allow generalizing results in terms of robotic appearance. Several re-tests are necessary to achieve a more general understanding of social acceptance of HRI in public settings.

4.10 Case Study 10: Cognitive Walkthrough

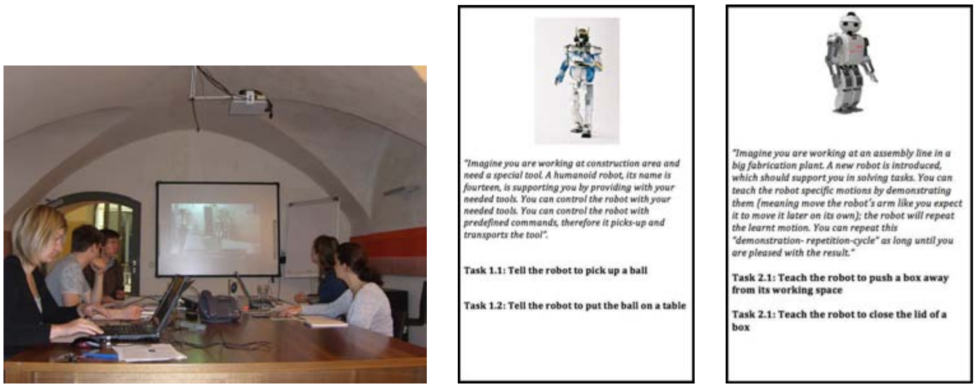
This case study was carried out based on the video material recorded during the two of the three laboratory-based user studies: (1) Case Study 3 with the HRP-2 robot (see Section 4.4), (2) Case Study 4 with the HOAP-3 robot (see Section 4.5). Four experts evaluated the two human-robot interaction scenarios in terms of usability with a special focus on learnability. The case study was conducted at the ICT&S Center, University of Salzburg, May 2009. More details on this case study can be found in Weiss et al (2009g).

The content based research question was "What are the major usability problems of HRP-2 and HOAP-3 occurring in the conducted collaboration tasks, particularly addressing learnability?". From a methodological point of view, the case study was conducted to fill the methodological gap of thoroughly addressing the usability indicator learnability.

4.10.1 Study Setting

The study was set-up as a group-based cognitive walkthrough using video material from two user studies ((see Section 4.5 and Section 4.6)). A moderator lead the walkthrough and four expert evaluators assessed the human-robot interaction scenarios in terms of learnability in accordance to the lead questions of the cognitive walkthrough technique (see SubSection 3.7.1). The video material had a duration of approximately three minutes each and was projected on a screen visible for all evaluators (see Figure 4.12). Two cognitive walkthrough sessions were conducted on two separate days: one for the HRP-2 scenario and one for the HOAP-3 scenario.

Figure 4.12: Case Study 10: Study Setting and Material



(a) Study Setting

(b) Scenario Card HRP-2

(c) Scenario Card HOAP-3

4.10.2 Instruments

For each interaction scenario, a separate cognitive walkthrough was conducted, both using the same instruments. Table 4.10 gives an overview of the addressed factor and indicator in this case study.

Video Stimulus Based on the recordings of the two user studies (see Section 4.5 and Section 4.6) a video was composed for each tasks where the ideal procedure of the interaction sequences could be watched. Thus, in total four videos were used in he cognitive walkthrough. (1) Commanding HRP-2 to pick up a ball, (2) commanding HRP-2 to put a ball on a table, (3) teaching HOAP-3 to push a box, (4) teaching HOAP-3 to close the lid of a box.

Scenario Cards To provide the expert evaluators with the necessary context of the interaction scenarios and the tasks, information for both was provided on so called scenario cards (see Figure 4.12).

Action Sequence Table A so called action sequence table was provided for the expert evaluators to support structured comparable note taking during the cognitive walkthrough and to ease the discussion between the evaluators. In the action sequence table, the task which should be evaluated was split up into

Table 4.10: Method Mix: Case Study 10

Factor	Addressed Indicator	Used Method
Usability		Cognitive Walkthrough
	Learnability	Cognitive Walkthrough

several action sequences. For each action sequence, a picture was provided and an area for each of the four cognitive walkhrough lead questions: (1) Will the user try to achieve the effect that the subtask has? (2) Will the user notice that the correct action is available? (3) Will the user understand that the wanted subtask can be achieved by the action? (4) Does the user get feedback?

4.10.3 Procedure

In the beginning of an analysis session, the moderator introduced the scenario to the evaluators and distributed the scenario card among the evaluators. Then, the moderator showed the first action sequence in the video. After each action sequence, the evaluators started a group discussion about their notes and tried to agree on if a user chose the correct interaction possibility at each step or if usability problems would occur. If usability problems were identified the expert evaluators tried to name the reasons and suggest improvements for the analyzed scenario.

In that way, the evaluators walked through the complete task and were supported by the moderator (e.g. if the audio quality of the recordings was bad the moderator read aloud the possible human commands and the robot answers). All sessions were audio recorded and transcribed. The transcript was then analyzed in terms of the cognitive effort and usability problems within each task.

4.10.4 Results

In total, 6 different main usability problems concerning learnability were identified for HRP-2 scenario and five for the HOAP-3 scenario, whereas four of the problems could be identified for both scenarios. The usability problems are presented in accordance to the definition of learnability from Dix et al (2004) saying that learnability is based on five principles: predictability, synthesizability, familiarity, genarlizability and consistency.

Insights on the Theoretical Framework This case study again showed the importance of feedback and visibility of system status to ensure a good predictability. The expert evaluators stressed the importance of consistency on the one hand (same feedback modality for the same action) but flexibility of the reaction of the robot (e.g. adaptation and variation in verbal answers) on the other hand. In the following paragraphs more detailed insights on learnability as usability indicator are presented.

Synthesizability and Consistency: Both interaction scenarios provided only one-dimensional feedback for the user. Furthermore, in both scenarios, moments occur where the user is left alone without any feedback. Regarding consistency, it would be valuable if the user always receives the same type of feedback on the same or similar action; e.g. a green blinking light whenever the robot understood a command. Regarding synthesizability, it would be recommended to support the user in assessing the effect of past actions by suitable feedback. In many situations, this would mean multidimensional feedback like a combination of a green blinking light with the audio feedback "command understood" to represent that a command is processed.

Predictability: In the HOAP-3 scenario, it was hard for the user to assess the robot's vision range, which was essential to solve the tasks successfully.It would be advisable to support the user in determining the effect of future actions based on the interaction history. Therefore, it is necessary to provide cues about the range of vision, e.g. on a separate ambient display or by coloring the area on the table where the robot is able to detect and manipulate objects.

Familiarity: In both interaction scenarios, it was hard to decide for the users if the robots are "ready for interaction", on "stand-by mode", or "switched off". Using the metaphor of a stand-by mode of a computer or a TV-set would support the learnability. The usage of familiar knowledge of computer-based or real world domains could be easily applied, e.g. by means of a stand-by LED or if the robot looks down with its head ("sleeping mode").

Generalizability: In the HOAP-3 scenario, the output of the robot in terms of phrases, body language, and light cues was hard to interpret for the users. It would be valuable to support the user when interacting within and across applications with recognizable output. Cues from human-oriented perception can be of high value, e.g. activating a robot by saying its name and short, easy to remember commands like in the HRP-2 scenario foster learnability and could be easily transferred to other interaction scenarios or humanoid robots.

Insights on the Methodological Framework The methodological adaption of a video-based cognitive walkthrough with a group of experts proved its feasibility to retrospectively analyze user studies on direct human-robot interaction scenarios with respect to learnability. Thus, it is a suitable way to close the gap in the investigation of the usability indicators.

The adaption to use videos instead of letting the evaluator interact personally with the robot was feasible as it allowed to repeat unclear action sequences and to discuss them in more detail with a team of experts. However, the two sessions also showed that it is advisable to provide scenario cards to increase the context awareness amongst the evaluators and to have at least one evaluator in the cognitive walkthrough team who was also part of the user study testing team. This experienced evaluator can provide more detailed information, e.g. if a usability problem which was assumed by another evaluator actually occurred in the user study.

The moderator role is relevant to lead the discussion back to learnability if it becomes of topic. Moreover, a positive side effect could be observed during the two sessions, namely that insights on other usability indicators were found, e.g. flexibility.

4.11 Case Study 11: Expert Interviews on Societal Impact

This case study was carried out as two consecutive expert interviews taking into account the methodological approach of a Delphi study. The first round of expert interviews was conducted with the industrial project partner DRAGADOS (a Spanish construction company), at their company site, Madrid, September 2008. The second round of interviews was conducted with the industrial project partner Space Application Services (SAS), a Belgium company interested in using robotic technology for space exploration)at their company site in Zaventem, Belgium, March 2009. Both rounds of interviews were conducted to gain deeper insights on the societal impact indicators form an industry point of view to complement the data from the in-depth user interviews. More details about these interviews can be found in Weiss et al (2009d) and in Weiss et al (2009g). The content based research questions for this study were to explore:

1. "How do systems have to be prepared not to influence the quality of life, health and security of citizens in a negative way?"

2. "How does the collaboration with robots impact working conditions and does it cause any changes in methods for employee retention?"

3. “How do new types of education have to be designed to cover the new requirements in the working life and is it important to groom workers for the new working conditions in a psychological manner?”
4. “Does the use of robots cause a change of the culture, does it affect fears and goals of people and does the collaboration with robots have an unequal impact on different cultures?”

From the methodological viewpoint of validating the proposed evaluation framework, this case study should close the gap of addressing societal impact only from a user’s point of view. Furthermore, the interviews should explore if any indicators will come up, which have not been taken into account for the theoretical framework.

4.11.1 Study Setting

The first round of interviews were conducted with two representatives of the company DRAGADOS: Miguel Segarra, head of the research department for buildings and processes, and Enrique Maizquez Gonzalez, expert in building prefabrication. The interviews were based on guidelines derived from the literature review on the societal impact indicators and lasted approximately between one and two hours each.

The second round of interviews was conducted with four representatives of the company SAS: Jeremi Gancet, Nicolas Descouvemont, Bernard Fontaine, and Elvina Motard. The guidelines for theses interviews were developed based on the results of the first round of interviews (to take into account the Delphi methodology) and lasted again between one and two hours in average.

Both sets of interviews were audio recorded and transcribed afterwards[1].

4.11.2 Instruments

Interview Guidelines Both rounds of expert interviews were semi- structured using pre-defined guidelines (the guidelines can be found in Weiss et al (2009g)) and in the Appendix: Annex 8), whereas the second round of guidelines was based on the results of the first round of interviews to take into account the Delphi approach. In both interviews, there was no chronological order of the questions. It was the responsibility of the interviewer to adapt the order of the questions in accordance to the interview situation. Table 4.11 gives an overview of which methods were used to assess which indicator in this case study.

[1]The interview protocols were authorized by the industrial partners.

Table 4.11: Method Mix: Case Study 11

Factor	Addressed Indicator	Used Method
Societal Impact		
	Quality of Life, Health, and Security Working Conditions and Employment Education Cultural Context	Expert Interview Delphi Study Expert Interview Delphi Study Expert Interview Delphi Study Expert Interview Delphi Study

4.11.3 Procedure

All interviews started with a short description of the of the FP6 EU-funded Project "Robot@cwe: Advanced robotic systems in future collaborative working-environments". Afterwards, the participants were introduced into the goal of the interview:

> "The main objective of this interview is to gain insights in the possible societal changes (chances and risks) which arise in the co-working of humans and robots. As our point of view is mainly a theoretical one, we need input from practitioners who will be on the front line of future collaboration with robots. In this interview, we would like to address diverse aspects of this future scenario. Please do not be afraid to give your own opinion as no one can foretell the future, no wrong answers exist."

After signing the permission for data utilization and the completion of a short questionnaire on demographic data, the semi-structured interview with open ended questions was conducted. Afterwards, the audio recordings of the interviews were transcribed and a qualitative content analysis was conducted (Flick et al, 2004). In the content analysis approach after Flick et al (2004), significant as well as on first sight not significant, parts of the transcript are decomposed and analyzed by a group of researchers (in theses cases three researchers for both interview rounds) in terms of manifest and latent meaning. Furthermore, the results from the expert interviews were compared with those from the user studies to reveal differences and similarities.

4.11.4 Results

Insights on the Theoretical Framework The case study showed that even though experts not only answered differently from the participants of the user studies but also heterogeneously depending on their industrial background, the Delphi approach allowed to compose one picture of the future robotic society. In the following, the cumulated results of both expert interview rounds will be presented which describe the picture of the future robotic society in accordance to the theoretical USUS indicators for societal impact.

Quality of Life, Health, and Security: Experts from DRAGADOS assumed that the life expectancy will increase due to robotic technology (e.g. tele-robotic surgery). Robots will not harm the physical health of citizens, but probably the physiological, as they reduce human-human relationships. Robots will be used as servants and support humans in many fields like care settings, education, and service jobs.

In general experts from SAS had a positive look into the future. They expect more leisure time for people and a better health care system. However, experts warned that robots are safety critical in direct human-robot collaboration (e.g. hurting humans).

Working Conditions and Employment: According to the experts from DRAGADOS, the societal changes of the working conditions due to robotic technology are on hand. There will be an increase in the unemployment rates above all for low educated people. The first applications of robots will be dirty, dull, and dangerous tasks. Experts expect no changes in income. However, they point out that more focus will have to be put on ensuring safety as a key aspect of future working conditions.

According to experts of SAS, robotic technology will have no negative impact on the employee situation in future. Robotic technology will produce many new sophisticated jobs. Moreover, there will still be a strong need for human workers in the future. SAS experts do not see robots just in dangerous environments, but in homes, assistant work, and even the arts.

Education: Regarding the question how the education system will change in the future, experts from DRAGADOS stated that education will simply become more important and that higher educated people will be preferred on the labor market. However, they do not expect massive changes in the education system and also do not think that this will be necessary. Instead, they suggested employee-training as a reasonable approach.

The experts from SAS showed a similar response behavior. They stated that the introduction of robots into people's working life should be step by step and that employees should receive specific trainings to handle robotic technology. However, they also do not see a need to teach "robotics" in school, but they stated that new disciplines at universities will evolve due to the technological progress.

Cultural Context: Regarding the influences of cultural context on robotics and vice versa, the experts from DRAGADOS pointed out that the value system will only change slowly (e.g. the value of human work will be defined differently over time). However, they do not assume any cultural limitations regarding the introduction of robots, meaning they expect the Western society equally to accept robots in working life as Asia, but it will probably take longer.

The experts from SAS see language as the cultural limit to introduce robots (English as basis for working with robots in terms of configuration and programming). They also stated that robots built for Western societies will have a different embodiment than robots for Asian countries. SAS experts also agreed that industrialized countries (above all Japan) are more ready for the introduction of robots. Moreover, robots as artists will have an impact on culture. Finally, SAS experts stressed that in future our thoughts will change what makes humans special.

Insights on the Methodological Framework Expert interviews proved to be of high value for gathering data on the societal impact indicators. Above all, having experts from two very diverse fields (construction company and space applications) provided a widespread area of answers which also differed profoundly from the assumptions participants stated at the end of the user studies (for more details on this result see Weiss et al (2009d)).

By means of theses two interview rounds it could not be validated if a Delphi study would provide a more holistic picture. To provide the SAS experts the opinion of the DRAGADOS experts (in style of a Delphi approach) did cause some methodological problems. SAS experts stated that it is hard for them to interpret answers of DRAGADOS experts as they are from a different industry.

Furthermore, difficulties occurred for the experts to have a common understanding of the term future. It worked out to give experts a clear time frame ("let's say future is in the next 50 years"), which was not helpful for user study participants.

It will be necessary to test the Delphi approach in its original fashion in the future to determine its feasibility for societal impact measurement. However, the data analysis showed that expert interviews provide in-depth heterogeneous data about the societal impact indicators which is not congruent to user study participant answers. Though this data could be used to develop a standardized questionnaires for future user studies.

4.12 Reflection on the Eleven Case Studies

This chapter presented the case studies which were conducted in terms of proving the feasibility of the proposed methodological framework of the USUS evaluation framework. Eleven studies were presented, which all provided building blocks to assess the suitability of methods to address specific USUS indicators. The sections on "Insights on the Methodological Framework" thus consisted of the advantages, limitations, and lessons learned for future study settings. I would like to reflect on the question, if I could prove the validity of the methodological framework. I would answer partly. I could show that the methods I originally proposed can be used to address the suggested indicators, but I cannot prove if it is the best method in terms of produced results. Future comparative studies would be required to prove this.

Moreover, I could show that specific study settings and tools (e.g. combination of specific questionnaires) can be used for a variety of human-robot interaction scenarios, but this does not guarantee that it can work for all interaction scenarios. For instance, studies with more complex tasks and with longer task durations will have to skip questionnaires and interview parts to stay in a reasonable time frame. However, the eleven case studies cannot prove the completeness of reasonable methods to address the theoretical indicators. The indicators of user experience and social acceptance were mainly assessed by means of questionnaires. Other methods should be considered in the future to gain additional data. In Chapter 6, I will give additional method proposals, based on methodological considerations and challenges and present a revised methodological framework.

To summarize the content-related findings, one could highlight the following aspects: Regarding usability, the task durations vary surprisingly among the study participants. Most interesting was the observation that there is a difference in task duration between Asian and Western participants due to the different problem solving strategies. As a novel methodological approach, I consider "cultural interaction scenarios" which take this finding into account to shape human-robot interaction based on the socio-cultural context.

In general, all evaluated human-robot interaction scenarios offered too little flexibility in terms of solving strategies. Human-robot interaction scenarios are still strictly ordered action sequences. More flexible scenarios will improve the user performance. Furthermore, there is also a lack in user support. If the goal is that novice users can independently interact with humanoid robots, user support systems must be developed that allow the user to recover him/herself from minor errors.

Regarding social acceptability, there are good news for roboticists and engineers, as the attitude towards robots ratings by means of the NARS questionnaire were significantly better (at least for one scale) after the interaction with the robotic system in all laboratory-based studies. This finding indicates that user-centered design approaches, like the USUS evaluation framework is a step into the right direction to allow a mutual shaping between societal requirements and technological possibilities. Moreover, the case study data, as well as the online survey data (see Section 5.3.2), indicated that the ratings of the social acceptance indicators are heavily dependent on independent variables like age, gender, and nationality. This finding again supports my suggestion of "cultural interaction scenarios".

To my conviction, the HRI community now needs to go one step further and not only analysis the differences regarding perception and acceptance of robotic agents, but use these findings to inform the design of interaction scenarios, e.g. especially for women or for children. The adaption robotic systems should not only happen on an individual basis during the actual interaction, but predefined on a general level. This could be of advantage in public setting, where robots not only interact with one specific person, but with groups of people who share similar characteristics.

Regarding user experience, it was interesting to observe that the arousal during the interaction varies a lot for novice users as well as for expert users. As already mentioned above, a next step could be to use this physiological data as input therefore a robotic system can react on this variations. Furthermore, a high correlation between the feeling of security and the feeling of control could be observed in the user studies. In other words, the more the user had the feeling of control over the interaction, the more he/she felt secure in the interaction. As it became obvious in the laboratory- and field-based studies, users in general share the belief that robots which are used for direct interaction with humans must be safe (users take their safety for granted). It will be important to find more cues to increase the feeling of control to guarantee the feeling of security also for the long-term (not experimental) usage of robots.

Regarding societal impact, a major finding was that the working context users prefer robots as sophisticated tools, not as working colleagues. Users expect robots in the working context to be functionally designed, for the specific tasks they should perform. In consequence for humanoid robots, this indicates that cues need to be identified that visualize the task assignment of the robot. For HRI researchers, it indicates that in the working context the focus should be put on acceptance factors like performance expectancy instead of attachment or reciprocity.

The final issue, however still remains, after eleven case studies it cannot be guaranteed that the USUS evaluation framework is complete in terms of indicators, factors, or methods. The case studies showed that human-robot interaction studies can be very challenging regarding theoretical grounding and methodological approach. Human-robot interaction studies are also very rewarding, if they are planned in detail and in accordance to a theoretical basis and lessons learned from previous studies. However, due to the small sample sizes in the user studies, statistical reliability and validity could only partly be addressed. To limit this aspect, the next chapter presents the results of a broad online survey with 398 participants.

Chapter 5

Reference Model: The Revised USUS Framework

5.1 Outline of the Chapter

This chapter presents a reference model on the synergies and influences between the proposed factors and indicators of the theoretical framework (see Chapter 3). The reference model is based on a cross-analysis of the quantitative data gathered in a broad online survey based on the UX- and SoAc-Questionnaire used in some of the case studies (see Chapter 4). The reliability and validity of the user experience and social acceptance indicators is tested by factor analysis and Cronbach's alpha. The results of the descriptive analysis of the derived indicator scales show that there is a strong interdependency within some user experience and social acceptance factors. Furthermore, the influence of six independent variables: "nationality" (expressions: Asian/Western), "gender", "age", "profession" (expressions: worked in HRI/not worked in HRI), "experiment" (expressions: already participated in HRI experiments/never participated in HRI experiments), and "robot" (expressions: HRP-2 robot/HOAP-3 robot) was analyzed. The outcome of this work is a revised framework for the factors user experience and social acceptance.

5.2 The UX- and SoAc-Questionnaire

Based on the literature review and on the results of the focus group study, two questionnaires were designed to measure the indicators of the user experience and the social acceptance factors: The User Experience-Questionnaire (subsequently UX-Questionnaire) and the Social Acceptance-Questionnaire (subsequently SoAc-Questionnaire). The development of the questionnaires began with a huge number of items, which were intended to cover the different indicators of the theoretical framework to be measured. From this pool of items two researchers decided collaboratively which items should be used in the final questionnaires. Furthermore, it was decided to measure the UX-Questionnaire on a 7 point Likert scale in accordance to the AttrakDiff questionnaire (see SubSection 3.7.3) and to measure the SoAc-Questionnaire in accordance to the UTAUT model on a 5 point Likert scale (see Section 3.4).

In the following, the two questionnaires are presented in their English version (for the second HOAP-3 user study the questionnaires were translated to Spanish by an accredited translator - see Appendix: Annex 9; for the mixed-reality user study they were translated to German - see Appendix: Annex 10, and for the HRC scenario with HRP-2 they were translated to Japanese - see Appendix: Annex 11). All items of the questionnaires were already preassigned to an indicator of the theoretical framework. Based on this allocation, the scales were calculated for the user studies. For the SoAc-Questionnaire, additional items for the UTAUT factor facilitating conditions were added, which were taken into account later on in the analysis of the online survey. The empirical validation of the scales could not be conducted until the online survey, due to the small sample sizes.

UX-Questionnaire

		Absolutely disagree						Absolutely agree
Indicator	Item	1	2	3	4	5	6	7
EMB	I liked the size of the robot.	☐	☐	☐	☐	☐	☐	☐
E	Interacting with the robot is fun.	☐	☐	☐	☐	☐	☐	☐
COE	When talking to the robot, I feel like talking to a human.	☐	☐	☐	☐	☐	☐	☐
E	I am happy when the robot understands my commands.	☐	☐	☐	☐	☐	☐	☐
FS	I think that the robot is vulnerable to hackers. (R)	☐	☐	☐	☐	☐	☐	☐
COE	I can interact with the robot like I interact with other humans.	☐	☐	☐	☐	☐	☐	☐
E	I am disappointed if the robot does not understand my commands. (R)	☐	☐	☐	☐	☐	☐	☐
EMB	I liked that the robot looked similar to a human.	☐	☐	☐	☐	☐	☐	☐
FS	I hesitate to use the robot for fear of making errors that will harm me. (R)	☐	☐	☐	☐	☐	☐	☐
COE	When working with the robot, I perceive it as working in a team.	☐	☐	☐	☐	☐	☐	☐
HOP	I perceive the robot as a social actor.	☐	☐	☐	☐	☐	☐	☐
EMB	I liked that the robot has human like features: face, ears, eyes, etc.	☐	☐	☐	☐	☐	☐	☐
FS	I fear to use the robot, as an error might harm the robot. (R)	☐	☐	☐	☐	☐	☐	☐
HOP	I liked that the robot detected my face.	☐	☐	☐	☐	☐	☐	☐
COE	I feel good when interacting with the robot.	☐	☐	☐	☐	☐	☐	☐
EMB	I liked the physical co-location of the robot.	☐	☐	☐	☐	☐	☐	☐
HOP	I perceive that the robot is intelligent.	☐	☐	☐	☐	☐	☐	☐
E	I am angry if the robot does not understand my commands. (R)	☐	☐	☐	☐	☐	☐	☐
EMB	I liked the design of the robot.	☐	☐	☐	☐	☐	☐	☐
HOP	I enjoyed talking with the robot	☐	☐	☐	☐	☐	☐	☐
COE	The robot could become a companion for me.	☐	☐	☐	☐	☐	☐	☐
FS	I feel secure when working with the robot.	☐	☐	☐	☐	☐	☐	☐
E	I felt afraid of the robot. (R)	☐	☐	☐	☐	☐	☐	☐
HOP	I liked that the robot understands my voice commands.	☐	☐	☐	☐	☐	☐	☐
FS	I perceive the robot as safe.	☐	☐	☐	☐	☐	☐	☐

25 items
R: reverse item (6)
EMB: Embodiment (5)
E: Emotion (5)
COE: Co-Experience (5)
FS: Feeling of Security (5)
HOP: Human-oriented Perception (5)

ICT&S Center

UNIVERSITÄT SALZBURG

SoAc-Questionnaire

		Strongly agree	Agree	undecided	Disagree	Strongly disagree	Do not know
Indicator	Item	5	4	3	2	1	99
PE	The deployment of robots will increase my chances of success in my job.	☐	☐	☐	☐	☐	☐
EE	It will be easy for me to become skillful in dealing with robots.	☐	☐	☐	☐	☐	☐
AT	Robots will make work more interesting.	☐	☐	☐	☐	☐	☐
FG	Robots will be part of our everyday work.	☐	☐	☐	☐	☐	☐
FR	I can imagine that I will care for the wellbeing of a robot.	☐	☐	☐	☐	☐	☐
A	I can imagine taking a robot into my heart.	☐	☐	☐	☐	☐	☐
FG	Robots will have a similar importance as human colleagues.	☐	☐	☐	☐	☐	☐
A	It would feel good if a robot was near me.	☐	☐	☐	☐	☐	☐
PE	I will solve tasks faster using robots.	☐	☐	☐	☐	☐	☐
SE	I could not be successful while working with robots under time pressure. (R)	☐	☐	☐	☐	☐	☐
PE	Robots will make my tasks easier.	☐	☐	☐	☐	☐	☐
FR	Humans and robots will be interdependent.	☐	☐	☐	☐	☐	☐
EE	Interacting with robots will be easy to understand.	☐	☐	☐	☐	☐	☐
AT	I would not like to work together with robots. (R)	☐	☐	☐	☐	☐	☐
EE	It would be difficult to learn how to handle robots.(R)	☐	☐	☐	☐	☐	☐
SE	I will never be able to solve a task together with a robot. (R)	☐	☐	☐	☐	☐	☐
PE	Robots will facilitate burdensome tasks we have now.	☐	☐	☐	☐	☐	☐
FG	Robots and humans will make a good team.	☐	☐	☐	☐	☐	☐
FR	The interaction with robots will be a mutual experience.	☐	☐	☐	☐	☐	☐
AT	Using robots is a bad idea. (R)	☐	☐	☐	☐	☐	☐
AT	I would be afraid to employ robots at work. (R)	☐	☐	☐	☐	☐	☐
FC	I would be motivated to integrate robots in my daily workday.	☐	☐	☐	☐	☐	☐

ICT&S Center

UNIVERSITÄT SALZBURG

		Strongly agree	Agree	undecided	Disagree	Strongly disagree	Do not know
Indicator	Item	5	4	3	2	1	99
FC	If problems with robots occur there would be people who could help me.	☐	☐	☐	☐	☐	☐
SE	I could not solve a task with the help of robots if no one was there to tell me what to do. (R)	☐	☐	☐	☐	☐	☐
PE	I would find robots in my job useful.	☐	☐	☐	☐	☐	☐
EE	It will be easy to use robots.	☐	☐	☐	☐	☐	☐
AT	Working with robots would be fun.	☐	☐	☐	☐	☐	☐
EE	Efforts to solve tasks together with robots will be a huge undertaking. (R)	☐	☐	☐	☐	☐	☐
SE	I could solve all problems which occurred during the interaction on my own.	☐	☐	☐	☐	☐	☐
FC	I have the necessary knowledge to handle robots.	☐	☐	☐	☐	☐	☐
A	I can imagine building a special relationship with robots.	☐	☐	☐	☐	☐	☐
AT	I do not think it is necessary to employ robots in daily working life.(R)	☐	☐	☐	☐	☐	☐
FR	The relationship with robots will be based on the principle of give and take.	☐	☐	☐	☐	☐	☐
PE	The utilization of robots will increase my productivity.	☐	☐	☐	☐	☐	☐
A	I would not like to imagine a world in which robots were not used.	☐	☐	☐	☐	☐	☐
FG	Robots will be an important part of our society.	☐	☐	☐	☐	☐	☐
SE	When a problem occurs with robots, I would not be able to continue with my work without help. (R)	☐	☐	☐	☐	☐	☐
FC	I do not have the necessary abilities to handle robots. (R)	☐	☐	☐	☐	☐	☐
FG	I would like to collaborate with robots.	☐	☐	☐	☐	☐	☐
FC	There will be enough information material available to help simplify the interaction with robots.	☐	☐	☐	☐	☐	☐

40 items
R: reverse item (11)
PE: Performance expectancy (6)
EE: Effort expectancy (5)
SE: Self efficacy (5)

A: Attachment (4)
AT: Attitude towards using technology (6)
FC: Facilitating conditions (5)
FG: Forms of grouping (5)
FR: Feeling of reciprocity (4)

5.3 The Online Survey

The online survey was based on the preliminary UX- and SoAc-Questionnaires and on two short videos (around 1.5 minutes) derived from the preliminary HRP-2 user study (see Section 4.4) and the first HOAP-3 user study (see Section 4.5). Only items which could be answered just by watching human-robot interaction in a video were chosen from the preliminary questionnaires. Because of that, the UX-Questionnaire was heavily reduced in its number of items (no adaptations have been executed for the SoAc-Questionnaire). Furthermore, participants had to answer some questions on socio-demographics. The online survey was provided in four different languages (German, English, French and Spanish). The French and the Spanish questionnaire were translated by accredited translators. Unfortunately, at this stage of the "Robot@cwe" project it was not possible to provide the questionnaire also in Japanese. The complete English survey can be found in the following and the translations in the Appendix: Annex 9, Annex 10, and Annex 12[1].

[1]For the French translation only the short version of the UX-Questionnaire is available.

ICT&S Center

UNIVERSITÄT SALZBURG

Online Survey

Questions on User Experience

	Absolutely disagree						Absolutely agree
	1	2	3	4	5	6	7
I liked the size of the robot.	☐	☐	☐	☐	☐	☐	☐
I think that the robot is vulnerable to hackers. (R)	☐	☐	☐	☐	☐	☐	☐
I would be disappointed if the robot did not understand my commands. (R)	☐	☐	☐	☐	☐	☐	☐
I liked that the robot looked similar to a human.	☐	☐	☐	☐	☐	☐	☐
I would hesitate to use the robot for fear of making errors that will harm me. (R)	☐	☐	☐	☐	☐	☐	☐
I perceived the robot as a social actor	☐	☐	☐	☐	☐	☐	☐
I liked that the robot has human like features: face, ears, eyes, etc.	☐	☐	☐	☐	☐	☐	☐
I would fear to use the robot, as an error might harm it. (R)	☐	☐	☐	☐	☐	☐	☐
I perceived that the robot is intelligent.	☐	☐	☐	☐	☐	☐	☐
I would be angry, if the robot did not understand my commands. (R)	☐	☐	☐	☐	☐	☐	☐
I liked the design of the robot.	☐	☐	☐	☐	☐	☐	☐
The robot could become a companion for me.	☐	☐	☐	☐	☐	☐	☐
I would feel secure when working with the robot.	☐	☐	☐	☐	☐	☐	☐
I felt afraid of the robot. (R)	☐	☐	☐	☐	☐	☐	☐
I perceived the robot as safe.	☐	☐	☐	☐	☐	☐	☐

Questions on Social Acceptance

	Strongly agree	Agree	undecided	Disagree	Strongly disagree	Do not know
Item	5	4	3	2	1	99
The deployment of robots will increase my chances of success in my job.	☐	☐	☐	☐	☐	☐
It will be easy for me to become skillful in dealing with robots.	☐	☐	☐	☐	☐	☐
Robots will make work more interesting.	☐	☐	☐	☐	☐	☐
Robots will be part of our everyday work.	☐	☐	☐	☐	☐	☐
I can imagine that I will care for the wellbeing of a robot.	☐	☐	☐	☐	☐	☐
I can imagine taking a robot into my heart.	☐	☐	☐	☐	☐	☐
Robots will have a similar importance as human colleagues	☐	☐	☐	☐	☐	☐
It would feel good if a robot was near me.	☐	☐	☐	☐	☐	☐
I will solve tasks faster using robots.	☐	☐	☐	☐	☐	☐
I could not be successful while working with robots under time pressure. (R)	☐	☐	☐	☐	☐	☐
Robots will make my tasks easier.	☐	☐	☐	☐	☐	☐
Humans and robots will be interdependent.	☐	☐	☐	☐	☐	☐
Interacting with robots will be easy to understand.	☐	☐	☐	☐	☐	☐
I would not like to work together with robots. (R)	☐	☐	☐	☐	☐	☐

It would be difficult to learn how to handle robots.(R)	☐	☐	☐	☐	☐	☐
I will never be able to solve a task together with a robot. (R)	☐	☐	☐	☐	☐	☐
Robots will facilitate burdensome tasks we have now.	☐	☐	☐	☐	☐	☐
Robots and humans will make a good team.	☐	☐	☐	☐	☐	☐
The interaction with robots will be a mutual experience.	☐	☐	☐	☐	☐	☐
Using robots is a bad idea. (R)	☐	☐	☐	☐	☐	☐
I would be afraid to employ robots at work. (R)	☐	☐	☐	☐	☐	☐
I would be motivated to integrate robots in my daily workday.	☐	☐	☐	☐	☐	☐
If problems with robots occur there would be people who could help me.	☐	☐	☐	☐	☐	☐
I could not solve a task with the help of robots if no one was there to tell me what to do. (R)	☐	☐	☐	☐	☐	☐
I would find robots in my job useful.	☐	☐	☐	☐	☐	☐
It will be easy to use robots.	☐	☐	☐	☐	☐	☐
Working with robots would be fun.	☐	☐	☐	☐	☐	☐
Efforts to solve tasks together with robots will be a huge undertaking. (R)	☐	☐	☐	☐	☐	☐
I could solve all problems which occurred during the interaction on my own.	☐	☐	☐	☐	☐	☐
I have the necessary knowledge to handle robots.	☐	☐	☐	☐	☐	☐
I can imagine building a special relationship with robots.	☐	☐	☐	☐	☐	☐
I do not think it is necessary to employ robots in daily working life.(R)	☐	☐	☐	☐	☐	☐
The relationship with robots will be based on the principle of give and take.	☐	☐	☐	☐	☐	☐
The utilization of robots will increase my productivity.	☐	☐	☐	☐	☐	☐
I would not like to imagine a world in which robots were not used.	☐	☐	☐	☐	☐	☐
Robots will be an important part of our society.	☐	☐	☐	☐	☐	☐
When a problem occurs with robots, I would not be able to continue with my work without help. (R)	☐	☐	☐	☐	☐	☐
I do not have the necessary abilities to handle robots. (R)	☐	☐	☐	☐	☐	☐
I would like to collaborate with robots.	☐	☐	☐	☐	☐	☐
There will be enough information material available to help simplify the interaction with robots.	☐	☐	☐	☐	☐	☐

Questions on Socio Demographics

Gender: male/ female
Age: _________
Education: obligatory school/ high school/ university
Nationality: ___________
Did you ever work in the research field of Human-Robot Interaction before? Yes/ No
Did you ever participate in a research study with a robot before? Yes/ No

The questionnaires were distributed via email on three scientific online platforms: CHI-announcement, robotics worldwide, and EUropean RObotics research Network (EURON). Moreover, it was distributed to all University of Salzburg employees via an online newsletter and to the "Robot@cwe" project partners with the request of further distribution. The survey had a two months run time (from March 2009 to May 2009). The survey was set-up with the online tool lime survey. In the first part participants had to watch the short movie either about the HRP-2 or the HOAP-3 robot (the allocation was counterbalanced by chance) and then answer the questions on user experience and social acceptance and finally on socio-demographics.

A total of 398 surveys were filled in. In this remaining sample participants were kept in the analysis if their questionnaires were completed and no suspect data pattern was present, e.g. questionnaires with more than 10% questions missing, patterns like 1234512345, or giving all one answers were discarded.

Participants from 37 different nations took part in the online survey. Of the total participants who took the survey, 246 filled in the German version, 123 the English, 9 the Spanish, and 20 the French questionnaire. As the average nationality distribution across all the distribution channels is unknown the exact numbers for the language representation cannot be fixed. It is clear that the offered languages were not sufficient, therefore all participants could not answer the questions in there mother tongue. The huge bias towards more German-speaking participants can be explained by the distribution through the University of Salzburg news letter and better mouth to mouth advertisement. The bias towards English-speaking participants can be explained by two reasons: (1) the questionnaire was not available in every mother language, (2) even if the questionnaire was available in the mother language participants with a scientific background preferred the English questionnaire as it is the language of the research community.

The mean age of the participants was 29.17 years ($SE : 0.438, n = 395$); the youngest participant was 19, the oldest 66 years old. A total of 47.2% of participants were female, 52.8% were male (n=394). Participants who only finished "obligatory school" composed 0.5%, 39.1% "highschool", and 60.4% of the participants had a "university degree" (n=394). Regarding the question on the pre-knowledge in HRI, 22.6% of the participants stated that they worked or still work in the reserach field of HRI, while 77.4% stated the opposite(n=390). Those participants who have already taken part an HRI experiment composed 33.6% and 66.4% did not (n=393).

5.3.1 The Context of the Survey

The distribution of an online survey on human-robot interaction based on video data of two user studies among robotic related email lists, is a good background and a sample of 398 participants can be considered large. However, as the participant data in the previous section showed, this survey is not representative in a strict statistical sense. A method to obtain a representative sample would have been "random sampling" (e.g. picking participants randomly from a list) or better from a so called strata, subsets of the population that share at least on common characteristics and whose sizes reflect the actual representation in the population ("stratisfied sampling").

As mentioned before the survey was distributed to people with a specific foreknowledge and interest in robotics which is likely to have injected a bias towards a more positive response behavior. However, with the sampling method used for this survey ("convenience sampling") and the videos from the user studies as stimulus, one gets an approximation of the truth.

Two limitations were inherent to this cross-cultural online survey due to its set-up and procedure:

1. Language is not the only boarder between different nationalities; also mentalities, cultures, and attitudes differ. The items of the questionnaires were in fact translated by accredited translators but a translation can always have slight variations in meaning. E.g. the German term "sozialer Akteur"might have a slightly different connotation than "social actor" in English. No adaptations and modifications of questionnaire features like content, format, or visual presentation of a question were conducted in order to better fit the needs of a specific population in terms of language and culture. Future work will definitely be required to improve the survey design in accordance to cross-cultural study design guidelines in order to limit confounding variables (e.g culturally different response behavior) and produce more reliable and comparable data (see Harkness et al (2003)). However, it is assumed that the words in the questions enjoy a relative consistent meaning across the languages and that this factor is not dominant. Thus, the results presented in the following sections give valuable first insights on user experience and social acceptance for a sample with divers participating nationalities.

2. The participants for the online survey were gained by online platforms which are used by people interested in technology in general and robotics in specific. In other words, the participants of the online survey were (semi) experts in the field, but were not interviewed in their role as experts, but as "normal survey participant". The strong link between the

interviewee's foreknowledge and their acceptance has already been observed in Hammel et al (1989) and Dario et al (1999). Participants who knew more about robots had a more positive answering behavior. Additional information on robotics or the possibility to see or interact with a robot, changed the mind of a majority of participants from negative or undecided to positive. This attitude change could also be observed case studies (see Chapter 4) and in the online survey data (see Section 5.3.2). Data analysis and the reference model derived from the data gathered in this online survey has to be read keeping these limitations in mind.

5.3.2 Data Analysis

The data analysis of the online survey was conducted to fulfill three aims: (1) Reviewing the theoretical factor indicator model; (2) Developing a statistically validated questionnaires measuring user experience and social acceptance of human-robot interaction; (3) Developing a reference model visualizing the interrelations and correlations between user experience and social acceptance indicators.

Factor Analysis

The dimensionality of the 15 items from the user experience measures and the 47 items of the social acceptance measure were analyzed using a maximum likelihood factor analysis. Three criteria were used to determine the number of factors to rotate: (1) the a priori hypothesis that the measures were not unidimensional, (2) the initial statistics of the principal component analysis, and (3) the interpretability of the factor solutions in accordance to the theoretical USUS framework.

The principal component analysis indicated that the initial hypothesis that the measures are not unidimensional was correct (see Figure 5.1 and Figure 5.2). Based on the principal component analysis, four factors were rotated using a Varimax rotation procedure for the user experience measures and five factors for the social acceptance measures[2]. The rotated solutions, as shown in Table 5.1 and Table 5.2 (the tables can be found at the end of this chapter) yielded four interpretable factors for user experience embodiment (EMB), feeling of security (FS), human-oriented perception (HOP), Emotion (E) and five interpretable factors for social acceptance: performance expectancy (PE), effort expectancy (EE), relationship expectancy (RE), perceived competences (PC), attitude towards robots (ATT). The user experience factors accounted

[2]originally five factors were theoretically assumed for user experience and eight for social acceptance in the theoretical USUS framework see Section 3.2.

for embodiment: 28%, feeling of security 16%, human-oriented perception 10%, emotion 8% of the item variance. The social acceptance factors accounted for performance expectancy 33%, effort expectancy 7%, relation expectancy 6%, perceived competences 5%, attitude towards robots 4% of the item variance.

Figure 5.1: Principal Component Analysis User Experience

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4,248	28,321	28,321	4,248	28,321	28,321
2	2,432	16,216	44,538	2,432	16,216	44,538
3	1,453	9,688	54,226	1,453	9,688	54,226
4	1,146	7,640	61,866	1,146	7,640	61,866
5	,883	5,885	67,751			
6	,818	5,454	73,205			
7	,673	4,488	77,693			
8	,559	3,729	81,422			
9	,530	3,532	84,954			
10	,515	3,431	88,386			
11	,459	3,063	91,449			
12	,401	2,671	94,120			
13	,363	2,423	96,543			
14	,279	1,862	98,405			
15	,239	1,595	100,000			

Extraction Method: Principal Component Analysis.

Subsequently, an internal consistency estimation of reliability was computed for all user experience and social acceptance measures: Cronbach's alpha. The values for Cronbach's alpha can be found in Table 5.3, at the end of this chapter, each indicating a satisfactory reliability. The consistency estimation revealed that the two items for the scale emotion were wrongly expected to be reverse and thus needed to be recoded again, namely these items "I would be angry if the robot did not understand my commands" and "I would be disappointed if the robot did not understand my commands". It is recommended for the future usage of the UX-Questionnaire to avoid questions with double negatives and to reformulate these two questions, e.g. "I would be happy if the robot understands my commands" and "I would be pleased if the robot follows my commands".

Figure 5.2: Principal Component Analysis Social Acceptance

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	13,190	32,975	32,975	13,190	32,975	32,975
2	2,806	7,015	39,990	2,806	7,015	39,990
3	2,431	6,078	46,068	2,431	6,078	46,068
4	1,983	4,957	51,025	1,983	4,957	51,025
5	1,617	4,042	55,066	1,617	4,042	55,066
6	1,255	3,136	58,203	1,255	3,136	58,203
7	1,096	2,741	60,943	1,096	2,741	60,943
8	,984	2,459	63,402			
9	,943	2,357	65,759			
10	,864	2,161	67,920			
11	,826	2,065	69,985			
12	,762	1,905	71,890			
13	,726	1,814	73,704			
14	,694	1,736	75,440			
15	,670	1,676	77,116			
16	,651	1,627	78,743			
17	,618	1,546	80,289			
18	,563	1,407	81,695			
19	,542	1,355	83,051			
20	,530	1,326	84,377			
21	,498	1,245	85,622			
22	,465	1,162	86,784			
23	,453	1,133	87,917			
24	,439	1,097	89,014			
25	,414	1,034	90,049			
26	,396	,989	91,037			
27	,370	,926	91,963			
28	,358	,894	92,858			
29	,343	,857	93,715			
30	,320	,801	94,516			
31	,297	,743	95,259			
32	,279	,698	95,956			
33	,270	,674	96,630			
34	,252	,631	97,261			
35	,237	,592	97,853			
36	,207	,517	98,370			
37	,193	,482	98,852			
38	,182	,455	99,307			
39	,150	,374	99,681			
40	,127	,319	100,000			

Extraction Method: Principal Component Analysis.

Differences and Correlation Analysis

General Insights on User Experience and Social Acceptance For further analysis of the user experience and the social acceptance measures, one scale variable was developed for each factor. The scales were computed by taking the mean across the according items. The descriptives for the theses scales can be found in Table 5.4, at the end of this chapter.

Regarding the user experience of robots, it is debatable to what extent robots, in general and in working scenarios in particular, should look like human beings. The uncanny valley theory of Mori (1970) already assumed that the correlation "the more human-like, the better" stops at some point. Nevertheless, embodiment was best rated among all other user experience indicators. The conclusion can be drawn that the HRP-2 robot and the HOAP-3 robot still maintain the necessary degree of visible artificiality to be assessed positive in terms of embodiment. It seems that the two videos presented in the online survey showed interaction scenarios in which the balance of function and form was adequate for the participants. However, the low rating of the human-oriented perception indicator, indicates that the participants mainly associate robots with "classical" qualities of machines, but not with attributes like intelligence and social skills. The low rating could also be due to the foreknowledge of the participants, knowing there is a long way to go until the technological development of "autonomous sociable robots". Considering that participants with a good image of robotics answered the questions on feeling of security with a mean value, is surprisingly low. However, the interaction with the robots did not seem to frighten the participants on a high degree, but they are skeptical towards the overall safety. The mean value for the indicator emotion indicates that the robots evoked expectations that, if not fulfilled, could lead to negative emotions for some of the participants.

Regarding the social acceptance of robots the results show a clear tendency towards traditional technology acceptance indicators, like effort, acceptance, and performance expectancy instead of aspects such as relationship expectancy. These results go in line with results from the case studies. For example, in the focus group case study (Section 4.2), participants discussed that robots should be treated like tools and not like colleagues and created a functional designed robot with the "Robot Building Set"), with results from the user experience questions, and state-of-the art literature (e.g. Arras and Cerqui (2005); Takayama et al (2008)).

It seems that a majority would accept robots in collaborative scenarios if they increase the performance with little extra effort. The general attitude towards robots was rated rather high, meaning that the participants are very positive attuned towards robots. It can be concluded that robots are no longer seen as "job killer technology". This finding is on one hand partly due to the sample characteristics, but on the other hand goes in line with the findings from Arras and Cerqui (2005). The answering behavior on the indicator perceived competency is rather positive as well. The response reveals that participants tend to feel skilled enough to interact with the robots presented in the videos without the assistance of another human. However, it has to be considered that several of the participants have previous experience with robots in their jobs or as subjects in user studies.

More insights will be given regarding deviations in the response behavior. A Kolmogorov-Smirnov test showed that the questionnaire data was not normally distributed, non-parametric Mann Whitney U-tests were calculated to analyze the data with respect to six independent variables: “nationality” (expressions: Asian/Western), “gender”, “age”, “profession” (expressions: worked in HRI/not worked in HRI), “experiment” (already participated in HRI experiments/never participated in HRI experiments), and “robot” (expressions:HRP-2 robot/HOAP-3 robot).

Differences between Western and Asian Participants The fact that a different cultural background influences the societal attitude towards robotic systems, it is often referred to the “two robotic cultures” in the East and the West (Hornyak, 2006). To analyze if there are differences in terms of user experience and social acceptance between Asian and Western participants, the variable nationality was transformed into Western (n=366) and Asian (n=22).

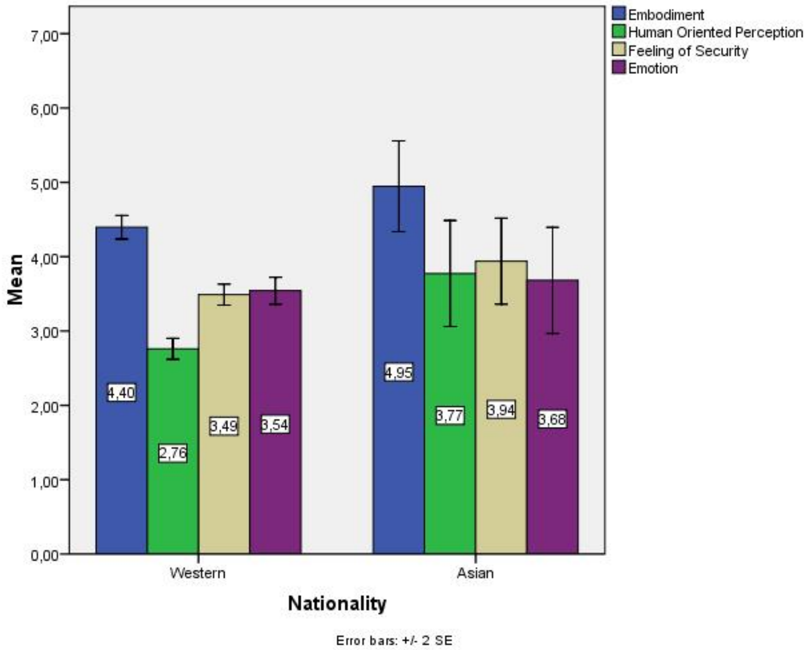
The analysis revealed for the user experience indicators human-oriented perception ($z = -2.76, p = 0.01$) a significant difference, indicating that Asian participants rated this indicator better than Western participants (see Figure 5.3). This result goes in line with the results of Bartneck (2008), who could show that conventional humanoid robots (not androids) are perceived more positively by Americans compared to Japanese.

The social acceptance rating differs significantly in terms of relationship expectancy ($z = -4.21, p = 0.00$) and performance expectancy ($z = -2.50, p = 0.01$) between Asian and Western participants. Thereby, the Asian citizens show higher social acceptance ratings than the Western participants (see Figure 5.4).

Regarding relationship expectancy, this results support the assumption of Kaplan (2004) who stresses that due to the fact that “artificial” and “natural” is not strictly separated in the Asian culture as Asians can more easily imagine to build a relationship to a robot. However, it is somehow contradictory that performance expectancy was also rated higher by Asian participants, as the focus on performance as a quality factor for a machine is assigned to Western participants.

It is concluded, that this response behavior could be explained to the fact that performance expectancy is the strongest indicator of social acceptance

Figure 5.3: Differences UX: Western and Asian



according to the factor analysis. People with an Asian cultural background show a higher acceptance in general (Hornyak, 2006).

Differences within Gender Regarding user experience (see Figure 5.5), only the factor feeling of security was rated significantly different by men and women ($z = -3.17, p = 0.02$). Female participants rated the indicator feeling of security higher than male participants. This positive attitude of women in terms of feeling of security, could be explained through the stereotype picture of genders and technology, arguing that women are less familiar with high-technology and robotics and thus less concerned about safety issues (similar to the results of the two "ACE" field trials, see Section 4.9, in which participants expected their safety for granted). This result mirrors Arras and Cerqui (2005), but it comes as a surprise in a survey context, in which mostly women with technological foreknowledge were expected to answer.

Further differences within gender also agrees with with the findings from Arras and Cerqui (2005). Regarding the social acceptance indicators, several significant differences could be identified depending on gender (see Figure 5.6). Male participants showed a significantly higher relationship expectancy ($z = -5.46, p = 0.00$) than female participants, as well as perceived competency ($z = -6.24, p = 0.00$), attitude towards robots ($z = -5.58, p = 0.00$), perfor-

Figure 5.4: Differences SoAc: Western and Asian

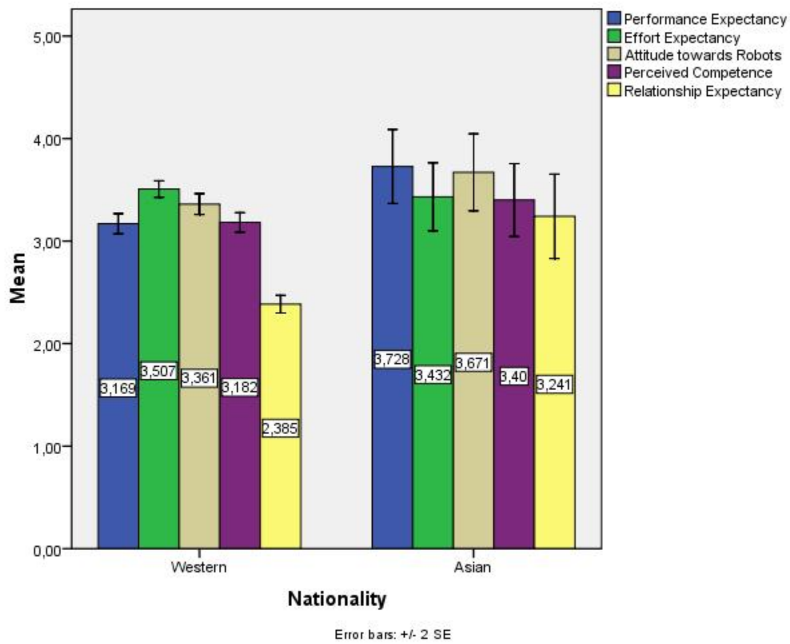
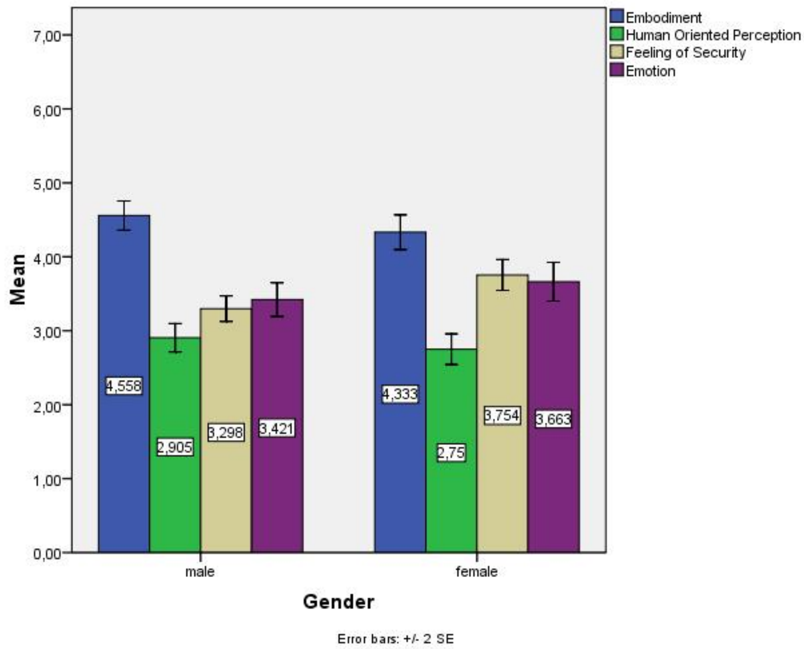
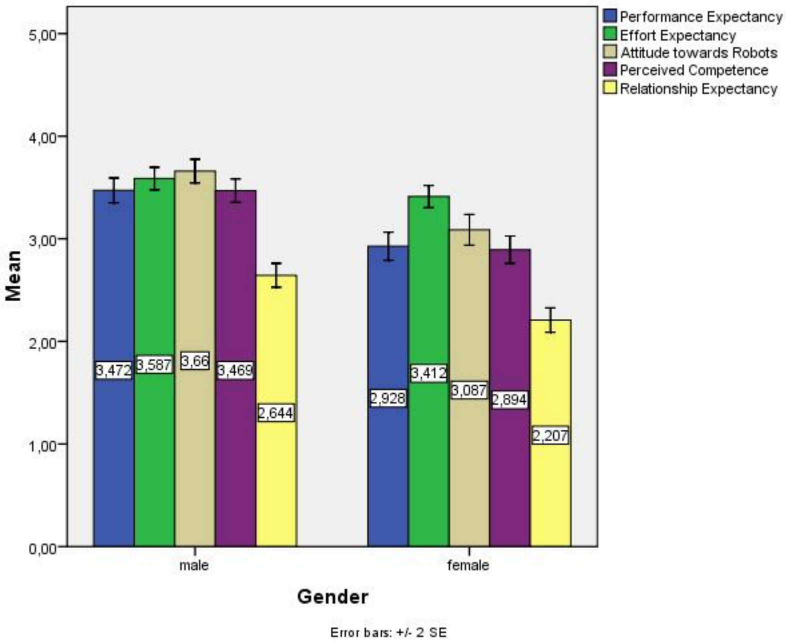


Figure 5.5: Differences UX: Male and Female



mance expectancy ($z = -5.77, p = 0.00$), and a significantly more positively rated effort expectancy ($z = -2.51, p = 0.00$). This tangible, more positive attitude of men towards robots and women as the more skeptical gender group, mirrors the results of the survey of Arras and Cerqui (2005). Thus, it supports the claim of the authors that robotics still requires promotion to be a valuable aid in everyday life, equally for men and women.

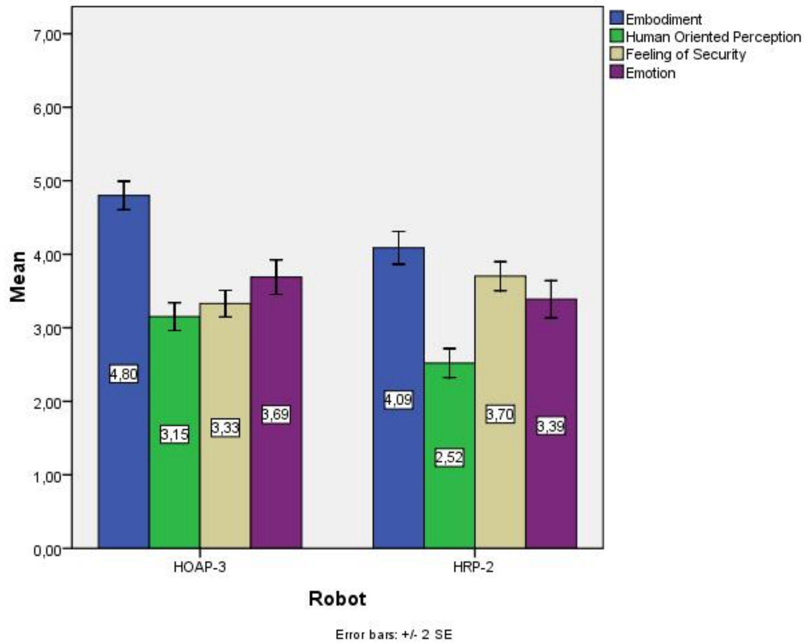
Figure 5.6: Differences SoAc: Male and Female



Differences between the HOAP-3 and the HRP-2 Robot The user experience indicators embodiment ($z = -4.80, p = 0.00$) and human-oriented perception ($z = -4.77, p = 0.00$), and emotion ($z = -1.97, p = 0.05$) were rated significantly higher for the HOAP-3 robot than for the HRP-2 robot (see Figure 5.7). These results correspond with data gathered in the "Retrospective Think Aloud" of the respective HOAP-3 user study in which the HOAP-3 robot was often mentioned as cute and likable. However, the indicator feeling of security was rated significantly better for the HRP-2 robot ($z = -2.51, p = 0.01$). This could be due to the chosen video material, as in the presented HRP-2 user study, no direct contact interaction happened between the user and the robot and the robot was always accompanied by an engineer to assure the users' safety.

In the user study with the HOAP-3 robot, no accompanying engineer was present and direct contact with the robot was required to solve the task.

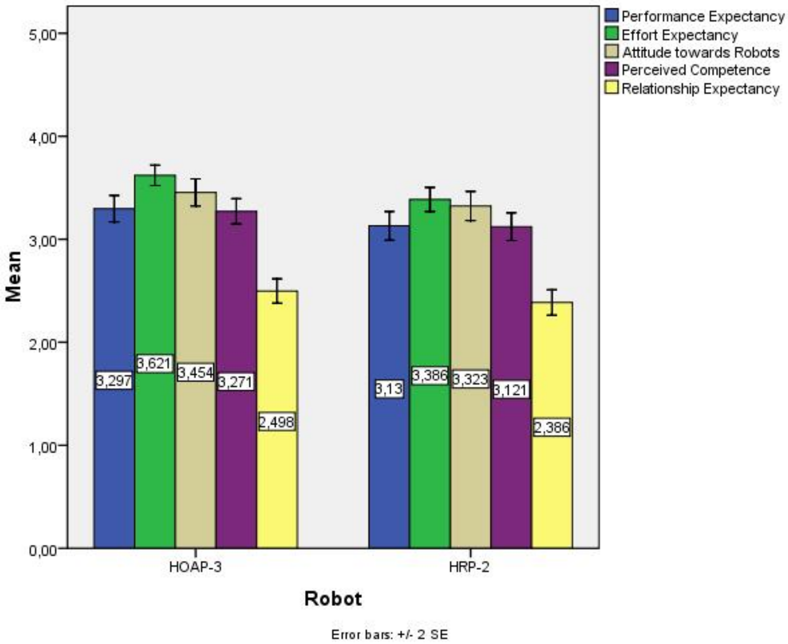
Figure 5.7: Differences UX: HOAP-3 and HRP-2



The social acceptance indicator effort expectancy ($z = -2.97, p = 0.00$) was also rated higher for the HOAP-3 robot than for the HRP-2 robot (see Figure 5.8) which mirrors the findings from the respective user study (see Section 4.5 and Weiss et al (2009e)), revealing that learning by demonstration is perceived as a very natural and intuitive way of interacting which costs little extra effort. Nevertheless, it was surprising that the effort expectancy of a learning by demonstration scenario was rated better than by a speech command based scenario, which is often referred to as easiest way of interaction. For the other social acceptance indicators, no significant differences could be revealed.

Correlations with Age Regarding the independent variable "age", no significant correlation could be found for the user experience and social acceptance indicators. Only the social acceptance indicator attitude towards robotics showed a weak but significant correlation (Spearman´s rho=0.106, p=0.04, n=395). The older the participant, the more positive the attitude towards robots was rated.

Figure 5.8: Differences SoAc: HOAP-3 and HRP-2



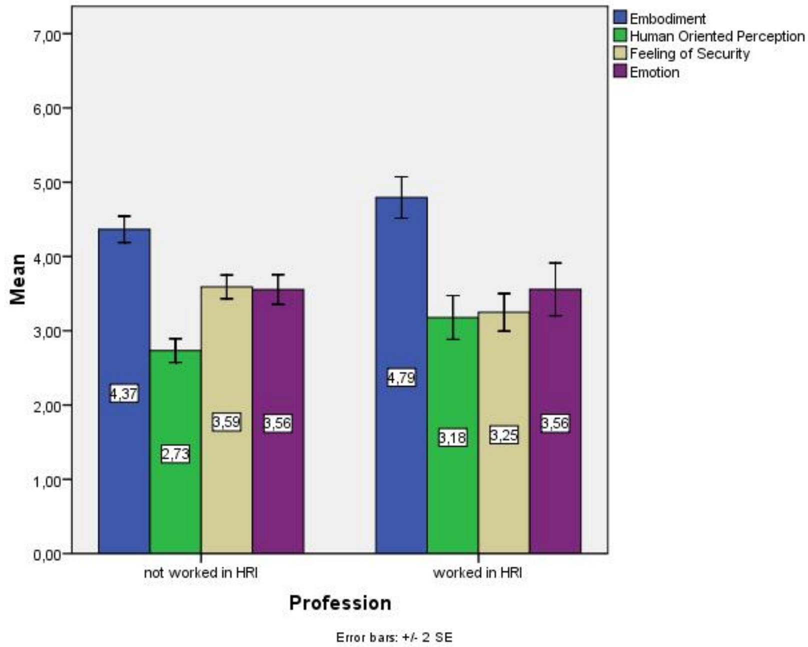
For the roboticists, this outcome could appear as a surprise, as normally young adults are expected to be the more positive attuned age group. Young adults use more automation and communication technology in there everyday life style and are attracted by the gadget character of robots considering robotic technology as status symbol. In comparison, elders are expected to reject robotic technology; as by acceptance they would have to admit their personal loss of autonomy and the need of a robot, which conducts specific activities instead of them (comparable to hearing devices).

Moreover, it is assumed that elders cannot keep track with technological development and thus are more skeptically attuned. There is a reasonable way to explain the positive correlation between age and attitude towards robots. The result corresponds with the findings from Arras and Cerqui (2005) again. The older people get, the more likely they accept robots which will assist them in loss of independence. In other words, robots could help to maintain everyday live autonomy and thus increase the quality of life.

Differences between "Not Worked" and "Worked in HRI" The analysis regarding the independent variable "profession" (meaning if participants already worked in HRI or not) revealed significant differences for the user experience indicators embodiment ($z = -2.33, p = 0.02$) and human-oriented perception ($z = -2.72, p = 0.01$). Both indicators are rated significantly higher by participants who have already worked in HRI (see Figure 5.9). This again shows the impact of foreknowledge. Participants who know more about robots, have a more positive answering behavior (see Hammel et al (1989)).

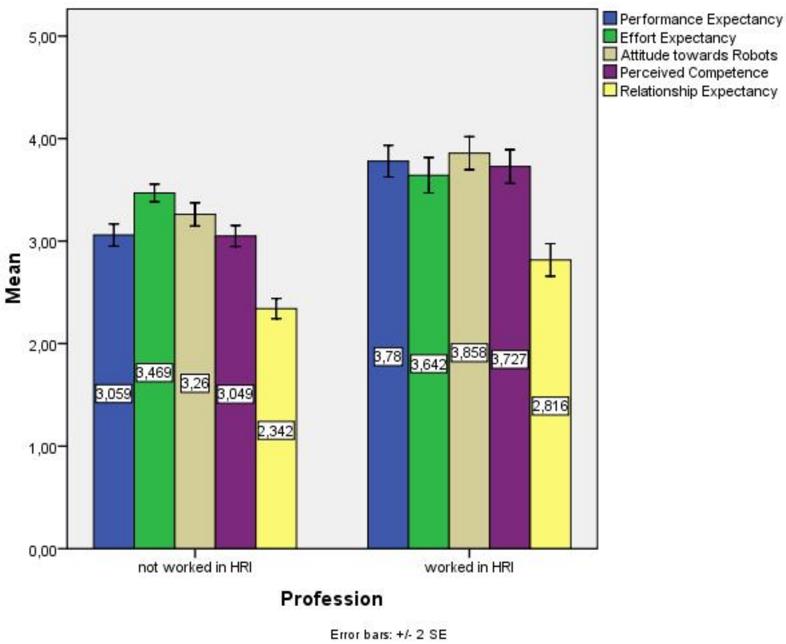
As the current tendency in the research community of HRI is to design humanoid robots, embodiment is rated higher by people who have already worked in that field. Similarly, having knowledge about the technological challenges of implementing human-oriented perception strategies into robotic design, could explain the significantly higher rating of this indicator by experienced participants. It is noteworthy that feeling of security is rated significantly higher by participants who have not worked in HRI ($z = -2.08, p = 0.04$). This can be explained with in a similar way as the higher rating of this indicator by female participants: Less technological background knowledge leads to a higher feeling of security (taking security for granted).

Figure 5.9: Differences UX: Not Worked and Worked in HRI



The strong link between the participants foreknowledge and their acceptance of robotic technology which was already found in the data of other studies (see Hammel et al (1989)) is also depicted in the results of the online survey (see Figure 5.10). Participants who have already worked in HRI rated all indicators of social acceptance significantly higher than participants who never worked in HRI (all $z > -2.33, p < 0.02$).

Figure 5.10: Differences SoAc: Not Worked and Worked in HRI



Differences due to Participation in HRI Experiments The analysis revealed highly significant differences within user experience and social acceptance indicators, depending on whether participants had already taken part in an HRI experiment before or not, which mirrors the results regarding the independent variable "profession" (see above and Figure 5.11; Figure 5.12). The user experience indicators embodiment ($z = -2.59, p = 0.01$) and human-oriented perception ($z = -3.50, p = 0.00$) were rated significantly better by participants who had already taken part in an HRI experiment, whereas feeling of security was rated significantly better by participants who did not take part in HRI experiments so far ($z = -2.67, p = 0.01$).

Furthermore, survey participants who already took part HRI user studies rated all social acceptance indicators significantly higher. They expected higher performance ($z = -7.61, p = 0.00$), less effort ($z = -2.69, p = 0.01$), as well as better relationship ($z = -6.27, p = 0.00$) when working with robots. There also showed a higher positive attitude towards robots ($z = -7.04, p = 0.00$) and perceived competence ($z = -6.31, p = 0.00$) than the participants who have not taken part in an HRI experiment to date. These results correlate with the data gathered in the user studies, where the interaction with the robots decreased the values of the Negative Attitude towards Robots Scale questionnaire.

Figure 5.11: Differences UX: Participation in HRI Experiments

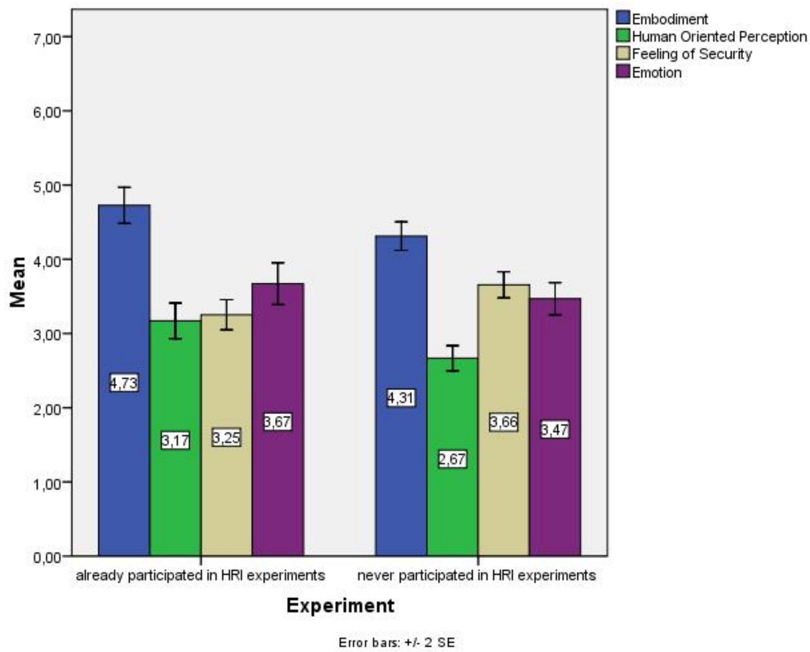
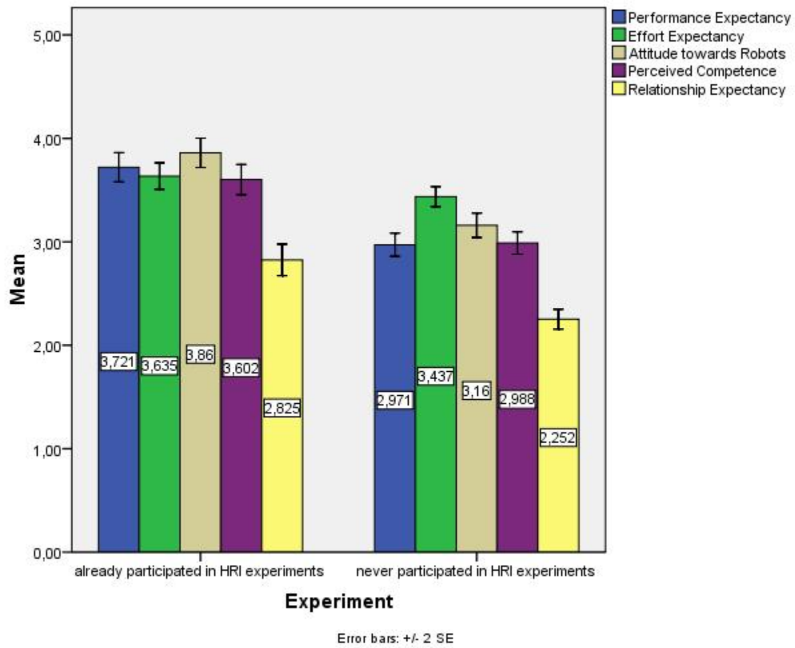


Figure 5.12: Differences SoAc: Participation in HRI Experiments



Correlations between the User Experience and Social Acceptance To investigate the interrelations between the user experience and social acceptance indicators, correlation coefficients (Spearman's roh) were computed among the nine indicator scales. The results of the correlation analyses presented in Table 5.5 (at the end of this chapter), show that 52 out of the 72 correlations were highly significant ($p < 0.01$) and 14 were greater or equal than 0.409. The correlations of the indicator emotion tended to be lower. In general, the results suggest that the indicators of social acceptance and user experience are highly interrelated, which goes in line with the opinion of some HCI researches that user acceptance is part of user experience. Furthermore, the results indicate towards the remark stated in Chapter 3 that the theoretical indicators of the USUS evaluation framework cannot be viewed independently.

5.3.3 The Reference Model

The next section is based on the findings of the factor, differences, and correlation analysis. The following reference model for the indicators of the user experience and social acceptance factor of the theoretical USUS framework can be derived (see Figure 5.13).

5.4 Reflection on the Reference Model

The initial conception of the USUS evaluation framework (see Chapter 3) was based on a literature review and existing evaluation concepts in HRI research to come up with relevant indicators to assess usability, social acceptance, user experience, and societal impact in human-robot interaction studies. However, an initial attempt of setting up a theoretical evaluation framework needs empirical validation. The eleven case studies presented in Chapter 4 had, as their main goal, to assess the feasibility of the proposed methodological framework to address the proposed indicators. These evaluation studies could only partly address the interrelations between the indicators. Therefore, the broad online survey was conducted on one hand to validate the UX- and SoAc-Questionnaire. On the other hand to investigate the relationship between the two concepts social acceptance and user experience.

The successful validation of the UX- and SoAc-Questionnaire as a result provides the HRI research community with a valid tool to assess a humanoid robot in terms of all revised indicators of social acceptance and user experience. However, even if an evaluation study should only focus on some of the indicators, e.g. only the user experience indicator embodiment plus the social acceptance indicators relationship expectancy and effort expectancy, the ref-

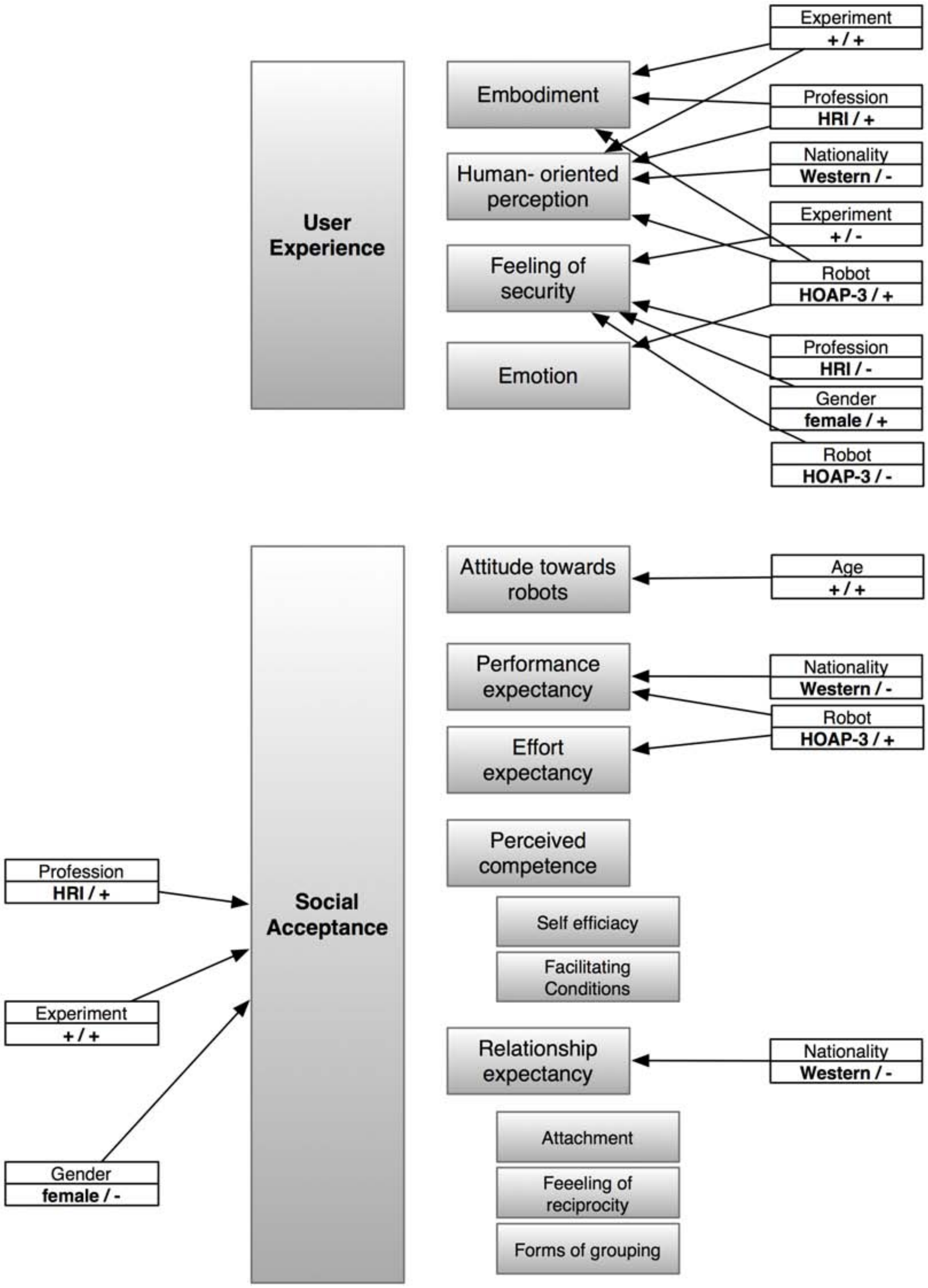


Figure 5.13: The Reference Model

erence model can be used. In that case, there is a need to adapt the revised theoretical model. Consequently, the questionnaires need to be revalidated taking into account the significant correlations between the user experience and social acceptance indicators. Processing the results of the questionnaire would then include: First, conducting a factor analysis with a rotation component matrix to check if the same items still "load" on the same factor. Establishing Cronbach's alpha for the items of each scale would be the second step. If a construct consists of more than two statements, it can be used to test the consequences of omitting an item (based on a theoretical solid assumption), especially if the alpha is not high enough. Final step, calculating the scores for each scale by averaging the scores on the items.

It will take more work to validate the complete USUS evaluation framework than the online survey and the eleven case studies presented in this PhD thesis. The given examples give a good indication of the possibilities of the evaluation framework, as well as it advantages and limitations. Above all, the inter-indicator relations and further influencing variables add to the possibilities.

The main advantage of the revised model is the conjunction between state-of-the-art literature and an empirical proof-of-concept study which enables other HRI researchers to conduct theory-based evaluation studies with validated measures. Thus, the model supports repetitive testing and allows simple investigation of quality factors like relationship expectancy and perceived competence. A first attempt was made to investigate relevant influences on the USUS evaluation framework, which is rarely done for other evaluation frameworks, namely the influence of the six independent variables: "nationality" (expressions: Asian/Western), "gender", "age", "profession" (expressions: worked in HRI/not worked in HRI), "experiment" (expressions: already participated in HRI experiments/never participated in HRI experiments), and "robot" (expressions: HRP-2 robot/HOAP-3 robot). The revised reference model can therefore support the recruitment procedure for user studies and a focus can be put on the data interpretation. The revised reference model can be effectively used to compare different humanoid robots with regard to the social acceptance and user experience indicators.

The main limitation is a need for a high number of participants if a revision of the model and the questionnaires are needed to address specific research questions. This is inherent to the statistical procedures of the factor analysis and Cronbach's alpha. However, the revised reference model can still be used for qualitative studies with a small number of participants, especially in combination with other methods proposed in the revised methodological framework (see Chapter 6).

Table 5.1: Factor Analysis: User Experience

Assumed Factor	Item				
	Embodiment (EMB)	EMB	FS	HOP	E
EMB	I liked the size of the robot.	**0.503**	-0.095	0.246	0.007
EMB	I liked that the robot looked similar to a human.	**0.793**	-0.080	0.194	-0.077
EMB	I liked that the robot has human-like features: e.g. face. ears. eyes etc.	**0.873**	-0.027	0.175	-0.064
EMB	I liked the design of the robot.	**0.598**	-0.150	0.273	-0.061
	Feeling of Security (FS)				
FS (R)*	I think that the robot is vulnerable to hackers.	0.008	**0.454**	-0.058	-0.080
FS (R)*	I hesitated to use the robot for fear of making errors that will harm me.	-0.023	**0.686**	-0.107	-0.069
FS (R)*	I feared to use the robot as an error might harm the robot.	-0.039	**0.658**	0.287	-0.038
FS	I perceived the robot as safe.	-0.226	**0.553**	-0.415	-0.019
	Human-Oriented Perception (HOP)				
HOP	I perceived the robot as a social actor.	0.355	-0.074	**0.518**	0.012
HOP	I perceived that the robot is intelligent.	0.229	0.035	**0.602**	-0.052
C*	The robot could become a companion for me.	0.288	-0.137	**0.487**	0.045
	Emotion (E)				
E	I would be angry if the robot did not understand my commands.	0.056	-0.145	0.131	**0.979**
E	I would be disappointed if the robot did not understand my commands.	-0.150	-0.085	-0.101	**0.614**
	Deleted Factors				
FS	I would feel secure when working with the robot.	0.210	-0.581	0.410	-0.005
E	I would feel afraid of the robot.	0.124	-0.594	0.028	0.061
*C:	Co-experience				
*(R):	revers				

Table 5.2: Factor Analysis: Social Acceptance

Assumed Factor	Factor/ Item	PE	AT	RE	EE	PC
	Performance Expectancy (PE)					
PE	The deployment of robots will increase my chances of success in my job.	**0.580**	0.314	0.109	-0.012	0.192
PE	I will solve tasks faster using robots.	**0.694**	0.100	0.014	0.218	-0.024
PE	Robots will make my tasks easier.	**0.796**	0.138	0.144	0.179	0.042
PE	Robots will facilitate burdensome tasks we have now.	**0.487**	0.282	0.067	0.170	0.029
FG*	Robots and humans will make a good team.	**0.477**	0.418	0.320	0.225	0.132
PE	I would find robots in my job useful.	**0.656**	0.323	0.187	0.146	0.232
PE	The utilization of robots will increase my productivity.	**0.768**	0.272	0.139	0.214	0.111
	Effort Expectancy (EE)					
EE	It will be easy for me to become skillful in dealing with robots.	0.137	0.329	0.207	**0.336**	0.302
EE	Interacting with robots will be easy to understand.	0.208	0.059	0.152	**0.784**	0.094
EE (R)*	It would be difficult to learn how to handle robots.	0.058	0.048	0.015	**0.654**	0.173
EE	It will be easy to use robots.	0.203	0.140	0.111	**0.758**	0.148
EE (R)*	Efforts to solve tasks together with robots will be a huge undertaking.	0.108	0.141	-0.018	**0.489**	0.070
FC*	There will be enough information material available to help simplify the interaction with robots.	0.191	0.228	0.150	**0.506**	0.065
	Relationship Expectancy (RE)					
FR*	The interaction with robots will be a mutual experience.	0.309	0.008	**0.459**	0.166	0.136
FR*	I can imagine that I will care for the wellbeing of a robot.	0.129	0.503	**0.404**	0.110	0.237
AT	I can imagine to take a robot into my heart.	0.000	0.309	**0.733**	0.029	0.033
FG*	Robots will have a similar importance as human colleagues	0.201	0.039	**0.655**	0.051	0.012
A*	It would feel good if a robot was near me.	0.221	0.383	**0.590**	0.132	0.115
A*	I can imagine building a special relationship with robots.	0.055	0.294	**0.798**	-0.001	0.042
FR*	The relationship with robots will be based on the principle of give and take.	0.059	0.049	**0.518**	0.042	0.061
FG*	Robots will be an important part of our society.	0.402	0.136	**0.413**	0.186	0.079
	Perceived Competency (PC)					
SE* (R)*	When a problem occurs with robots. I would not be able to continue with my work without help.	0.021	0.102	-0.037	0.111	**0.549**
FC* (R)*	I do not have the necessary abilities to handle robots.	0.149	0.139	0.050	0.052	**0.685**
SE*	I could solve all problems which occurred during the interaction on my own.	0.085	0.079	0.150	0.118	**0.600**
FC*	I have the necessary knowledge to handle robots.	0.269	0.225	0.188	0.028	**0.811**
SE* (R)*	I could not solve a task with the help of robots if no one was there to tell me what to do.	-0.065	0.110	0.032	0.222	**0.424**
	Attitude towards Robots (AT)					
AT (R)*	I would not like to work together with robots.	0.342	**0.593**	0.189	0.141	0.205
FG*	I would like to collaborate with robots.	0.406	**0.715**	0.341	0.208	0.149
AT	Robots will make work more interesting.	0.483	**0.462**	0.263	0.124	0.094
AT	Working with robots would be fun.	0.284	**0.663**	0.320	0.228	0.209
A*	I would not like to imagine a world in which robots were not used.	0.159	**0.350**	0.199	0.092	0.213
AT (R)*	Using robots is a bad idea.	0.272	**0.649**	0.114	0.204	0.153
AT (R)*	I would be afraid to employ robots at work.	0.142	**0.564**	0.038	0.367	0.225
	Deleted Factors					
(R)*	I could not be successful while working with robots under time pressure.	0.129	0.191	0.050	0.227	0.149
	Robots will be part of our everyday work.	0.483	0.274	0.315	0.188	0.073
	Humans and robots will be interdependent.	0.403	-0.003	0.313	0.048	0.093
(R)*	I will never be able to solve a task together with a robot.	0.280	0.319	0.101	0.325	0.145
	I would be motivated to integrate robots in my daily workday.	0.472	0.562	0.268	0.269	0.197
	If problems with robots occur there would be persons who could help me.	0.126	0.264	0.048	0.343	-0.012
(R)*	I do not think it is necessary to employ robots in daily working life.	0.369	0.294	0.007	0.230	0.043
*FC:	Facilitating Conditions					
*SE:	Self Efficacy					
*FR:	Feeling of Reciprocity					
*A:	Attachment					
*FG:	Forms of Grouping					
*(R):	revers					

Table 5.3: Internal Consistency: UX- and SoAc-Scales

Scale	Cronbach's Alpha	Items
Embodiment	0.83	4
Human-Oriented Perception	0.70	3
Feeling of Security	0.68	4
Emotion	0.72	2
Performance Expectancy	0.88	7
Effort Expectancy	0.80	6
Attitude towards Robots	0.89	7
Perceived Competence	0.78	5
Relationship Expectancy	0.80	6

Table 5.4: Descriptive Statistics: UX- and SoAc-Scales

Indicator	min	max	mean	S.E.	S.D.
Embodiment	1	7	4.43	0.08	1.51
Human-Oriented Perception	1	7	2.83	0.07	1.39
Feeling of Security	1	7	3.53	0.07	1.35
Emotion	1	7	3.54	0.09	1.72
Performance Expectancy	1	5	3.21	0.05	0.94
Effort Expectancy	1	5	3.51	0.04	0.77
Attitude towards Robots	1	5	3.38	0.05	0.96
Perceived Competence	1	5	3.20	0.05	0.90
Relationship Expectancy	1	5	2.45	0.04	0.85

Table 5.5: Correlations: UX- and SoAc-Indicators

		EMB	HOP	FS	E	PE	EE	AT	PC	RE
EMB	roh	1.00	0.45**	-0.20**	-0.12*	0.34**	0.30**	0.43**	0.05	0.42**
	p		0.00	0.00	0.01	0.00	0.00	0.00	0.38	0.00
	n	398	398	398	396	397	397	397	392	397
HOP	roh	0.45**	1.00	-0.10*	-0.06	0.42**	0.21**	0.39**	0.07	0.48**
	p	0.00		0.04	0.27	0.00	0.00	0.00	0.19	0.00
	n	398	398	398	396	397	397	397	392	397
FS	roh	-0.20**	-0.10*	1.00	-0.14**	-0.26**	-0.44**	-0.41**	-0.36**	-0.25**
	p	0.00	0.04		0.01	0.00	0.00	0.00	0.00	0.00
	n	398	398	398	396	397	397	397	392	397
E	roh	-0.12*	-0.06	-0.14**	1.00	-0.03	-0.02	-0.05	0.05	-0.12*
	p	0.01	0.27	0.01		0.60	0.68	0.33	0.37	0.01
	n	396	396	396	396	395	395	395	390	395
PE	roh	0.34**	0.42**	-0.26**	-0.03	1.00	0.47**	0.71**	0.30**	0.56**
	p	0.00	0.00	0.00	0.60		0.00	0.00	0.00	0.00
	n	397	397	397	395	397	397	397	392	397
EE	roh	0.30**	0.21**	-0.44**	-0.02	0.47**	1.00	0.57**	0.44**	0.40**
	p	0.00	0.00	0.00	0.68	0.00		0.00	0.00	0.00
	n	397	397	397	395	397	397	397	392	397
ATT	roh	0.43**	0.39**	-0.41**	-0.05	0.71**	0.57**	1.00	0.42**	0.62**
	p	0.00	0.00	0.00	0.33	0.00	0.00		0.00	0.00
	n	397	397	397	395	397	397	397	392	397
PC	roh	0.05	0.07	-0.36**	0.05	0.30**	0.44**	0.42**	1.00	0.27**
	p	0.38	0.19	0.00	0.37	0.00	0.00	0.00		0.00
	n	392	392	392	390	392	392	392	392	392
RE	roh	0.42**	0.48**	-0.25**	-0.12*	0.56**	0.40**	0.62**	0.27**	1.00
	p	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	
	n	397	397	397	395	397	397	397	392	397

Chapter 6

Application of the USUS Evaluation Framework

6.1 Outline of the Chapter

This chapter will in the beginning describe methodological challenges which had to be faced during the conduction of the eleven case studies presented in Chapter 4. Based on this, some methodological considerations will be presented which should be taken into account when applying the proposed framework on other HRI studies. The scenarios on how the application can look like, how it can be supported by evaluation guidelines, and how the framework is positioned in the user-centered design cycle will conclude this chapter.

6.2 Methodological Challenges

The major methodological challenges which were encountered during the work on this PhD thesis were the following:

System Functionalities: The humanoid robots investigated for this PhD thesis offered a broad variety of functionalities. However, it is not possible in one case study to investigate all functionalities in a single experimental design, as this would produce a never ending list of confounding variables. Therefore, it is necessary to limit the functionalities of a robot in controlled experiments. This, however, stands in contradiction with the expectations participants have to a humanoid robot, as they expect, for instance, that if a robot understands one voice command, it will understand voice commands in general. One has to walk a tightrope in the experimental design to limit the robot's functionalities without influencing the participants' expectations too much. Furthermore, experimental settings should still be close enough to reality to obtain valuable results.

System Comparability: In traditional HCI studies, the comparison between different similar systems is a common approach to conduct benchmark studies. This approach is partly possible in Human-Robot Interaction studies. Robotic platforms are very different in their capabilities, particularly humanoid robots, as they are task independently designed and usable in a variety of different interaction scenarios. This offers a high flexibility in study settings and measures to be taken. On the other hand this limits the comparability of study results. It will still need a lot of effort for the research field to establish common measures, metrics, and benchmarks to ensure comparable study results and to foster the standardization of robotic system (parts).

Observing Natural HRI: A major issue was also to gather reasonable data, as some of the indicators of the theoretical framework were hard to collect in lab settings. Thus, a lot of effort has been put in combing different methods and in combined data analysis like the interpretation of interview data together with video data analysis. However, most of the data gathered in this PhD thesis is self-reported participant data from a reflective level. For future research, additional emphasis should be put on field trails, in situ methods like physiological feedback, and testing with autonomous robots in more loose settings.

Participant Sampling: Participant sampling was a critical issue through the whole "Robot@cwe project". The main criteria which always had to be fulfilled were no pre-experience in robotics and counterbalancing in gender. However, since in several of the conducted studies the participant recruiting was taken over by project partners, the criteria were not always met exactly. A major problem was that in many cases students were recruited (as this is easy to do for university partners), meaning it is difficult to balance education (a furtherer influence factor was the language barrier, which is described in the following). Additionally, frequently students from related university departments were recruited, meaning they had no pre-experiences in robotics, but a strong background in computer science or engineering.

Language Barrier: A big challenge in the conduction of the case studies presented in this thesis (see Chapter 4) were the different languages the participants spoke. Even though most of the time questionnaires were available in the respective mother tongue, all studies were conducted in English or German. This precondition on one hand influenced the characteristics of potential participants who could be recruited, as good English or German skills were required. On the other hand, interview data is limited in its depth to expression possibilities. It would have been ideal if all of the experiments could have been conducted in the mother tongue of the participants and then be translated into German as the loss of data due to translation would have been lower than the loss due to communication difficulties.

6.3 Methodological Considerations

The major methodological considerations which should be taken into account when applying the framework in other research projects are summarized in the following :

Context and Environment Factors: For a successful laboratory- based user study, the interaction environment should realistically reflect the real setting so that participants respond in a "natural" manner. As it is in some cases not possible to place the participant in the actual interaction context, like in the second user studies with the HOAP-3 robot which was situated in a space environment (see Section 4.6). In such cases it is important to simulate that context as close a s possible. For example, simple cues like ambient construction site noise, supported the contextual situatedness in the mixed-reality prototype study (see Section 4.7).

Adaptation of the Theoretical Framework: For different robots and interaction scenarios/contexts, a place to start is with a qualitative approach to reshape the factor indicator model. It needs to be checked if all required indicators are identified. In general, the selection of the robot is a key element of successful HRI studies. Clearly, the selection must be in accordance with the application and the domain which is investigated. One step further is to investigate in a comparative study setting if the embodiment of the robot influences other USUS indicators, similar to the analysis of the online survey (see Section 5.3.2). Future studies in this direction would allow to generalize the impact of USUS indicators and to identify which effects are specific to particular robots (e.g. HOAP-3 vs- HRP-2) or to robot types in general (e.g. zoopomorphic vs. anthropomorphic).

Participant Sampling: Recruiting the right participants and having a representative sample is a challenge all social research studies have to face.

As it is a matter of resources (mainly participant compensation and time) to conduct HRI studies with big sample sizes, participant profiles should be defined as precisely as possible.

As the preliminary user study with the HRP-2 robot (see Section 4.4) showed, even the profession of participants could have a strong influence on the soft factors, user experience, social acceptance and societal impact. To increase the sample size, video-based studies can be beneficial, as two case studies (see Section 4.2, Section 4.3) and also the online survey (see Section 5.3) demonstrated. Similar approaches can also be found in literature, e.g Woods et al (2006).

Tools and Equipment: The choice of data capturing equipment is not only crucial for the successful conduction of a user study, but also for a structured data analysis. For instance, it became obvious during the data analysis of the second HOAP-3 user study that it is not only necessary to determine beforehand the required video perspectives, but to ensure that the videos can be synchronized to allow a time lined analysis of the comprehensive video material and not only of single perspectives.

It is advisable for the data analysis of various observational data sources (video data, screen logging data, HRV data, think aloud data etc.) to follow a systematic time lined approach similar to the SInA approach developed by researchers of the University of Bielefeld Lohse et al (2009). Using this approach for the data analysis of the Human-Robot Collaboration scenario with HRP-2 (see Section 4.8) allowed a more comprehensive data analysis and thus provided the most valuable recommendations for the engineers to improve the HRI scenario.

Experimental Settings: Difficult indicators should be addressed by a method mix. A combination of qualitative and quantitative methods along with a combination of self-reporting measures on the reflective level and in situ methods at the visceral level is the best ease mix. A single evaluation method is not sufficient to capture all relevant aspects of HRI scenarios. Furthermore, within a single method (e.g. questionnaire, interview etc.) multiple indicators should be addressed; ideally considering various factors which have rarely been done for the USUS evaluation framework thus far.

However, the first trial in the focus group study suggested that this would be a valuable approach to better understand the interrelations between the indicators and to identify missing indicators. The case studies proved that a task-based approach is valuable to investigate HRI scenarios and that a learning and explanation phase before the actual task conduction, produces more meaningful data. When testing with novice users, it is advisable to investigate an attitude change before and after interacting with the robot.

Using the UX and SoAc-Questionnaire: All questionnaires provided in this thesis should be checked on scale reliability and factor coherence, described in Section 5.4. Questionnaire translations should at least be conducted by accredited translators or, in the ideal case, by native speakers of the respective language (as only native speakers know the true meanings of phrases in their language and with the back translation method).

Furthermore, it is recommended for the future usage of the UX-Questionnaire to avoid questions with double negatives (namely "I would be angry if the robot did not understand my commands" and "I would be disappointed if the robot did not understand my commands") and to reformulate these two questions, into e.g. "I would be happy if the robot understands my commands" and "I would be pleased if the robot follows my commands". Moreover, for cross-cultural usage of the UX-and SoAc-Questionnaire the adherence of cross-cultural guidelines for the survey design is recommended (see Harkness et al (2003)).

Data Analysis: Data gathered on the theoretical framework should not be analyzed or isolated for one factor, but possible synergies, interrelations, and dependencies between indicators and factors should be considered. The interpretation of quantitative data should always be accompanied by knowledge of qualitative data and the relevant literature. Furthermore, it is important to be aware that the results of evaluation activities are grounded within the experimenter's personal judgment.

6.4 Revised Methodological Framework

Choosing the right method for an indicator and composing a reasonable method mix can be a major pitfall in the application of the USUS evaluation framework. In Section 3.7 six methods were proposed to address all indicators of the theoretical framework. Based on lessons learned in the case studies and the general considerations and challenges a revised methodological framework can be proposed (see Table 6.1).

The revised methodological framework takes into account the theoretical reference model from Figure 5.13), including the assignment of methods to indicators which were already investigated in terms feasibility by means of the eleven case studies (see Chapter 4), and newly proposed methods. Approved methods are marked with an X and novel methods which still need to be approved in future work, with an N.

However, it is essential for the newly proposed methods (just as much as it was for the originally proposed ones) to adapt them to the evaluation context and the system capabilities. Similarly also the newly proposed methods should not be used as stand-alone methods, but in combination with others to achieve a balanced method and indicator mix. All newly proposed methods are described in more detail in the following sections.

Factor	Indicator	Expert Eval.	User Studies	Questionnaires	Physio. Measures	Focus Groups	Interviews	Partic. Design
Usability	Effectiveness	X	X					
	Efficiency	X	X					
	Learnability	X	X					
	Flexibility	X	X					
	Robustness	X	X					
	Utility			N			X	
Social Acceptance	Performance							
	Expectancy			X		N	N	
	Effort Expectancy			X		N	N	
	Attitude toward						N	
	using Technology			X		X	N	
	Perceived Competence			X		N	N	
	Self Efficacy			X		X	N	
	Facilitating Conditions			X		N	N	
	Relationship Expectancy			X		N	N	
	Forms of Grouping			X		X	N	
	Attachment			X		X	N	
	Reciprocity			X		X	N	
User Experience	Embodiment			X		N	N	N
	Emotion			X	X	N	N	
	Human-oriented							
	Perception			X		X	N	N
	Feeling of Security			X	N	N	N	
	Co-Experience			X		N	N	N
Societal Impact	Quality of Life			N		N	X	
	Working Condition							
	and Employment			N		N	X	
	Education			N		N	X	
	Cultural Context			N		N	X	

Table 6.1: The Revised Methodological Framework

6.4.1 Participatory Design

Due to the success of the “Robot Building Set” for assessing the user experience indicator embodiment and due to the fact that user-centered design proved its value for HRI in general, participatory design is added as a novel methodological approach to the framework. Participatory design is a methodological design approach which is also called the “Scandinavian approach” in HCI (as it was developed by Scandinavian researchers). In this approach designers develop new systems together with potential users in a playful manner. Thus, this approach enhances the communication and teamwork through shared enjoyment of working in a game-like setting and it improves the knowledge about the user requirements.

Application

Participatory design should be integrated in a very early phase of the design process, as it should foster the mutual learning process between all active participants, such as engineers, computer scientist, psychologists, and potential users. For this design approach low- as well as high-fidelity prototypes can be used to enhance the user involvement, such as videos, mixed-reality prototypes, and mock-ups. Storytelling techniques and creativity techniques can be used for developing and iterating the prototypes.

Advantages and Disadvantages

This method could offer the designers of HRI scenarios new insights, leading to the improvement of user experience indicators, such as embodiment, human-oriented perception, and co-experience. However, a major limitation of this method is that it only reveals qualitative design suggestions and no measurable quantitative results.

6.4.2 Focus Groups: World Cafè Approach

The world cafè approach is an approach similar to that of a focus group. It is a conversational process based on a set of integrated design principles (for more details see http://www.theworldcafe.com/what.htm). This specific form of discussion groups could give additional insights on the social acceptance and user experience indicators. Moreover, if the focus of the world cafè is the development of an imaginary narrative of the “future robotic society”, it could be a valuable approach to investigate the indicators of societal impact.

Application

In a world cafè setting all participants are sitting around discussion tables (4-5 participant for each table, at least 3 tables). The table is covered with flip chart paper, therefore participants can take notes during the discussion. In one world cafè session (lasting between 15-30 minutes) each table discusses one topic or question, which was presented by the world cafè moderator beforehand. After each session the participants move to another table. Only the table host stays at the same table, to summarize the discussion results of the previous group for the next participants. A world cafè ends, after all participants talked about all topics and the table hosts presented the overall table results to the group.

Advantages and Disadvantages

The world cafè is a suitable method for getting a discussion about open ended questions started. It is also suitable to reflect on topics. The method is especially effective, if the discussion groups are heterogeneously composed. The disadvantages of the method are similar to those of the focus group method (see SubSection 3.7.5): Only the opinion of a small group of people can be represented and it is highly important to select a representative sample for a successful discussions. Furthermore, good facilitation skills of the table hosts are required to communicate the contributions of the previous participants.

6.4.3 User Study Analysis

As mentioned already in Section 6.3, the SInA approach Lohse et al (2009) is recommended to analyze behavior observation data gained in user studies. It also allows the combined analysis of several synchronized video perspectives and physiological data.

Application

The SInA approach is based on a three step procedure. First, an interdisciplinary team of researchers takes a look on the video material gathered during a user study. They perform an interaction analysis to reveal the relevant interaction episodes. Based on this a task analysis can be conducted to identify "prototypical interaction sequences". In a second step, cases which deviate from the prototypical interaction sequences are determined and "deviation patterns" are defined. Based on the quantification of the prototypical interaction sequences and the deviation patterns, implications for design can be derived to cope with interaction problems.

Advantages and Disadvantages

The SInA approach is very useful for a systematic analysis of observational data, gathered in user studies. It helps to bridge the gap between evaluation and implementation, as guidelines for the next iteration can be directly derived, by comparing prototypical interaction sequences with deviation patterns. The main disadvantage is that the approach is very time consuming due to the required annotation of video material by an interdisciplinary research team. Moreover, the critical point is that the coding scheme for the video annotation needs to be interpreted as similar as possible by all researchers to guarantee a valid identification of prototypical interaction sequences and deviation patterns.

6.4.4 Questionnaire on Societal Impact

Regarding societal impact an improvement would be the development of a standardized questionnaire to measure the effect of robots on society. A first step into this direction was already undertaken in the FP6 EU-Project "Robot-@cwe:Advanced robotic systems in future collaborative working environments" Weiss et al (2009g). However, the first developed questionnaire was only a pretest with ten employees of the DRAGADOS company. After an iteration of this questionnaire a bigger survey should be conducted to check the scale reliability and validity.

Application

The ideal application for a structured questionnaire on societal impact would be a context, in which robots are already introduced for collaboration. This context would allow getting insights on societal aspect from a perspective of experienced users.

Advantages and Disadvantages

The usage of a questionnaire on "societal impact of robots" in a working context could, on one hand, provide a large scale sample size and thus allow an in-depth statistical data analysis. On the other hand, it has to be considered that the survey context could bias the results, as employees could feel forced to give socially desired answers, therefore they will not get unemployed.

6.5 Guideline-based Application Process

As already pointed out in objective four of this PhD thesis, a major focus is put on the facilitation of HRI evaluation that goes beyond usability assessment and actually takes into account the holistic interaction experience of human-robot interaction. In the following, guiding research questions are provided for each indicator, which should support other researchers in the application of the theoretical and methodological USUS framework. Theses guiding research questions should assist researchers with the planning, design, execution, and data analyzes of human-robot interaction.

6.5.1 Guiding Research Questions: Usability

1. Effectiveness: Are the means of interaction provided by the robot effective for the human?

   Determine the relevant tasks the user should be able to perform with the robot and offer the required functionalities, therefore the user can carry them out. Assess the effectives of the human-robot interaction scenario, e.g. with measures such as "success rate", "error rate", and "task completion rate". Alternatively, you can use expert evaluation methods.

2. Efficiency: Do the provided means of interaction allow efficient task solving for the human?

   Provide intuitive input modalities, therefore the user can solve the tasks in a resource-saving way. Assess the efficiency with measures such as "task duration" and "required clicks". Alternatively, you can use expert evaluation methods.

3. Learnability: How easy can the human learn the interaction with the humanoid robot?

   Provide an intuitive, easy to learn interaction paradigm, which fits the human-robot interaction scenario (e.g. speech interaction is not necessarily the ideal choice for every scenario). Use expert- or user-based evaluation methods to assess the users' cognitive load during the interaction.

4. Flexibility: Does the robot allow enough ways for solving a collaborative task?

   Provide several ways how users can complete a task, to support various mental models. Use expert- or user-based evaluation methods to assess the flexibility of your human-robot interaction scenario and to identify other possible completion strategies for the defined tasks.

5. Robustness: Does the robot system provide enough support for the user to successfully solve a collaborative task?

   Provide the users with help functionalities, which support them in recovering from errors without the help of an expert. Assess, if the visibility of the system status allows the user to recover, by means of expert evaluations. Moreover, investigate by means of user studies, how users solve interaction problems in complex situations.

6. Utility: Is it possible to achieve planned collaborative tasks with the robot in terms of the provided functionalities?

   Compile a list of all required functionalities that your human-robot interaction scenario should support. Iterate this list during your user-centered design approach, to close the gap between evaluation and implementation.

6.5.2 Guiding Research Questions: Social Acceptance

1. Performance Expectancy: To which degree do the users believe that the collaboration with the robot improves their work?

   Bear in mind that performance expectancy is a key aspect of social acceptance in collaborative interaction scenarios. Try to assess peoples' expectations with qualitative approaches, such as focus groups or interviews. Moreover, try to assess the degree of performance expectancy after the interaction with a robotic systems, e.g. with a questionnaire.

2. Effort Expectancy: To which degree does the human believe that the collaboration with the robot is free of extra effort?

   Effort expectancy is a similar important indicator for social acceptance as performance expectancy. It should be assessed at early stages of the development process by means of qualitative methods, as well as after a user study, e.g. with a questionnaire.

3. Attitude toward Using Technology: Will the general attitude towards technology have an influence on the collaboration with the robot?

   Try to assess the attitude change of users before and after interacting with a robotic system, e.g by means of the NARS questionnaire (SubSection 3.7.3). Take into account that age has an influence on the attitude. Older participants often show more positive attitude ratings.

4. Perceived Competence: To which degree does the human perceive him/herself as competent to collaborate with the robot?

   Characterize your user profile as detailed as possible for your human-robot interaction scenario. Try to stick as close as possible to your user profile when you are recruiting participants for a user study. Try to assess the perceived competence with qualitative methods as well as with questionnaires, taking into account, that this indicator summarizes aspects of self efficacy and facilitating conditions:

   - Self Efficacy: Will the self efficacy have an influence on the collaboration with the robot?
   - Facilitating Conditions: Will the facilitating conditions have an influence on the collaboration with the robot?

5. Relationship Expectancy: To which degree does the human expect to establish a relationship with the robot?

   Take into account that relationship expectancy is less important for collaborative scenarios, than e.g. for a assistive scenario. However, try to assess by means of qualitative and quantitative approaches in how far users can imagine to establish a relationship with the robot. This could give you insights about the sustainability of the collaboration. Bear in mind that relationship expectancy is composed of the indicators forms of grouping, attachment, and reciprocity:

- Forms of Grouping: To which degree does the human feel to carry out a task in a team with the robot?
- Attachment: Does the robot support a feeling of attachment?
- Reciprocity: Does an experience of reciprocity between the human and the robot evolve?

6.5.3 Guiding Research Questions: User Experience

1. Embodiment: Is a “biologically inspired” designed robot preferred for the task in comparison to a “functionally designed”one?

   Independently of the embodiment of robot in your interaction scenario, try to identify, if the chosen embodiment is perceived as ideal for the scenario by you potential users. This can be done by creative techniques, such as participatory design, but also with standardized questionnaires.

2. Emotion: Which emotions does the robot evoke during the interaction with the human?

   Emotion is a key aspect of user experience. Try to take this indicator into account by a variety of methods (qualitative and quantitative), like interviews, questionnaires, and as well physiological measurements. Try to identify which emotions (positive or negative) the robot evokes and in how far they influence other evaluation indicators.

3. Human-Oriented Perception: Does an interaction based on paradigms of Human-Human Interaction (HHI) support the collaboration?

   Perform comparative user studies (e.g. Wizard of Oz studies) in which you analyze the impact of different HHI cues (e.g. gaze, gestures, and turn taking) on the interaction. Moreover, assess the degree of human-oriented perception after a user study, to review your implementation of the cues.

4. Feeling of Security: Does the human feel secure during the collaboration with the robot?

   Perform field trials with autonomous robots to achieve the most reliable data on feeling of security. Questionnaires, as well as physiological measures, can be used to evaluate this indicator. Moreover, bear in mind that feeling of security is closely related to feeling of control, an indicator which can be reasonably assessed by qualitative methods, such as interviews after a user study, as well as with standardized questionnaires.

5. Co-Experience: What kind of co-experience arises in collaboration with robots?

   Perform interviews and surveys after user studies to explore, if a feeling of co-experience evolves in the interaction with the robot. Try to identify cues that foster the feeling of co-experience and implement them into your system to increase the sustainability of human-robot interaction in collaborative scenarios on the long term.

6.5.4 Guiding Research Questions: Societal Impact

1. Quality of Life, Health, and Security: How do human-robot interaction scenarios have to be prepared in order to not influence quality of life, health, and security of citizens in a negative way?

   Determine human-robot interaction scenarios which positively influence the quality of life of the targeted user group (e.g. by means of the world cafè approach Section 6.4). Therefore, take into account the knowledge about peoples' imaginaries about the "future robotic society" and cultural specific requirements. Assess the users' perception of quality of life by means of qualitative methods, such as interviews (e.g. after a user study) or with questionnaires (in a robot related context).

2. Working Conditions and Employment: How does the collaboration with robots impact working conditions?

    Try to identify relevant areas in working life which could change, due to your human-robot interaction scenario. Discuss with experts and users the possible impacts robots could have on these identified areas. Derive implications for the design of your interaction scenario to limit negative effects.

3. Education: How do new types of education have to be designed to cover the new requirements in the working life?

    Try to identify required training phases in your human-robot interaction scenario and integrate them into your experimental setting of a user study. Discuss with participants how they could imagine a future education system with regard to robotics and integrate the derived knowledge in your system design, with respect to the foreknowledge of your users.

4. Cultural Context: Does the collaboration with robots have an unequal impact on different cultures?

    Try to conduct comparative studies (surveys, user studies, focus groups) to determine cultural differences in the interaction with robots. Take into account research guidelines of cross-cultural studies to ensure the comparability of your gathered data.

These guiding research questions should assist other researchers in determining the relevant aspects when applying the USUS evaluation framework. However, for a successful application of the framework it is also relevant to take into account the phases of the user-centered design cycle, which will be described in the following Section 6.6.

6.6 Integration into the UCD Cycle

User-centered design, as described in detail in Chapter 2, is an iterative design process which should give guidance throughout the entire design life cycle of an interactive system. Some of the major points user-centered design focuses on are:

1. Involving the potential end-users throughout the design life cycle to ensure their needs are adequately addressed.

2. Understanding the context in which the interactive system will be used.

3. Understanding the capabilities of both the interactive system and the user.

While this standard provides a general framework for interactive robot design, it does not provide practical guidelines how to apply it. Additional to the general user-centered design framework with its focuses, the integration of theUSUS evaluation framework can be included and employed iteratively throughout the user-centered design cycle. The focuses of the USUS evaluation framework in the user-centered design cycle are:

1. Assess multiple factors (operationalized into indicators) throughout the design cycle.

2. Combine qualitative and quantitative methods (self-reporting and in situ; user-based and expert-based).

3. Evaluate the attitude change of the user throughout the design cycle.

In HCI, it has become more and more popular in the last years to evaluate factors of technology usage going beyond usability, such as user experience and user acceptance. Several studies in HRI also have been conducted following a user-centered design (see SubSection 2.2.1). However, the core element of the USUS evaluation framework is to integrate human responses to Human-Robot Collaboration scenarios as early as possible into the design cycle by means of adequate method usage. This is important as it is hard to change robotic physical hardware and input/output modalities in a late phase of the design cycle. Furthermore, the design of Human-Robot Collaboration scenarios with humanoid robots is until now an ongoing process, meaning that the stage of a formative evaluation is only achieved in the latest cases, thus, user-centered design is even more prominent.

The application of the USUS framework as demonstrated by means of the case studies in Chapter 4, demonstrated that it is possible to integrate users by means of qualitative approaches. Integrating focus groups or even quantitative prototypical mixed-reality experiment in an early development stage can thereby address more evaluation goals than only usability problems.

During the application of the USUS evaluation framework, if it becomes apparent that some evaluation indicators or even complete factors are negatively assessed in a human-robot interaction study, it is important to figure out why and to recommend solutions that do not negatively impact other indicators/factors. A core element of the application of the USUS framework is the attempt of the interpretive understanding of human-robot interaction in order to thereby arrive at a causal explanation of its course and effects (in accordance to Max Weber's definition of Sociology). For example, in one of the case studies, one participant stated that she would never want to have a robot in a teaching role and mentioned an above average number of reasons for that. In the interpretation of the data, it could be seen that the participants was also afraid of robots increasing the unemployment rate; as having the profession of a primary school teacher. In another case study, one participant needed an extremely high task duration to solve a task successfully. In the final interview, it was explained that the participant was used to computer games with force feedback and expects a robot (which is a kind of technology that is much more sophisticated as a computer game) to have force feedback, even though nobody asserted that the robot has it.

The general application of the USUS evaluation framework within a user-centered design cycle could look like the following (see Figure 6.1):

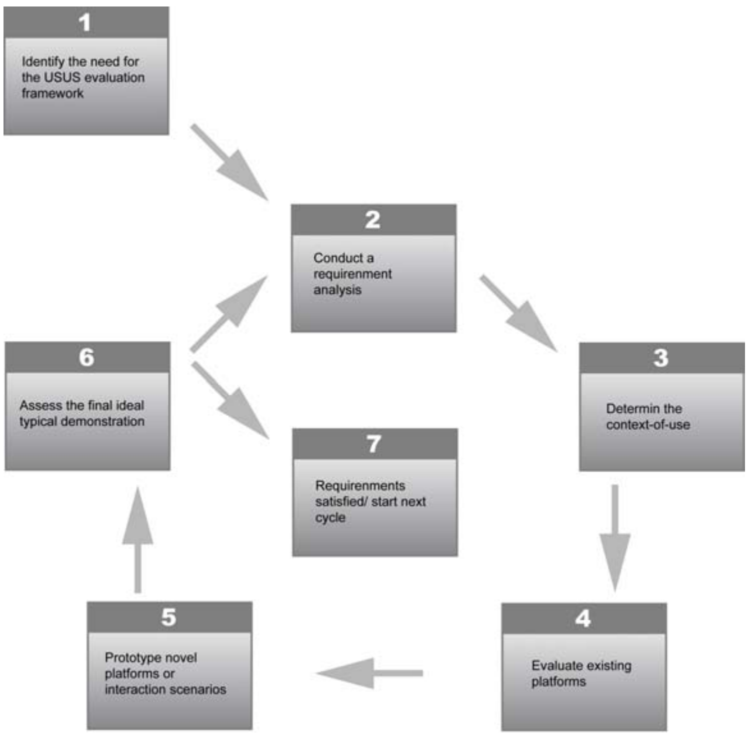


Figure 6.1: Application of the USUS Evaluation Framework

1. The need of the USUS evaluation framework needs to be identified. Is it the project's goal to go beyond classical performance measures like

precision, robustness, reliability etc. and include aspects like user satisfaction or societal impact? Are there any scenarios envisioned where robots enter human's everyday life and humans and robots have direct contact interaction?

2. The first requirement analysis should be preferentially conducted with qualitative methods like interviews and focus groups investigating the soft factors of the USUS framework (user experience, social acceptance and societal impact) to better understand the target group. The results of this phase should be taken as input for phase 4.

3. If a need for the USUS evaluation framework is identified, the context-of-use of robotic technology has to be identified. Is it a sensitive private setting or is it a public work setting? How many users and robots are involved in the interaction scenarios? Which type of robot is investigated, e.g. humanoid, zoopomorohic, or even robotic body extensions? In accordance to the definition of the context-of-use, the theoretical (and subsequently probably the methodological framework needs to be revised). For instance, indicators like human-oriented perception or performance expectancy are not ideal when investigating child-robot interaction with zoopomorphic robots. In this phase of the life cycle, the target group also has to be defined. Are novice users, elderly, children or even tele-surgery experts the main point of interest?

4. The fourth phase is based on the interaction scenarios defined in the third phase. Currently, existing robotic platforms can be used and evaluated in lab-based settings which are experimentally design in accordance to the scenarios. Quantitative instruments like questionnaires and physiological measurements can be used to assess a first baseline for the theoretical indicators.

5. In the fifth phase, novel platforms or interaction paradigms can be prototyped in accordance to the findings from the evaluation studies conducted in phase 4. As result of this a phase, an ideal typical demonstration for the defined interaction scenarios should be developed.

6. In the last phase, the final demonstration should be assessed and benchmarked against the evaluation results gained in the pre-studies.

7. If the requirements are satisfied, the technical development will continue to implement the robotic systems in everyday life situations. However, if further huge technological changes influence this procedure a next USUS-guided user-centered design cycle can be initiated.

Chapter 7

Discussion and Outlook

7.1 Discussion of the Evaluation Framework

Humanoid robots encourage social interaction and stimulate a unique interaction experience which is different from "normal" interactive technology. The field of social science, HCI, and also HRI provide several validated evaluation methods to assess several aspects of the interaction with robotic systems. Little work has been done so far to investigate methodological limitations and to provide an evaluation framework that targets the holistic and contextual nature of human-robot interaction. The overall goal of this thesis was to investigate evaluation approaches from a human-centered HRI view point and to offer a more holistic evaluation strategy for this research area. In the following sections it will be summarized in how far the research objectives were met in this PhD thesis. The following four research objectives were derived:

1. Proposing a theoretical and methodological framework for evaluating usability, social acceptance, user experience, and societal impact for working scenarios with humanoid robots.

2. Assessing the feasibility of the proposed evaluation framework by means of empirical case studies.

3. Developing a reference model for USUS evaluation framework to reveal interrelations between the factors and indicators of the evaluation framework.

4. Providing guidelines for the application of the proposed framework to support other evaluators.

7.1.1 Objective 1: Evaluation Framework

The aim to set up a unified evaluation framework for human-robot interaction with humanoid robots in a collaborative work scenario was achieved by proposing the USUS evaluation framework (see Chapter 3 and Weiss et al (2009a)). The main goal of the USUS evaluation framework was to guide Human-Robot Interaction research on how people experience robots as a support for collaborative work and accept them as part of society. Thus, the framework should go beyond traditional usability issues and open the view towards bigger concepts, like user experience, acceptance and impact factors.

The framework was set up as a multi-level factor/indicator evaluation model consisting of a theoretical framework covering four main factors: usability, social acceptance, user experience, and societal impact. In addition a methodological framework proposing a variety of methods to address these factors. In the first proposal, the four evaluation factors, the USUS factors, were split into several indicators to operationalize them, based on a intense literature review taking into account not only HRI literature, but also HCI literature, psychological and sociological literature. The combination of the HCI literature and sociological literature, for instance, enabled the development of the social acceptance indicators, including aspects like attachment and reciprocity. The inclusion of post-humanism literature allowed a structured definition of relevant key aspects to investigate imaginations of the impact of robotic technology for the future society.

The definition and description of the factors and indicators set the focus on HRI specific evaluation problems, e.g. user studying human-robot interaction with novice users. Additional examples involve, investigating social acceptance with different cultural backgrounds and investigating users' experiences with differently embodied robots. However, the first proposal of the theoretical USUS evaluation framework was based on generating reasonable indicators without taking into account their interrelations. A strict separation between the indicators (and even the factors) was never assumed, but it was a relevant precondition to set a baseline for an unbiased exploration of these interrelations.

The proposal for the methodological framework derived from methods which are already used in HRI (see Chapter 2) but also from methods neighborhood disciplines like HCI, psychology, and sociology. Regarding usage of evaluation methods, the USUS evaluation framework proposes a method mix, taking into account the advantages and limitations of evaluation techniques. The use of multiple methods helps to establish reliable results, in particular if methodological adaptations are made to meet the specific requirements of a study setting.

As Chapter 2 showed, developing frameworks for HRI research is nothing new, but the central point of this PhD thesis was to combine the theoretical and methodological state-of-the-art perspectives in HRI. The theoretical framework itself does not address only unique human-robot interaction evaluation factors. However, the methods and instruments proposed in the methodological framework assure targeting the indicators from a HRI-centric perspective. The approach of the USUS evaluation framework closes the gap in HRI research that human-robot interaction needs to be assessed in holistic view at all stages of the development process of robotics systems by researchers and engineers coming from various disciplines. Thereby, the theoretical framework offers the theory-guidance which is relevant to avoid evaluation studies "out of the blue" and the metrics and methods to generate reliable data and reproducible study settings.

7.1.2 Objective 2: Feasibility of the Framework

The feasibility of the USUS evaluation framework could partly be proved by eleven case studies (see Chapter 4) and a broad online survey (see Chapter 5). These studies took place during the three year project duration of the EU-funded FP6 project "Robot@cwe: Advanced robotic systems in future collaborative working environments" and gave insights on the theoretical as well as on the methodological framework. Each case study can be seen as a building block regarding the assessment of the suitability of the methods, initially proposed in the USUS evaluation framework. Lessons could be learned in all of the studies, in terms of required methodological adaptations for human-robot interaction studies. More details for each case study are given in the following sections.

The first case study (see Section 4.2 and Weiss et al (2009h)) used a qualitative focus group approach to investigate user experience and social acceptance factors. Only videos were used as discussion stimulus and a "Robot Building Set" was used to investigate the indicator embodiment.

The methodological adaptation used in this case study could show the high value of qualitative approaches to gain a better understanding of people's expectations, hopes, and fears regarding robots as working colleagues in which a "real" robot is not even used. Similarly, the creative "Robot Building Set" approach demonstrated a novel and easy way of exploring the desired embodiment for a human-robot interaction scenario. Thus, it could be proved that the USUS evaluation approach can be integrated into the very beginning of the user-centered design cycle of robotic systems. However, it still needs to be investigate in how far the type of robot used in the video stimulus and the sociocultural context of the participants influences the results of the focus group approach.

The second case study (see Section 4.3 and Weiss et al (2009f)) was based on an expert evaluation methods adapted from HCI, the heuristic evaluation. The main adaptation of the method laid in the use of video scenarios for remote evaluation guided by an evaluation manual and that the evaluators were not usability experts, but interdisciplinary HRI experts (engineers, psychologists, computer scientists etc.).

It could be shown that this expert evaluation approach was valuable to identify severe usability problems in human-robot interaction scenarios. This approved the suitability of the USUS evaluation framework for early stages of the user-centered design cycle. Moreover, video-based evaluation approaches can be more economical than in situ studies, as they can be conducted remotely. The evaluation manual turned out to be the critical aspect of this methodological approach. However, further development of such evaluation manuals will decrease the amount of study results with limited reliability in future HRI studies.

The third case study (see Section 4.4 and Weiss et al (2009c)) was the first laboratory-based user studies in which usability measures like task completion and task duration were assessed in combination with a "Retrospective Think Aloud", questionnaires, and interviews.

Even with a small sample size of only four participants, a tendency could be seen for the feasibility of the multiple method approach. Particularly, the linking of the questionnaire data with the interview data facilitated the clustering and interpretation of the data. The suitability of the AttrakDiff and SUS questionnaires to assess human-robot interaction scenarios could also be proved in this study. Two major critical aspects for human-robot interaction studies could be identified in this case study: interaction only based on speech commands was not experienced as collaboration for the participants and feeling of security can only be partially assessed in laboratory-based studies, as participants expect their safety as granted.

The fourth case study (see Section 4.5 and Weiss et al (2009e)) used the same user study design as the previous case study but for a completely different interaction scenario (learning by demonstration) which was the first indication of the flexibility of the USUS evaluation framework.

Due to the bigger sample size of twelve participants, more usability problems could be identified and a wider responses spectrum in the interview data could be achieved. However, this study revealed that training phases could improve the study setting of human-robot interaction studies and should be integrated into the user study methodology.

The fifth case study (see Section 4.6 and Weiss et al (2009f)) again used the same user study design but for a human-robot interaction scenario in which the robot was remote controlled via a computer interface and guided through a maze (representing a moon environment). No direct contact interaction happened between the user and the robot, but the human experienced the navigation on the moon though the “robot’s eyes” which was considered as close collaboration by the participants. In this case study, a training phase was integrated in the study set up.

In general, this case study gave the second indication for the flexibility and feasibility of the USUS evaluation framework as it could be applied successfully in a human-robot interaction study without direct contact interaction and voice control. Moreover, this case study gave valuable insights on the importance of the simulation of context factors for user studies and that the tasks need to be demanding enough to generate representative and comparable results. An additional lesson learned from this case study was the relevance of time-aligned analysis (in which the video material of the robot’s movements, the participants behavior and the clicks on the interface can be synchronized) to identify the reasons for usability problems.

The sixth case study (see Section 4.7 and Weiss et al (2009b)) also used the same user study design for a mixed-reality prototype scenario in which the robot was only simulated and displayed on a projection wall. In this study, it was explored how different interaction techniques influence the perception of the robot. The interface between the real and the simulated world was a board with which the user could interact with the robot and different feedback modalities were implemented (visual, auditive and haptic). The reaction of the robot was completely controlled by a human wizard. Nevertheless, the interaction scenario could be reasonably assessed with the USUS evaluation framework.

However, the study revealed the relevance of a high-fidelity of the prototype and the simulation of context factors (e.g. ambient construction site noise) to allow a reasonable application of the USUS evaluation framework. The findings of an evaluation study like this one might not be generalizable/identical for the same scenario with the actual robot, but they can be a valuable indication e.g. which feedback modalities to focus, if the goal is to improve user experience aspects.

The seventh case study (see Section 4.8 and Weiss et al (2009g)) was the largest laboratory-based user study presented in this PhD thesis. In this study, a collaboration scenario in which the human had to pick-up, carry, and put down a table together with the HRP-2 robot was investigated. Half of the study participants had a Western cultural background and the other half, an Asian cultural background. All evaluation indicators were addressed during the study with eleven different instruments.

The case study proved the high value of the theoretical guidance of the USUS evaluation framework as research questions on differences in experience and acceptance due to cultural background could be derived. Moreover, the feasibility of the methodological framework could be demonstrated in terms of the combination of questionnaires with physiological data to overcome the issue of socially desired answers, as participants hardly can manipulate their nervous system. The arousal feature calculated based on heart rate data made visible which moments during the interaction caused an arousal (which was similar for almost all participants independent of their cultural background), whereas the questionnaire revealed a significantly higher ratings regarding emotions for Western participants (which could be due to the understatement response behavior of Asian participants).

The eighth and ninth case study (see Section 4.9 and Weiss et al (2008, 2010)) was conducted as a field trial with the ACE robot with the aim to investigate social acceptance and societal impact in a natural setting. In these studies, the ACE robot was integrated as social actor in a public place without an warning for the pedestrians. Both studies were conducted as sociological breaching experiments in combination with a street survey which was introduced as novel research approach in the field of Human-Robot Interaction.

The case studies showed the suitability of the USUS evaluation framework for human-robot interaction evaluations "in the wild". It closed the methodological gap of how to assess social acceptance indicators which can hardly be observed in the laboratory, as forms of grouping and perceived competence. However, also the limitations of this type of field trial became obvious, such as participant sampling and socially desired response behavior in the street survey.

The tenth case study (see Section 4.10 and Weiss et al (2009g)) was the second proof that expert evaluation methods from HCI, this time the cognitive walkthrough approach, can be reasonably adapted for the assessment of human-robot interaction. Again, video-based scenarios derived from case study 3 (Section 4.4) and case study 4 (Section 4.5) were used as evaluation basis. A group of expert evaluators walked through these video scenarios to identify usability problems with a special focus on learnability issues.

The advantages of repeating single video sequences as often as required by the expert evaluators to identify the reasons for usability problems supported the process of proposing design alternatives. Providing the evaluators with scenario cards helped them to be aware of the evaluation context. However, this case study showed that it is valuable to include one expert into the evaluation team who was also part of the experiment from which the video-based scenario was derived. Regarding the USUS evaluation framework, the case study demonstrated how expert evaluation methods can also be reasonably used in later stages of the user-centered design cycle. The cognitive walkthrough approach can be adapted in a way to analyze learnability in HRI and that the theory-guided focus on a single evaluation indicator can close explanatory gaps.

The eleventh case study (see Section 4.11 and Weiss et al (2009d)) was a row of in-depth interviews (taking into account the Delphi approach) conducted with robotic experts from industry with the aim to gain more insights about the indicators of societal impact. This case study showed again the importance of qualitative methods in the USUS evaluation framework to guarantee in-depth data about "societal phenomenas" besides pure performance data of robotic systems. Having the opinion from experts from very diverse fields, opened the view for "future robotic society" imaginations.

Moreover, the expert opinions proved to be complementary to the participants' assumptions in the user studies and were not redundant. The case study revealed the tendency that a Delphi approach could generate a very clear picture of the future robotic society. However, methodological limitations became obvious, e.g. generating a common understanding of the term future or bridging knowledge gaps from one industry expert to the other. Thus, another method was proposed for the iteration of the methodological USUS evaluation framework: the world cafè approach (see Chapter 6).

In general, the case studies showed that the first proposed USUS evaluation framework was suitable for evaluation studies at all levels of the user-centered design cycle, as it combines expert evaluation approaches with user-based evaluation studies. Furthermore, it combines qualitative approaches (focus groups, in-depth interviews, and expert evaluation) as well as quantitative approaches (task duration and completion rate, pre-structured questionnaires, and physiological measurements). The case studies proved that the context dependency of the framework is rather low but the adaptation of the methods is a key element for successful evaluation studies. The framework can be used for user studies with autonomous systems and also Wizard of Oz studies in both lab- and field-based settings.

All eleven case studies revealed advantages, limitations, and lessons learned for future HRI study settings. However, the feasibility of the USUS evaluation framework could only be proved by means of the case studies (see also Section 4.12). They could show that the originally proposed methods (if adapted reasonably) can be used to evaluate the respective indicators, but they could not prove if these methods are best to achieve the most accurate results. Future comparative studies will be necessary to approve this. Moreover, it cannot be guaranteed that the method mix, which was successful for the interaction scenarios presented in this PhD thesis, can work for all kind of interaction scenarios, e.g. children interacting with zoopomorphic robots in a therapeutic context. Another limitation in the proof of feasibility is that the indicators of user experience and social acceptance were mainly assessed by means of questionnaires. Other methods should be considered and tested for these factors in future.

The main issue remaining after the empirical validation of the USUS evaluation framework is that the completeness neither of the theoretical nor of the methodological framework can be guaranteed in terms of factors, indicators and methods. However, the case studies showed that human-robot interaction studies can be challenging regarding theoretical grounding and methodological set up, but also very rewarding, if they are planed in detail and in accordance to a theoretical basis and lessons learned from previous studies.

7.1.3 Objective 3: Reference Model

The interrelations between the evaluation factors and indicators could be revealed only for the user experience and social acceptance indicators by means of a cross-analysis of the quantitative data gathered in a broad online survey with the UX- and SoAc-Questionnaire (Section 5.2).

The online survey provided a huge sample size (n=398) which allowed the prove of the questionnaire scales (and thus the prove of the factor/indicator model) in terms of reliability and validity by factor analysis and Cronbach's alpha. Furthermore, the interrelations between the indicators were analyzed by means of correlation coefficients (Spearman's roh) and the influence of six independent variables "nationality" (expressions: Asian/Western), "gender", "age", "profession" (expressions: worked in HRI/not worked in HRI), "experiment" (expressions: already participated in HRI experiments/never participated in HRI experiments), and "robot" (expressions: HRP-2 robot/HOAP-3 robot) was investigated by non parametric Mann Whitney U-tests (see also Weiss et al (2009g)).

The results of the data analysis revealed that the initially assumed granularity was too detailed for the social acceptance indicators. Thus, the number of social acceptance indicators was reduced from eight to five indicators: performance expectancy, effort expectancy, relationship expectancy, perceived competency, and attitude towards robots. The new indicator perceived competency comprises self efficacy and facilitating conditions; the indicator relationship expectancy comprises attachment, feeling of reciprocity ad forms of grouping.

The indicators for user experience did not change after the factor analysis, however the indicator emotion could only partially be addressed in the online survey and the indicator co-experience was not addressed at all. Furthermore, the influence of six independent variables revealed the high impact of independent variables on both evaluation factors and in particular on the user experience indicators. The correlation analysis between all user experience and social acceptance indicators demonstrated the strong interplay between the indicators (52 of 72 possible correlations were highly significant).

As the eleven case studies could only validate the feasibility of the proposed methodological framework addressing the respective evaluation indicators, the broad online survey was required to prove the assumed interplay between them. Due to he successful validation of the UX- and SoAc-Questionnaires survey, the HRI community can now be provided a reliable tool to assess a humanoid robots regarding their effects of user experience and social acceptance. Even if another researcher wants to use the reference model plus the questionnaires for a study only focusing on some of the indicators, it can be useful. However, in that case a check of validity and reliability is recommended (see Section 5.4).

The main main advantage of the revised model is its conjunction of state-of-the-art literature with an empirical proof-of-concept study which enables other HRI researchers to conduct theoretically-guided evaluation studies with validated measures. Thus, the reference model supports repetitive testing and simple investigation of HRI evaluation factors. Due to the investigation of the effect of the independent variables, the revised USUS evaluation framework can support the participant recruitment procedure and data interpretation, e.g. if only elderly participants are recruited it has to be taken into account that elders tend to have a more positive attitude towards humanoid robots than younger ones.

However, two limitations have to be considered regarding the explanatory power of the revised model. Guidelines of cross cultural surveys were only considered in the questionnaire translation process, but not in the choice of items and the presentation of the questionnaire. Participants of the online survey were (semi) experts in the field (due to the recruitment via robotics-related email lists), but were not interviewed in their role as experts, but as "normal survey participant". Thus, future work will be required to investigate if these confounding factors influence the validity of the reference model.

Up to now, only the interrelations of the user experience (excluding the indicator co-experience) and social acceptance indicators were validated. There is still the need to validate the interrelations within the factors usability and social acceptance and their relation towards the other evaluation factors. The current state of the revised USUS evaluation framework is presented (see figure in Figure 7.1).

7.1.4 Objective 4: Application Guidelines

The fourth aim of the USUS evaluation framework was to provide guidelines that support other researchers in answering research questions on how people experience robots in collaborative scenarios and if they accept them as part of society. This aim could be achieved by posing central research questions which can be answered by applying the revised USUS evaluation framework. The application of the USUS evaluation framework and its integration into the user-centered design cycle was also demonstrated through the case studies.

Chapter 6 gives an overview on the biggest challenges in applying the USUS evaluation framework which were encountered during the execution of the case studies and the online survey, namely: (1) Choice of system functionalities to investigate, (2) comparability of different systems, and (3) observation of "natural" human-robot interaction in laboratory-based settings, participant sampling, and the language barrier in cross cultural studies.

Following, the main methodological considerations which should be taken into account when applying the USUS evaluation framework were summarized: (1) simulation of context and environment factors (e.g. ambient noise), (2) adaptation of the theoretical framework in accordance to the investigated domain, (3) participant sampling (in particular recruitment characteristics), (4) tools and equipment (including advices for data analysis tools), (5) experimental design (taking into account the need of a method mix in HRI,(6) use of the UX- and SoAc-Questionnaire, and (7) data analysis taking into account possible inter factor/indicator synergies.

Based on the lessons learned from the eleven case studies, a revised methodological framework was proposed including four new methodological approaches:

(1) participatory design, (2) world cafè, (3) SInA data analysis, and (4) a standardized questionnaire for societal impact. These newly proposed methods still need empirical validation, but enrich the researcher's "tool box" for the method mix. Finally, the proposed central research questions can guide current activities within a research project and support other researchers on a more general level how to investigate the USUS factors in human-robot interaction.

The integration of the USUS evaluation framework into the user-centered design cycle focuses on (1) the assessment of multiple evaluation factors (operationalized into several indicators) throughout the design cycle, (2) the combination of quantitative and qualitative methods, and (3) the users' attitude change towards the investigated robotic system during the design cycle. Thereby, the user-centered design cycle consists of seven (instead of the original six stages): (1) identify the need for the USUS evaluation framework, (2) conduct a requirement analysis, (3) determine the context of use, (4) evaluate existing platforms, (5) prototype novel platforms or interaction scenarios, (6) assess the final ideal typical demonstration, and (7) requirements satisfied or start the next cycle.

The motivation behind the development of application guidelines for the USUS evaluation framework was to give researchers in HRI irrespectively of the disciplinary background a "reference", "manual", or "handbook" for ensuring that interaction with their robots is of high quality in terms of usability, social acceptance, user experience, and societal impact. This will hopefully help ensure the sustainable integration of robotic systems in everyday life.

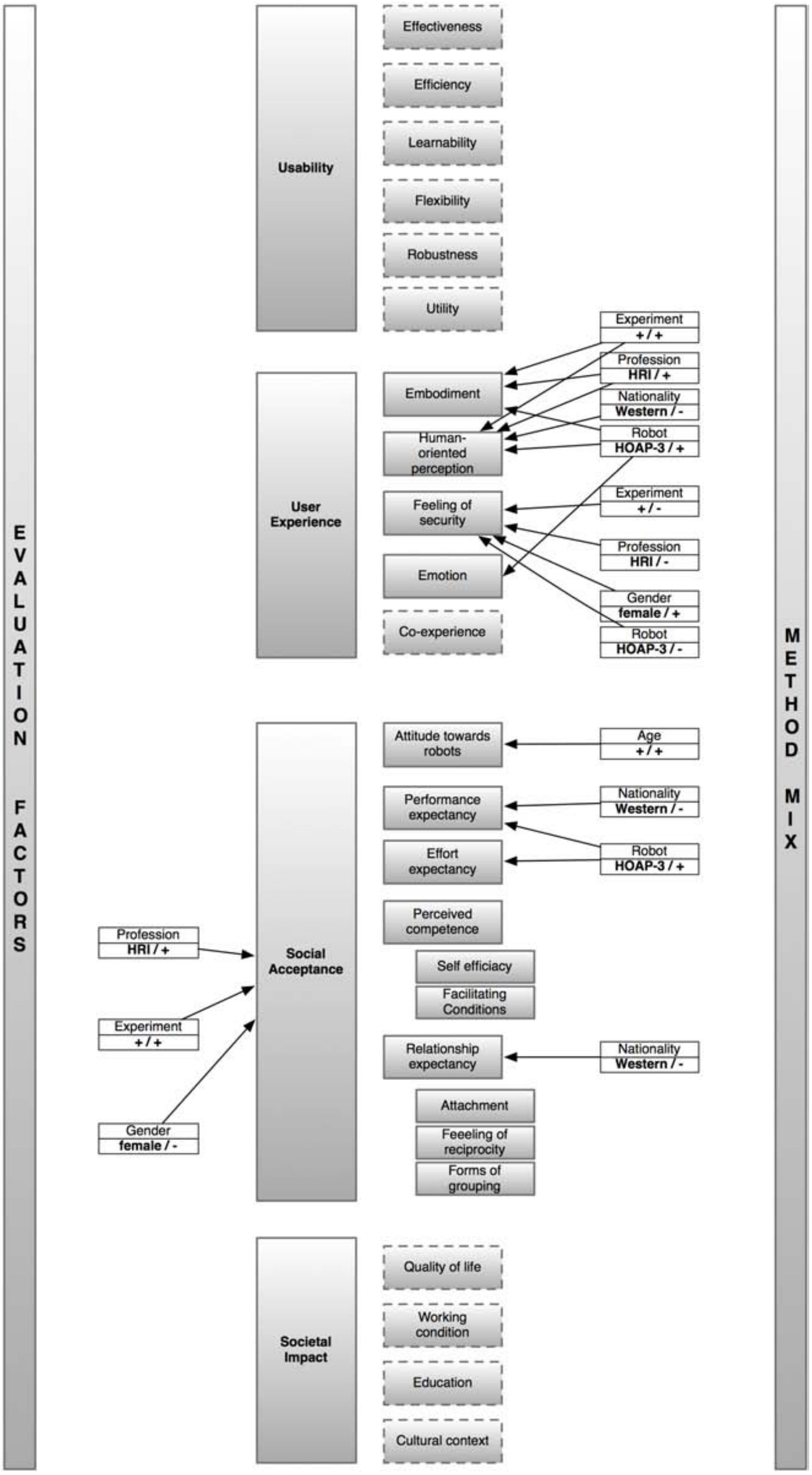
7.1.5 Reflection on the Outcome

While the fields of psychology, sociology, HCI, and HRI provide many well tested evaluation techniques (see Chapter 2), I was concerned with how these should be applied to HRI in a way that acknowledges and targets a holistic interaction context going beyond usability aspects.

The work in this PhD thesis proposed an evaluation framework not only addressing usability, but also social acceptance, user experience, and societal impact - the USUS evaluation framework. It was originally based on the superficial level of operationalizing evaluation factors by means of several indicators derived purely from literature review and could be partly validated by means of empirical studies. By assessing the indicators with a broad method mix in several case studies and cross analyzing the quantitative data, the evaluation framework could be revised and improved.

I am convinced that this framework can help to strengthen the small methodological, theoretical, and conceptual basis of the currently evolving research field of HRI (Dautenhahn (2007a)). Due to the numerous case stud-

Figure 7.1: Revised USUS Evaluation Framework



ies presented in detail in this PhD thesis, results are reproducible by other research groups. This will help the research filed to make progress in creating a foundation of knowledge and insights others can build upon.

7.2 Contributions of the Thesis

As mentioned in the introduction this PhD thesis aimed to contribute to three different research areas: (1) Human-Robot Interaction, (2) Methods of empirical social research, and (3) Sociology of Technology. In the following sections it will be reflected in how far these aims have been met.

7.2.1 Human-Robot Interaction

Regarding the contribution to HRI research, this PhD thesis can be seen as "reference", "manual" or "handbook" for robotic engineers (developers and researchers) to consider their human-robot interaction problems. If engineers know the way of evaluation, as a nature, they will try to develop products with high evaluation score and thus increase usable, positively experienced and accepted human-robot collaboration.

7.2.2 Methods of Empirical Social Research

Regarding the contribution to methods of empirical social research this thesis presented a variety of methodological adaptations and study settings which can serve as impulses for other researchers in this area. Moreover, methodological considerations and challenges are shown which will assist other researchers with the planning, design, and executing of user-centered studies.

7.2.3 Sociology of Technology

Regarding the contribution to the research field of sociology, it has to be said that predictive information about robots and the future society is extremely hard to gain. For example, 20 years ago no one dreamed of cell phones (a phone you can put in your pocket, isn't that crazy...!), but nowadays they are everywhere. However, gathering information about novice users' estimation of the future in comparison with the estimations of the industry, gave insights about people's attitudes, fears, and hopes. These thinkings can be addressed in future when introducing robots into the mass market. It could be shown that sociological theory gave valuable insights on how people in general interact with objects which could be adapted for HRI research. Furthermore, this thesis could demonstrate the importance of combining sociological thinking with technological development to drive robot design.

7.3 Outlook

This PhD thesis can be conceived of as series of case studies investigating several research questions on how users perceive the usability, social acceptance, user experience, and societal impact of humanoid robots in collaborative scenarios. The execution of these case studies was guided by the USUS evaluation framework, which offered a theory-grounding and a method proposal for the case studies. The conduction of the case studies gave on one hand valuable insights on usability, social acceptance, user experience, and societal impact of human-robot interaction in collaborative work settings with humanoid robots. On the other hand, also areas could be identified where future research is required:

Usability Task duration varies remarkably between participants with Asian and Western cultural background due to different problem solving strategies. A novel methodological approach I consider for user studies are "cultural interaction scenarios". These scenarios take into account the knowledge about cultural differences and try to shape interaction scenarios with respect to it. Another aspect of usability is the lack of flexibility in current human-robot interaction scenarios. To gain more insights about the influence of this evaluation indicator on the users' performance, but also on user experience and social acceptance, comparative Wizard of Oz studies with varying degrees of flexibility should be conducted. The last aspect in terms of usability, which is relevant for future research, is the area of user support. Currently human-robot interaction scenarios are lacking of options that help the user to recover on his/her own from minor interaction errors. Research in this direction will help engineers to develop systems that are more robust in scenarios with a human in the loop and identify measures for the usability indicator robustness in HRI.

Social Acceptance The results regarding social acceptability are in general good news for robotic engineers, as the acceptance ratings were higher in the user studies after participants interacted with the robots. Moreover, this finding supports the user-centered design approach of the USUS evaluation framework, as a mutual shaping between social requirements and technological possibilities will increase the acceptance of robotic technology. Qualitative approaches like the world cafè could help to better understand in which cases relationship expectancy and perceived competency are relevant for successful integration of robots into everyday life. For this newly proposed method in the USUS evaluation framework (see Chapter 6) as well as for interview techniques; reasonable, methodological adaptations need to be developed and the feasibility of the methods to assess social acceptance needs to approved.

User Experience The results of the user experience assessments were interesting regarding the physiological data measuring in the HRC user study with the HRP-2 robot (see Section 4.8). The arousal level varied a lot during the interaction. Future work will be needed to learn how to interpret the data correctly (are we measuring emotion or feeling of security?). In a next step, the data could be used as input for the robot to react adequately towards human behavior. This would help to increase the positive user experience in the interaction. Moreover, user studies should be conducted to further investigate the correlation between feeling of security and feeling of control (which could be a relevant for a next iteration of the theoretical USUS framework). If the assumed positive correlation between these user experience indicators can be verified, user experience design guidelines for feeling of control could improve the feeling of security for a wide range of robotic platforms. As a last point, future work in the area of user experience should focus on the integration of participatory design approaches in HRI, interviews, and the world cafè approach to inform the design of human-robot interaction studies at an early stage of the user-centered design cycle. Here again, reasonable methodological adaptations and their proof of feasibility are required to support the HRI research community.

Societal Impact The main finding regarding societal impact was that for the working contexts, people prefer robots which are sophisticated tools and not colleagues. This finding underlines the importance of societal impact studies to get rid of "classical assumptions" regarding human-robot interaction like "the more emotional intelligent the better". There is a need to more thoroughly investigate the imaginations of potential users and experts to understand the difference between usage contexts of robots (work vs. home vs. therapy etc.) and to focus on ethnographic approaches to better understand these contexts. Moreover, more cross-cultural (taking into account cross-cultural study design, see http://ccsg.isr.umich.edu/-instrdev.cfm) studies with standardized measures (e.g. societal impact questionnaires) are needed to characterize cultural user groups beyond Western vs. Asian. Little knowledge exists so far about cultural differences regarding human-robot interaction between European countries or Eastern countries. However, technology does not appear by chance. Robotic systems are designed in accordance to what they are supposed to do. The more detailed "cultural interaction scenarios" can be defined, the better the robotics community can design for their specific requirements.

Regarding the USUS evaluation framework future research will be required to conduct a holistic validation of the evaluation framework in terms of completeness of the theoretical indicators and the amount of reasonable methods to address them. As mentioned, for user experience the indicator feeling of control could be added to the theoretical framework. Similarly, it may be useful to extend the USUS evaluation framework on other impact factors like economics, physical environment, or ethics. Moreover, new methods (see Table 6.1) like participatory design and the world cafè approach have been added to the methodological framework. These newly proposed methods still need to be approved in terms of feasibility. Furthermore, the figure of the revised USUS evaluation framework visualizes by means of dashed lines, which indicators still need to be revised regarding the influence of independent variables and their interrelations to other evaluation indicators (see Figure 7.1).

Up to now the framework was only used for collaborative scenarios, but interaction scenarios where "robot-group interaction" or "groups of robots-human interaction" were not executed in case studies. Moreover, it would be of interest to investigate if the framework can also guide HRI research for human-robot interaction scenarios in other context, like in a home context, an educational context, and a therapeutic context. Similarly, future work will be necessary to identify, whether robotic applications are existing that are not suitable for this evaluation framework. Does it only work for humanoid robots or can it also be taken into account for functionally designed or zoopomorphic robots?

The revised USUS evaluation framework (see Figure 7.1) is one conceptual model for exploring human-robot interaction. It has become clear that there are still many concerns that need to be investigated. However, the evaluation framework proved its value as a new more holistic approach to assess human-robot interaction. User-centered approaches, guided by the USUS evaluation framework, can drive robotic design and foster its social acceptability and positive societal impact. The potential of the USUS evaluation framework can be realized best, if it guides the development of human-robot interaction scenarios from the very beginning until the very end of a project, as long as the project team consists of experts from various disciplines who have the same goal: improving human-robot interaction.

Bibliography

Alben L (1996) Quality of experience: defining the criteria for effective interaction design. interactions 3(3):11–15

Altmaninger M (2008) Erweiterung der methoden heuristische evaluierung und user studien durch videoprototypen zur evaluierung von human-robot interaktion. PhD thesis, University of Applied Sciences Salzburg, Salzburg, Austria

Apostolos MK (1985) Exploring user acceptance of a robotic arm: A multidisciplinary case study (choreography, aesthetics, disabled). PhD thesis, Stanford University, Stanford, CA, USA

Arras K, Cerqui D (2005) Do we want to share our lives and bodies with robots? A 2000 people survey. Tech. rep., nr. 0605-001

Badmington N (2000) Posthumanism. Readers in cultural criticism, Palgrave, Houndmills, Basingstoke, Hampshire , New York

Bandura A (1997) Self-efficacy: The exercise of control. Freeman, New York, NY

Bartneck C (2008) Who like androids more: Japanese or US Americans? In: RO-MAN2008: Proceedings of the 17th IEEE International Symposium on Robot and Human Interactive Communication, IEEE, Munich, pp 553–557

Bartneck C, Forlizzi J (2004) A design-centred framework for social human-robot interaction. pp 591–594

Bartneck C, Nomura T, Kanda T, Suzuki T, Kato K (2005) A cross-cultural study on attitudes towards robots. In: HCI International, Las Vegas, USA

Bartneck C, Suzuki T, Kanda T, A Nomura T (2006) The influence of people's culture and prior experiences with AIBO on their attitude towards robots. AI & Society The Journal of Human-Centered Systems 21(1)

Bartneck C, Croft E, Kulic D (2009) Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. International Journal of Social Robotics 1(1):71–81

Battarbee K (2003) Defining co-experience. In: DPPI '03: Proceedings of the 2003 International Conference on Designing Pleasurable Products and Interfaces, ACM, New York, NY, USA, pp 109–113

Bowlby J (1958) The nature of the child's tie to his mother. International Journal of Psychoanalysis 39:350–373

Breazeal C, Scassellati B (1999) How to build robots that make friends and influence people. In: IROS '99: IEEE/RSJ International Conference on Intelligent Robots and Systems, vol 2, pp 858–863

Breazeal C, Hoffman G, Lockerd A (2004) Teaching and working with robots as a collaboration. In: AAMAS2004: Proceedings of the 3rd International Joint Conference on Autonomous Agents and Multiagent Systems, IEEE Computer Society, pp 1030–1037

Brooke J (1996) SUS: A "quick and dirty" usability scale. In: Usability evaluation in industry, pp 189–194

Brooks R (2002) Humanoid robots. Commun ACM 45(3):33–38

Capek J (1920) Rossum's Universal Robots. Prague, CZ

Clarkson E, Arkin RC (2007) Applying heuristic evaluation to human-robot interaction systems. In: FLAIRS: Proceedings of the The 20th International FLAIRS Conference, pp 44–49

Courage C, Baxter K (2004) Understanding your users: A practical guide to requirements methods, tools and techniques. Morgan Kauffman, San Francisco, California

Dahlbäck N, Jönsson A, Ahrenberg L (1993) Wizard of oz studies: why and how. In: IUI '93: Proceedings of the 1st International Conference on Intelligent User Interfaces, ACM, New York, NY, USA, pp 193–200

Dario P, Guglielmelli E, Laschi C, Teti G (1999) Movaid: a personal robot in everyday life of disabled and elderly people. Technology and disability 10:55–71

Dautenhahn K (1999) Robots as social actors: AURORA and the case of autism. In: CT99: Proceedings of the 3rd International Cognitive Technology Conference, Mexico City, pp 359–374

Dautenhahn K (2007a) Methodology and themes of human-robot interaction: A growing research field. International Journal of Advanced Robotic Systems 4(1):103–108

Dautenhahn K (2007b) Socially intelligent robots: dimensions of human-robot interaction. Philosophical Transactions- Royal Society of London Series B Biological Sciences 362:679–704

Dautenhahn K, Walters ML, Woods S, Koay KL, Nehaniv E Chrystopher Land Sisbot, Alami R, Siméon T (2006) How may i serve you?: A robot companion approaching a seated person in a helping context. In: Goodrich MA, Schultz AC, Bruemmer DJ (eds) HRI2006: Proceedings of the 1st ACM/IEEE Conference on Human-Robot Interaction, ACM, Salt Lake City, Utah, USA, pp 172–179

Decker M (2007) Can humans be replaced by autonomous robots? Ethical reflections in the framework of an interdisciplinary technology assessment. In: ICRA2007: Proceedings of the 2007 IEEE International Conference on Robotics and Automation, 2007

Dillon A (2001) User acceptance of information technology. In: Karwowski W (ed) Encyclopedia of Human Factors and Ergonomics, Taylor and Francis, London

Dix A, Abowd GD, Finlay JE (2004) Human-Computer Interaction (3rd Edition). Prentice Hall

Dobson R (2004) Meet Rudy, the world's first "robodoc". BMJ 329(7464):474, `http://www.bmj.com/cgi/reprint/329/7464/474-b.pdf`

Drury JL, Scholtz J, Yanco HA (2003) Awareness in human-robot interactions. In: Proceedings of the 2003 IEEE Conference on Systems, Man and Cybernetics, Washington, pp 912– 918

Drury JL, Hestand D, Yanco HA, Scholtz J (2004) Design guidelines for improved human-robot interaction. In: CHI '04: CHI '04 extended abstracts on Human factors in computing systems, ACM, New York, NY, USA, pp 1540–1540

Drury JL, Yanco HA, Scholtz J (2005) Using competitions to study human-robot interaction in urban search and rescue. interactions 12(2):39–41

Epstein RG (1994a) The case of the killer robot (part 1). SIGCAS Comput Soc 24(3):20–28

Epstein RG (1994b) The case of the killer robot (part 2). SIGCAS Comput Soc 24(4):12–32

Flick U, von Kardorff E, Steinke I (2004) A companion to qualitative research. Sage Publications

Fong T, Nourbakhsh I, Dautenhahn K (2003) A survey of socially interactive robots. Robotics and Autonomous Systems 42:143–166

Fong TW, Thorpe C, Baur C (2001) Collaboration, dialogue, and human-robot interaction. In: Proceedings of the 10th International Symposium of Robotics Research, Lorne, Victoria, Australia, Springer-Verlag, London

Forlizzi J (2005) Robotic products to assist the aging population. interactions 12(2):16–18

Forlizzi J (2007) How robotic products become social products: An ethnographic study of cleaning in the home. In: HRI2007: Proceedings of the 2nd ACM/IEEE International Conference on Human-Robot Interaction, ACM, New York, NY, USA, pp 129–136

Goetz J, Kiesler S (2002) Cooperation with a robotic assistant. In: CHI '02: CHI '02 extended abstracts on Human factors in computing systems, ACM, New York, NY, USA, pp 578–579

Gouldner AW (1960) The norm of reciprocity: A preliminary statement. American Sociological Review 25:161–178

Gray CH (2001) Cyborg citizen: Politics in the posthuman age. Routledge, New York

Green S, Billinghurst XQ Mand Chen, Chase G (2008) Human-robot collaboration: A literature review and augmented reality approach in design. International Journal of Advanced Robotic Systems 5(1):1–18

Hammel J, Hall K, Lees D, Leifer L, Van der Loos M, Perkash I, Crigler R (1989) Clinical evaluation of a desktop robotic assistant. J Rehabil Res Dev 26(3):1–16

Harkness J, van de Vijver F, Mohler PP (2003) Cross-Cultural Survey Methods. John Wiley & Sons Inc.

Hassenzahl M (2001) The effect of perceived hedonic quality on product appealingness. Int J Hum Comput Interaction 13(4):481–499

Hassenzahl M (2003) The thing and I: Understanding the relationship between user and product. In: Blythe M, Overbeeke C, Monk AF, Wright PC (eds) Funology. From Usability to Enjoyment, Kluwer, Dordrecht, pp 31–42

Hassenzahl M, Tractinsky N (2006) User experience – a research agenda. Behaviour and Information Technology 25:91–97

Heerink B Mand Kröse, Wielinga B, Evers V (2006a) Studying the acceptance of a robotic agent by elderly users. International Journal of Assistive Robotics and Mechatronics 7(3):25–35

Heerink M, Kröse B, Evers V, Wielinga B (2006b) The influence of a robot's social abilities on acceptance by elderly users. In: RO-MAN2006: Proceedings of the 15th IEEE International Symposium on Robot and Human Interactive Communication, pp 521–526

Heerink M, Kröse B, Evers V, Wielinga B (2008a) The influence of social presence on enjoyment and intention to use of a robot and screen agent by elderly users. In: RO-MAN2008: The 17th IEEE International Symposium on Robot and Human Interactive Communication, pp 695–700

Heerink M, Kröse B, Wielinga B, Evers V (2008b) Enjoyment intention to use and actual use of a conversational robot by elderly people. In: HRI2008: Proceedings of the 3rd ACM/IEEE International Conference on Human-Robot Interaction, ACM, New York, NY, USA, pp 113–120

Hinds PJ, Roberts TL, Jones H (2004) Whose job is it anyway? A study of human-robot interaction in a collaborative task. Human-Computer Interaction 19(1-2):151–181

Hornyak TN (2006) Loving the machine: The art and science of Japanese robots. Kodansha international, Tokyo, New York, London

Hüttenrauch H (2007) From HCI to HRI: Designing interaction for a service robot. PhD thesis, KTH, Numerical Analysis and Computer Science, Stockholm, Sweden

ISO 13407 (1999) Human-Centered Design Process for Interactive Systems. International Organization for Standardization

ISO 9241-11 (1998) Ergonomics of Human System Interaction - Part 11: Guidance on Usability. International Organization for Standardization

Kahn J PH, Freier N, Friedman B, Severson R, Feldman E (2004) Social and moral relationships with robotic others? In: RO-MAN2004: Proceedings of the 13th IEEE International Workshop on Robot and Human Interactive Communication, pp 545–550

Kaplan F (2000) Free creatures: The role of uselessness in the design of artificial pets. In: Proceedings of the 1st Edutainment Robotics Workshop

Kaplan F (2001) Artificial attachment: Will a robot ever pass the ainsworth's strange situation test? In: Humanoids2001: Proceedings of the IEEE-RAS International Conference on Humanoid Robots, pp 125 – 132

Kaplan F (2004) Who is afraid of the humanoid? Investigating cultural differences in the acceptance of robots. International Journal of Humanoid Robotics 1(3):465–480

Kaplan F (2005) Everyday robotics: Robots as everyday objects. In: sOc-EUSAI '05: Proceedings of the 2005 Joint Conference on Smart Objects and Ambient Intelligence, ACM, pp 59 – 64

Kelley JF (1984) An iterative design methodology for user-friendly natural language office information applications. ACM Trans Inf Syst 2(1):26–41

Kiesler S, Hinds P (2005) Introduction to this special section on human-robot interaction. Human-Computer Interaction, Spec Issue on Human-Robot Interaction (1-2)

Knorr-Cetina K (1997) Sociality with objects: Social relations in postsocial knowledge societies. Theory Culture Society 14(4):1–30

Koay KL, Walters ML, Dautenhahn K (2005) Methodological issues using a comfort level device in human-robot interactions. In: RO-MAN2005: Proceedings of the 14th IEEE International Workshop on Robot and Human Interactive Communication, pp 359–364

Kozima H, Yasuda Y, Nakagawa C (2007) Social interaction facilitated by a minimally-designed robot: Findings from longitudinal therapeutic practices for autistic children. In: RO-MAN2007: Proceedings of the 16th IEEE International Symposium on Robot and Human Interactive Communication, pp 599–604

Krueger RA, Casey M (2000) Focus groups: A practical guide for applied research. Sage Publications, Thousand Oaks, CA

Kulic D, Croft E (2003) Strategies for safety in human robot interaction. In: ICAR2003: Proceedings of the IEEE International Conference on Advanced Robotics, Coimbra, Portugal, pp 644–649

Kulic D, Croft EA (2005) Anxiety detection during human-robot interaction. In: IROS2005: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, pp 616–621

Kulic D, Croft EA (2007) Affective state estimation for human-robot interaction. IEEE Transactions on Robotics 23(5):911–1000

Law E, Roto V, Vermeeren AP, Kort J, Hassenzahl M (2008) Towards a shared definition of user experience. In: CHI '08: CHI '08 extended abstracts on Human factors in computing systems, ACM, New York, NY, USA, pp 2395–2398

Lee MK, Forlizzi J, Rybski PE, Crabbe F, Chung W, Finkle J, Glaser E, Kiesler S (2009) The Snackbot: Documenting the design of a robot for long-term human-robot interaction. In: HRI2009: Proceedings of the 4th ACM/IEEE International Conference on Human-Robot Interaction, ACM, New York, NY, USA, pp 7–14

Lohse M, Hahnheide M (2008) Evaluating a social home tour robot applying heuristics. ROMAN2008: Workshop on Robots as Social Actor (position paper)

Lohse M, Hanheide M, Rohlfing KJ, Sagerer G (2009) Systemic interaction analysis (sina) in hri. In: HRI '09: Proceedings of the 4th ACM/IEEE international conference on Human robot interaction, ACM, New York, NY, USA, pp 93–100

MacDorman KF, Ishiguro H (2006) The uncanny advantage of using androids in social and cognitive science research. Interaction Studies 7(3):297–337

Marble JL, Few DA, Bruemmer DJ (2004) I want what you've got: Cross platform usability and human-robot interaction assessment. In: PerMIS '04: Proceedings of the Performance Metrics for Intelligent Systems Workshop, Gaithersburg, MD, USA

Minge M (2005) Methoden zur erhebung emotionaler aspekte bei der interaktion mit technischen systemen. PhD thesis, Freie Universitaet Berlin, Berlin

Mori M (1970) The uncanny valley. Energy 7:33–35

Mudry PA, Degallier S, Billard A (2008) On the influence of symbols and myths in the responsibility ascription problem in roboethics - A roboticists perspective. In: RO-MAN2008: Proceedings of the 17th IEEE International Symposium on Robot and Human Interactive Communication, pp 563–586

Mutlu B, Forlizzi J (2008) Robots in organizations: The role of work flow, social, and environmental factors in human-robot interaction. In: HRI2008: Proceedings of the 3rd ACM/IEEE International Conference on Human-Robot Interaction, ACM, New York, NY, USA, pp 287–294

Mutlu B, Shiwa T, Kanda T, Ishiguro H, Hagita N (2009) Footing in human-robot conversations: how robots might shape participant roles using gaze cues. In: HRI2009: Proceedings of the 4th ACM/IEEE International Conference on Human-Robot Interaction, ACM, New York, NY, USA, pp 61–68

Nielsen J (1992) Finding usability problems through heuristic evaluation. In: CHI1992: Proceedings of the ACM/SIGCHI Conference on Human Factors in Computing Systems, ACM Press, pp 373–380

Nomura T, Kanda T, Suzuki T (2006) Experimental investigation into influence of negative attitudes toward robots on human-robot interaction. AI Soc 20(2):138–150

Norman DA (2004) Emotional design: Why we love (or hate) everyday things. Basic Books, New York, USA

Olsen DR, Wood SB (2004) Fan-out: measuring human control of multiple robots. In: CHI2004: Proceedings of the ACM/SIGCHI Conference on Human Factors in Computing Systems, ACM, New York, NY, USA, pp 231–238

Pedhazur E, Schmelkin E Land Pedhazur (1991) Measurement, Design, and Analysis: An Integrated Approach. Lawrence Erlbaum Associates

Preece J, Rogers Y, Sharp H (2002) Interaction Design. John Wiley & Sons, Inc., New York, NY, USA

Reeves B, Nass C (1996) The media equation: How people treat computers, televisions, and new media like real people and places. Cambridge University Press, New York

Riek L, Robinson P (2009) Affective centered design for interactive robots. In: AISB2009: Proceedings of the 23rd AISB convention:Symposium on New Frontiers in Human-Robot Interaction, SSAISB, pp 102–108

Riek LD, Rabinowitch TC, Chakrabarti B, Robinson P (2009) How anthropomorphism affects empathy toward robots. In: HRI2009: Proceedings of the 4th ACM/IEEE International Conference on Human-Robot Interaction, ACM, New York, NY, USA, pp 245–246

Robins B, Dautenhahn K, Te B, Billard A (2005) Robotic assistants in therapy and education of children with autism: can a small humanoid robot help encourage social interaction skills? Univers Access Inf Soc 4(2):105–120, 1104626

Robowatch (2007) http://www.robowatch.de

Rosson MB, Carroll J (2002) Usability engineering: Scenario-based development of Human-Computer Interaction. Morgan Kauffman, San Francisco et al.

Scholtz J (2002) Evaluation methods for human-system performance of intelligent systems. In: PerMIS'02: Performance Metrics for Intelligent Systems Workshop, Gaithersburg, MD, USA

Schräder P (2005) Konsequenzen des evidenzberichts und der einzelfallbewertung der methode Robodoc für die begutachtungspraxis des MDK, das behandlungsfehlermanagement der krankenkassen und den grundsätzlichen umgang mit innovationen. Results of the hta-report on robodoc and 44 individual surveys and the consequences for the medical advisory service, the management of malpractices and the principles of future handling of innovations 6:389–395

Severinson-Eklundh K, Green A, Hüttenrauch H (2003) Social and collaborative aspects of interaction with a service robot. Robotics and Autonomous Systems 42(3-4):223–234

Shiwa T, Kanda T, Imai M, Ishiguro H, Hagita N (2008) How quickly should communication robots respond? In: HRI2006: Proceedings of the 3rd ACM/IEEE International Conference on Human- Robot Interaction, pp 113–120

Song TH, Park JH, Chung SM, Hong SH, Kwon KH, Lee S, Jeon JW (2007) A study on usability of human-robot interaction using a mobile computer and a human interface device. In: MobileHCI2007: Proceedings of the 9th international conference on Human computer interaction with mobile devices and services, ACM, New York, NY, USA, pp 462–466

Sparrow R, Sparrow L (2006) In the hands of machines? The future of aged care. Minds Mach pp 141–161

Staudte M, Crocker MW (2009) Visual attention in spoken human-robot interaction. In: HRI2009: Proceedings of the 4th ACM/IEEE International Conference on Human-Robot Interaction, ACM, New York, NY, USA, pp 77–84

Steinfeld A, Fong T, Kaber D, Lewis M, Scholtz J, Schultz A, Goodrich M (2006) Common metrics for human-robot interaction. In: HRI2006: Proceedings of the 1st ACM/IEEE Conference on Human-Robot Interaction, ACM, New York, NY, USA, pp 33–40

Stone D, Jarrett C, Woodroffe M, Minocha S (2005) User interface design and evaluation. Morgan Kauffman, Amsterdam et al.

Stubbs K, Hinds PJ, Wettergreen D (2007) Autonomy and common ground in human-robot interaction: A field study. IEEE Intelligent Systems 22(2):42–50

Takayama L, Ju W, Nass C (2008) Beyond dirty, dangerous and dull: What everyday people think robots should do. In: HRI2008: Proceedings of the 3rd ACM/IEEE International Conference on Human-Robot Interaction, ACM, New York, NY, USA, pp 25–32

Thrun S (2004) Toward a framework for human-robot interaction. Human-Computer Interaction 19(1):9–24

Turkle S (1984) The second self: Computers and the human spirit. Simon & Schuster, Inc., New York, NY, USA

Turn R (1984) Courses on societal impacts of computers. SIGCAS Comput Soc 13, 14(4, 1-3):14–16

US Department of Labor (1994) Dictionary of occupational titles

Venkatesh V, Davis FD (2000) A theoretical extension of the technology acceptance model: Four longitudinal field studies. Manage Sci 46(2):186–204

Venkatesh V, Morris MG, Davi GB, Davis FD (2003) User acceptance of information technology: Toward a unified view. MIS Quarterly 27(3)

Wada K, Shibata T (2006) Robot therapy in a care house - Its sociopsychological and physiological effects on the residents. In: ICRA2006: Proceedings of the 2006 IEEE International Conference on Robotics and Automation, pp 3966–3971

Watson D, Clark L, Tellegen A (1988) Development and validation of brief measures of positive and negative affect: The PANAS scales. J Pers Soc Psychol 54(6):1063–1070

Wehmeyer K (2007) Assessing users' attachment to their mobile devices. In: Proceedings of the international Conference on the Management of Mobile Business

Weiss A, Bernhaupt R, Tscheligi M, Wollherr D, Kuhnlenz K, Buss M (2008) A methodological variation for acceptance evaluation of Human-Robot Interaction in public places. In: RO-MAN 2008: Proccededings of the 17th IEEE International Symposium on Robot and Human Interactive Communication, pp 713–718

Weiss A, Bernhaupt R, Lankes M, Tscheligi M (2009a) The USUS evaluation framework for Human-Robot Interaction. In: AISB2009: Proceedings of the Symposium on New Frontiers in Human-Robot Interaction, SSAISB (ISBN - 190295680X), Edinburgh, Scottland, pp 158–165

Weiss A, Bernhaupt R, Schwaiger D, Altmaninger M, Buchner R, Tscheilig M (2009b) User experience evaluation of multimodal interaction techniques in human-robot collaboration using a wizard of oz approach. In: Humanoids2009: Proceedings of the 9th IEEE-RAS International Confernece on Humanoids Robotics, accepted for publication

Weiss A, Bernhaupt R, Tscheligi M, Yoshida E (2009c) Addressing user experience and societal impact in a user study with a humanoid robot. In: AISB2009: Proceedings of the Symposium on New Frontiers in Human-Robot Interaction, SSAISB (ISBN - 190295680X), Edinburgh, pp 150–157

Weiss A, Igelsböck J, Wurhofer D, Tscheligi M (2009d) Looking forward to a "Robotic Society"? - Imaginations of Future Human-Robot Relationships. In: HRPR2009: Proceedings of the 2nd International Conference on Human-Robot Personal Relationships

Weiss A, Igelsböeck J, Calinon S, Billard A, Tscheligi M (2009e) Teaching a humanoid: A user study on learning by demonstration with HOAP-3. In: Ro-Man2009: Proceedings of the 18th IEEE International Symposium on Robot and Human Interactive Communication, accepted for publication

Weiss A, Wurhofer D, Igelsböck J, Buchner R, Förster F, Pierro P, Tscheligi M (2009f) Deliverable 3.4-7@m30: Preliminary report on usability, social acceptance, user experience and societal impact evaluation. Tech. rep., http://www.robot-at-cwe.eu/

Weiss A, Wurhofer D, Igelsböck J, Buchner R, Förster F, Tscheligi M (2009g) Deliverable 4.6@m36: User evaluation report. Tech. rep., http://www.robot-at-cwe.eu/

Weiss A, Wurhofer D, Lankes M, Tscheligi M (2009h) Autonomous vs. tele-operated: How people perceive human-robot collaboration with HRP-2. In: HRI '09: Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction, ACM, New York, NY, USA, pp 257–258

Weiss A, Igelsböck J, Tscheligi M, Bauer A, Kühnlenz K, Wollherr D, Buss M (2010) Robots asking for directions – the willingness of passers-by to support robots. In: HRI '10: Proceedings of the 5th ACM/IEEE International Conference on Human Robot Interaction, ACM, New York, NY, USA, accepted for publication

Wharton C, Bradford J, Jeffries R, Franzke M (1992) Applying cognitive walkthroughs to more complex user interfaces: Experiences, issues, and recommendations. In: CHI1992: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, New York, NY, USA, pp 381–388

Woods S, Walters M, Koay K, Dautenhahn K (2006) Comparing human robot interaction scenarios using live and video based methods: towards a novel methodological approach. In: AMC2006: Proceedings of the 9th IEEE International Workshop on Advanced Motion Control, pp 750–755

Yanco HA, Drury JL, Scholtz J (2004) Beyond usability evaluation: Analysis of human-robot interaction at a major robotics competition. Human-Computer Interaction 19(1,2):117–149

Appendix A

List of Annex Documents

All the referenced appendices can be found on the accompanying CD.

Annex 1: The NARS Questionnaire

Annex 2: The AttrakDiff Questionnaire

Annex 3: The SUS Questionnaire

Annex 4: The PANAS Questionnaire

Annex 5: The Focus Group Guidelines

Annex 6: The Heuristic Evaluation Manual and Template

Annex 7: The Questionnaires for the ACE Field Trials

Annex 8: The Interview Guidelines of the Expert Interviews

Annex 9: The UX- and SoAc-Questionnaire in Spanish

Annex 10: The UX- and SoAc-Questionnaire in German

Annex 11: The UX- and SoAc-Questionnaire in Japanese

Annex 12: The UX- and SoAc-Questionnaire in French (Online Survey)

Appendix B

List of Abbreviations

3D dull, dirty, dangerous
ACE Autonomous City Explorer
ACM Association for Computing Machinery
AIST National Institute of Advanced Industrial Science and Technology
ANOVA Analysis of Variances
ATT Attractiveness
CHI Computer-Human Interaction
CNRS Centre National de la Recherche Scientifique
CSCW Computer Supported Cooperative Work
DRAGADOS Spanish Construction Company
E Emotion
ECG Electrocardiogram
EE Effort Expectancy
ELAN Video Annotation and Language Analysis Tool
EMB Embodiment
EPFL Ecole Polytechnique Federale de Lausanne
EURON EUropean RObotics research Network
FS Feeling of Security
HCI Human-Computer Interaction
HHI Human-Human Interaction
HOAP-3 Humanoid Robot used by EPFL and UC3M
HOP Human-Oriented Perception
HQI Hedonic Quality Identification
HQS Hedonic Quality Stimulation
HRC Human-Robot Collaboration
HRI Human-Robot Interaction
HRP-2 Humanoid Robot used by CNRS
HRV Heart Rate Variability
ICT&S Information and Communication Technologies & Society
IEEE Institute of Electrical and Electronics Engineers
IFR International Federation of Robotics
IRIT Institute de Recherche en Informatique de Toulouse
LAAS Laboratory for Analysis and Architecture of Systems
LASA Learning and Algorithms and System Laboratory
LED Light Emitting Diode
LSD Least Significant Difference Test
MAUSE Towards the MAturation of Information Technology USability Evaluation
MSTC Manufacturing Science and Technology Center
NARS Negative Attitude towards Robots Scale
OG Human Operator in Germany
OJ Human Operator in Japan
PANAS Positive and Negative Affect Schedule
PC Perceived Competency
PE Performance Expectancy
PhD Doctor of Philosophy
PQ Pragmatic Quality

RE Relationship Expectancy
RH-1 Humanoid Robot used by UC3M
RUR Rossmus Universal Robots
SAS Space Application Services
SD Standard Deviation
SoAc Social Acceptance
SCR Skin Resistance Response
SRR Skin Resistance Response
SRT System Response Time
STSM Short Term Scientific Mission
SUS System Usability Scale
TP Test Participant
TV Television
UC3M Universidad Carlos III de Madrid
UCD User-Centered Design
USUS Usability, Social Acceptance, User Experience, and Societal Impact
UTAUT Unified Theory of Acceptance & Use of Technology
UX User Experience
VDMA Verband Deutscher Maschinen- und Anlagenbau
WOZ Wizard of Oz

Appendix C

Curriculum Vitae

Astrid Weiss

Education

2005-2010: PhD in Human-Computer Interaction at the ICT&S Center, University of Salzburg, Austria.
PhD thesis on "Validation of an evaluation Framework for Human-Robot Interaction".
Supervisors: Prof. Martin Weichbold and Prof. Manfred Tscheligi.
Oral exam: March 17th, 2010.

2000-2005: MA in Sociology, specialization in methods of empirical social research and apllied statistics, University of Salzburg, Austria.
MA thesis on "Innovationen im Österreichischen Alpentourismus"(eng: *Innovations in the Austrian Alp Tourism*).
Supervisor: Prof. Reinhard Bachleitner.
Oral exam: March 4th, 2005.

Areas of Expertise

- Human-Computer Interaction theories and methodologies
- Human-Robot Interaction theories and methodologies
- Social Interfaces
- Methods of empirical social research and applied statistics

Research Experience

2007-2010: Research and Teaching Assistant, ICT&S-Center, Human-Computer Interaction & Usability Unit, led by Prof. Manfred Tscheligi, University of Salzburg, Austria.

Oct. 2009: Visiting research fellow at the CNRS-AIST Joint Japanese-French Robotics Laboratory (JRL), led by Prof. Abderrahmane Kheddar, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Japan (2 weeks).

Sep. 2008: Short Term Scientific Mission (cost-action294: MAUSE) at the IHCS-IRIT, led by Prof. Philippe Palanque , University of Toulouse, France (1 month).

2006-2007: Project collaborator for "ITV4all" project, ICT&S-Center Human-Computer Interaction & Usability Unit, University of Salzburg.

Oct. 2005 - Jun. 2006: Internship, ICT&S-Center Human-Computer Interaction & Usability Unit, University of Salzburg (3 months).

Oct. 2003 - Jun. 2004: Collaboration on the interdisciplinary study-project: "Leben 2014" (eng: *Life 2014*, development of future plans for the Austrian region Upper-Pinzgau with the sociological scenario technique), Institute of Sociology, led by Prof. Reinhard Bachleitner (6 months).

Academic Service

Workshop Organizations:

- Workshop on Societal Impact: How Socially Accepted Robots Can be Integrated in our Society to be held at HRI2009.
  Co-organizers: Prof. Manfred Tscheligi (University of Salzburg, AUT), Prof. Aude Billard (EPFL, Lausanne, CH), Prof. Kerstin Dautenhahn (University of University of Hertfordshire, UK)
- Workshop on Robots as Social Actors: Evaluating Social Acceptance and Societal Impact of Robotic Agents hold at IEEE RO-MAN 2008.
  Co-organizers: Prof. Manfred Tscheligi (University of Salzburg, AUT), Prof. Aude Billard (EPFL, Lausanne, CH)

Affiliations to Research Projects:

- European Specific Targeted Research Project IURO (Interactive Urban Robot). 3-years project, started in 2010, under contract FP7-ICT-248314, involving 4 universities and 1 industrial partner.
- Christian Doppler Laboratory "Contextual Interfaces". 5-years project, started in 2009, involving University of Salzburg and 2 industrial partners.

- European Specific Targeted Research Project ROBOT@CWE (Advanced robotic systems in future collaborative working environments). 3-years project (2006-2009) under contract FP6-IST-034002, involving 7 universities and 2 industrial partners.
- FFG project iTV4All (User friendly design of remote controls for interactive TV). 2-years project (2006-2007), involving the ICT&S Center, University of Salzburg and an Austrian SME.

Teaching Experience

Teaching assistance at the University of Salzburg, Department of Computer Science, Courses in Human-Computer Interaction (since 2007).

- WS 2009/10: VP User Interface Design
- WS 2008/09: VP User Interface Design, SE HCI Studio
- SS 2008: PS Usability and User Experience Engineering, VP HCI Anwendungen (eng: *HCI Applications*), SE HCI Innovation
- WS 2007/08: PS Grundlagen der HCI (eng: *Basics in HCI*), PS Usability and User Experience Engineering
- SS 2007: SE HCI Studio II (Tutor of Prof. Tscheligi)

Appendix D

Declaration

This thesis is a presentation of my original research work. Wherever contributions of others are involved, every effort is made to indicate this clearly, with due reference to the literature, and acknowledgment of collaborative research and discussions. The work was done under the guidance of Prof. Martin Weichbold, at the department of Political Science and Sociology, University of Salzburg, Austria and Prof. Manfred Tscheligi, at the HCI & Usability Unit of the ICT&S Center, University of Salzburg, Austria.

Mag. Astrid Weiss

In my capacity as supervisor of the candidate's thesis, I certify that the above statements are true to the best of my knowledge.

Prof. Martin Weichbold/Prof. Manfred Tscheligi
Date: