SEBASTIAN SCHNEIDER

SOCIALLY ASSISTIVE ROBOTS FOR EXERCISING SCENARIOS

SOCIALLY ASSISTIVE ROBOTS FOR EXERCISING SCENARIOS

SEBASTIAN SCHNEIDER

Studies on group effects, feedback, embodiment and adaptation

A doctoral thesis presented for the degree of
Doctor of Engineering (Dr.-Ing.) at

Faculty of Technology
Bielefeld University
Inspiration 1
33619 Bielefeld
Germany

REVIEWERS
Prof. Dr. Franz Kummert
Prof. Dr. Elisabeth André

BOARD
Prof. Dr. Ulrich Rückert
Dr. Laura Hoffmann

DEFENDED AND APPROVED
January 28, 2019

Printed on permanent paper as per ISO 9706.

ABSTRACT

Even though positive effects of being physically active are commonly known, only a few parts of the world population are sufficiently active. The World Health Organization (WHO) states that this problem affects 31% of the adult's world population and 80% of the adolescent population. Appropriate levels of physical activity (PA) are essential to prevent obesity in childhood and to keep a Quality of Life (QOL) in old age but are also essential to prevent other Noncommunicable Diseases (NCDs). Thus, physical inactivity is growing into a severe problem globally, and there is a growing need to motivate people to become more physically active during their lifetime. One primary cause that raises PA levels is having a peer or help from professionals. However, having assistance is not possible in every situation. It might be challenging to find and schedule with a partner or to commute to other places. Roboticist introduced Socially Assistive Robot (SAR) as an assistive tool for exercising, cognitive or rehabilitation tasks. This thesis explores SAR in the context of exercising along four features that have been partly targeted but not yet thoroughly investigated. These features are a) the social role of the robot, b) encouragement c) embodiment and d) adaptation. First, this thesis looks at the motivational effects of exercising with SAR concerning features a) - c). Second, this thesis questions how a system can adapt to the user, and whether adaptivity or adaptability is enough to close the gap between user needs and system behavior. I conducted studies that test the different features by assessing subjective ratings of the robot as well as measurable motivational variables (e.g., time spent exercising with the robot) in a bodyweight workout scenario.

The results show that features a) - c) have a positive influence on user's exercising time. Additionally, users perceive a robot companion as more likable than a robot instructor or a human partner. Furthermore, an adaptive robot increases the associated competence and quality of relationship compared to an adaptable robot. However, the results also show that the robot does not always have to exercise along with the user. In situations where it is not possible, the robot could also only give encouraging feedback. This thesis backs up earlier findings of using SAR by replicating motivational group exercising effects found in Human-Human Interaction (HHI). Thus, the evidence that SARs are a suitable tool for rehabilitative interventions increases which may convince health experts to consider SAR as a useful therapeutic tool. Nevertheless, this thesis evaluated the effects only during short-term interactions. Thus, proving that the found effects are long-lasting is essential for future studies.

PUBLICATIONS

[Sch14] S. Schneider et al. "How Socially Assistive Robots Supporting on Cognitive Tasks Perform." In: *Proceedings of the 50th Anniversary Convention of the AISB*. 2014.

[Sch15a] S. Schneider et al. "Long-Term Feedback Mechanisms for Robotic Assisted Indoor Cycling Training." In: *Proceedings of the 3rd International Conference on Human-Agent Interaction*. HAI '15. Daegu, Kyungpook, Republic of Korea: ACM, 2015, pp. 157–164. DOI: 10.1145/2814940.2814962.

[Sch15b] S. Schneider et al. "Reusable Motivational Instruction Patterns for Socially Assistive Robots." In: *Workshop Towards Intelligent Social Robots ? Current Advances in Cognitive Robotics*. Korea, 2015.

[Sch16a] S. Schneider et al. "Exercising with a humanoid companion is more effective than exercising alone." In: *Humanoid Robots (Humanoids), 2016 IEEE-RAS 16th International Conference on*. IEEE. 2016, pp. 495–501.

[Sch16b] S. Schneider et al. "Motivational Effects of Acknowledging Feedback from a Socially Assistive Robot." In: *Social Robotics: 8th International Conference, ICSR 2016, Kansas City, MO, USA, November 1-3, 2016 Proceedings*. Ed. by A. Agah et al. Cham: Springer International Publishing, 2016, pp. 870–879. ISBN: 978-3-319-47437-3. DOI: 10.1007/978-3-319-47437-3_85.

[Sch17a] S. Schneider et al. "A framework for designing socially assistive robot interactions." In: *Cognitive Systems Research* 43 (2017), pp. 301–312.

[Sch17b] S. Schneider et al. "Does the User's Evaluation of a Socially Assistive Robot Change Based on Presence and Companionship Type?" In: *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. ACM. 2017, pp. 277–278.

[Sch17c] S. Schneider et al. "Exploring Embodiment and Dueling Bandits for Preference Adaptation in Human-Robot Interaction." In: *Proceedings of the 26th IEEE International Symposium on Robot and Human Interactive Communication*. 2017.

[Sch18a] S. Schneider et al. "Adaptive versus adaptable robots: Effects on the perceived competence and relationship." in preperation. 2018.

[Sch18b] S. Schneider et al. "Embodied socially assistive robots are more motivating than virtual agents." in preperation. 2018.

[Sch18c] S. Schneider et al. "PREC 2018: Personal Robots for Exercising and Coaching." In: *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. ACM. 2018, pp. 401–402.

ACKNOWLEDGMENTS

Thanks to Franziska Wäschle, from the time we first met, you supported and cared for me. You helped me keep my balance and enjoy all the small things in life: cooking, reading, gardening, dancing, laying in the grass, strolling around. You contributed a lot to my personal, literate and critical development during this time. Thus, you encouraged me to try to understand myself better. This thesis would have never existed if we had never met.

Thanks to my parents and family, who always asked me when I will finish. I know it was a time where I was not so much there for you as you would have liked. Mainly, I would also like to thank you for raising me to a person who stays relaxed and calm even in the face of writing a thesis. I guess you did a good job there.

From an academic point of view, I would like to thank Marc Hahneide for giving me the opportunity to work for him as a student research assistant. He raised my interest in robotics, and without him, I would have probably become a genome researcher. Thanks to Sebastian Wrede I had the opportunity to be one of the first in our institute to work with him on the Nao platform, back in 2009. He also encouraged and supported me to stay in academia. Thanks a lot to Franz Kummert for supervising my thesis and giving me the academic freedom to pursue my research interests. I am really grateful for the trust and support he provides me. Also, thanks to Britta Wrede for letting me work in this great group and giving me the opportunity to teach social robotics to students. Besides, many thanks to all my colleagues - with you the time during my thesis felt more like fun than like work. Thanks for the support you gave me and your time, always when I entered your offices and had questions, ideas or just wanted to chat. I would also like to thank Prof. Elisabeth André for reviewing my thesis as an external referee. Thanks also to the CITEC Graduate School for funding my research project and allowing me to travel around the world to go to conferences and summer schools. These times will always be memorable for me. Therefore, I would also like to thank all the people I met in different venues.

At last, a big thanks to all my friends. I am pleased to have such an intimate network of close people. There are (unfortunately) too many I would like to thank personally. However, I would like to thank Norman for being a companion, since we started studying 12 years ago. Because of you, the past years have been full of memories of joy and fun. Moreover, I would also like to thank Cecilé, Nina, Miriam, Florian, Vera, Torben and Christian for giving me feedback on my thesis.

CONTENTS

Appendices

LIST OF FIGURES

LIST OF TABLES

LIST OF HYPOTHESES

1 INTRODUCTION

A Quality of Life (QOL) during the lifespan requires appropriate physically active [Biz07]. Even though everyone knows that being physically active is good for one's health, only 31% of the world population is sufficiently active [Hal12]. A variety of factors influence people to be physical active. These determinants are a) demographic and biological factors, b) psychological, cognitive and emotional factors, c) behavioral attributes and skills, d) social and cultural factors, e) environment and f) physical activity characteristics. Having social support was consistently found to be an important factor for adults to become physically active [Tro02]. However, appropriate assistance from peers, coaches or physicians, which could facilitate starting and sticking to a workout, is not available for everybody. It includes finding and scheduling with the associates and often to commute to other places. Recently, various types of technology have been introduced to assist people in their daily life. However, few of these agents have an embodiment that could create the perception of an equal exercising partner.

Hence, Socially Assistive Robots (SARs) have been introduced as a suitable tool to facilitate motivation in rehabilitative or exercising tasks [Fei05]. The rationale for the usage of robot is because people are likely to anthropomorphize non-biological artifacts and treat media and technology human-like [Epl07; Ree97]. Consequently, it allows transferring motivational group effects in exercising observed in Human-Human Interaction (HHI) and apply them in Human-Robot Interaction (HRI) scenarios. Using robots as exercising peers might help people to exercise more. However, there are some lacks in understanding (motivational) effects when exercising with a robotic partner on the dimensions of the social role, encouragement, embodiment and adaptivity, which I will explain later in this chapter.

Besides robots, there are multiple ways to target the problem of keeping appropriate levels of physical activity. The changing lifestyle of people (i.e., working habit, means of transportation) and environmental factors (i.e., urbanization, air pollution) also influence the levels of activity [Hal12]. This thesis presents perspectives and applied research on how computer science and robotics could contribute to overcoming the widespread problems of being physically active.

1.1 THESIS MOTIVATION

The World Health Organization (WHO) affirms in their published key facts about physical activity[1] that insufficient physical activity is one of the leading risk factors for death worldwide and a key risk factor for Noncommunicable Diseases (NCDs) such as cardiovascular diseases, cancer, and diabetes. A low-level of physical activity is estimated to be the leading cause of approximately 21-25% of breast and colon cancers, 27% of diabetes and 30% of ischaemic heart disease burden. On the other side, sufficient physical activity has significant health benefits and contributes to preventing NCDs. It reduces the risk of hypertension, coronary heart disease, stroke, diabetes, breast and colon cancer, depression and the risk of falls. It improves bone and functional health and is fundamental for energy balance. Nevertheless, one in four adults is not active enough, and more than 80% of the world's adolescent population is insufficiently physically active. Knowing about the importance of physical activity does not necessarily help to increase it. Even though 4 of 5 cardiac patients know the importance of physical activity, only 39% stick to an exercise regimen [Tat15]. These facts lead to the conclusion that people who are insufficiently active have an increased risk of death compared to people who are sufficiently active [WHO18].

However, getting people motivated to increase their physical activity is a challenging problem [Bau02]. Some of the earliest technologies to promote physical activity at homes were exercise videotapes (with aerobic videos by Jane Fonda as one of the most prominent representatives [DeB87]). Later, there appeared pedometers, accelerometers, heart rate monitors and global positioning system as tools for individuals to motivate them to be more active by providing feedback about a user's activity performance, e.g., [Bou13]. Moreover, some more interactive technologies appeared like Active Video Game (AVG) (e.g., Wii Sports or Wii Fit). Those technologies are designed to create an engaging game experience at home but are additionally used in schools, community and senior centers to promote physical activity. Furthermore, hospitals and physical therapy centers use them in their rehabilitation programs also [Jun09]. In conclusion, the emergence of attempts to help people to become more active by using increasingly advanced technologies complemented by the global data on physically inactivity indicates that this demand will likely grow in the future.

Therefore, this thesis explores new technologies that can be used to motivate people to exercise. Motivation is usually defined as the process of starting, maintaining and repeating a goal-oriented behavior. It is the cause of one's behavior direction and adherence to this behavior. To increase one's physical activity level requires all of these

1 http://www.who.int/mediacentre/factsheets/fs385/en/, retrieved 08/31/2018

aspects, however this thesis mainly focusses on the element of maintaining an exercise. This aspect of maintaining refers to the question whether robots could be used to motivate people to increase one's exercising duration, repetition number, intensity or effort. This limitation to the maintaining aspect of motivation neglects the initial problem of starting a behavior that increases physical activity or the long-term problem of repeating this behavior. Still, it is a crucial aspect that might contribute to an increase of one's physical activity. Chapter 2 will introduce a selection of important theories for exercising motivation.

1.2 RESEARCH QUESTIONS

The presented data from the WHO shows that there is an overall need for rehabilitation and motivational programs to assist people in increasing their physical activity. This is where HRI research could contribute to the ongoing efforts to enhance physical activity and build robotic tools that help users to exercise more in their daily life. This section outlines the research questions of this thesis and motivates why robots as exercising tools could be a promising approach.

This potential of using robots as exercising peers stems from observations in HHI where being part of a group with superior members increases one's motivation [Web07]. The interdependence in group task results in the so called Köhler Effect which has shown to improve effort in exercising [Fel11; Irw12]. In this effect "the least capable group member exhibits a motivation gain (relative to individual performance) when performing as part of a group on effort-based tasks" [Fel14, p. 99]. Kerr et al. [Ker07] have identified that this effect relies on two factors: 1. unfavorable social comparison with superior group members and 2. being necessary for the group success. Previous research shows that having a superior social companion that exercises co-actively in a conjunctive task can be beneficial for increasing motivation to exercise without any aversion to the task [Fel11; Irw12]. However, a human partner might not always be available to everyone; also there are people with social (physique) anxieties which prevent them from participating in group exercises or activities [Hau04]. While there have been different approaches to use technology as a mediator to promote an increase in physical activity (as presented in the previous section), this thesis investigates the potentials to use a SAR as a motivational tool and is looking at the motivational effects of using robots in exercising scenarios.

Those effects could be behavioral measures such as how often or how long a person exercises or subjective impressions of the user measured with questionnaires assessing how much people like to exercise. The usage of robots as facilitator to motivate people in rehabilitation or cognitive tasks is not a brand new research question [cf. Fas12;

Sch12; Ley14b]. Moreover, as already mentioned above, SARs have been defined by Feil-Seifer et al. [Fei05] as robots that assist people just by their mere presence. Those robots could be used to coach and instruct people to do different exercises, to track one's progress, to remind on schedules, to give encouraging feedback, exercise jointly or to give recommendations. All of these would be suitable applications for SAR, and some research works already explored them. Kidd et al. [Kid08b] used a stationary social robot as a personal dialogue coach for weight management and as an exercising reminder. Fasola et al. [Fas12] used a SAR to coach elderly in a stationary arm raising task and Swift-Spong et al. [Swi16] used a SAR as a partner for circuit training. These works compared different behaviors, backstories or embodiments of the robot, but few investigated the benefits of having a SAR compared to a baseline (I will go into a deeper analysis regarding this issue in section 2.2). The lack of research showing that having a SAR is of advantage for the user compared to having no robotic exercising partner leads to the question whether SARs enhance people's motivation to exercise compared to not having an exercising partner.

RESEARCH QUESTION 1 *Is the Köhler Effect effect replicable with robot exercising partners?*

This question raises the matter of how the robot's presence generally could affect a human's exercising motivation. It is a fundamental question, but not thoroughly answered yet. On the first sight, one can argue that it seems quite evident that having a partner will increase exercising motivation. Looking at the research from Reeves et al. [Ree97] on the Media Equation it appears to be given that people treat media and technology as human. They showed that people treat media human-like by replicating theories from social psychology with computers as interaction partners. Thus, their results support that the Köhler Effect is replicable with SAR as partners. Though, it is not sure whether this is true and if the magnitude of the effect is the same with robot partners as with human partners. The implementation of an artificially created social artifact could also lead to varying side effects like discomfort due to the presence of the system. Furthermore, one can argue that it is a quite challenging engineering task to build a robot that is capable of exercising as humans do. It will be easier for many applications if the presence of the robot is sufficient for the human's motivation.

Thus, it is necessary to distinguish between different social role of the robot. The social role in the context of exercising could be either defined as an instructor role, where the robot is structuring a workout or with a partner role, where the robot is exercising with the human as a partner. In the rest of this thesis, robots with an instructor role are called Robot Instructors (RIs) and robots that have the role of a companion are named Robot Companions (RCs). Both have different advantages and disadvantages. On the one hand, a RC could lead to the same motivational effects as a human exercising partner would, but is harder to realize from an engineering perspective. On the other hand, a RI is less challenging to build and deploy, but the effectiveness concerning motivation might not be as high as with a companion. Hence, the next research question is:

RESEARCH QUESTION 2 *What are the effects of having a Robot Companion compared to a Robot Instructor on the users exercise performance and evaluation of the system?*

Regardless of the robot's social role, the deployment of such a SAR enables the possibility to motivate users and give them encouraging verbal feedback. Other researchers approached the question of feedback from robots (e.g. [Mid09; Swi15]), but often those works lack a baseline condition, and they instead compare the effects of different types of feedback. Equipping the types as mentioned earlier of robots with verbal encouraging feedback mechanisms results in two combination of social role and feedback. Throughout this thesis, robots that are companions and give feedback are named a Robot Companion with Feedback (RCF) and robots that are instructors and give feedback are called Robot Instructor with Feedback (RIF). Thus, the following question is:

RESEARCH QUESTION 3 *What are the motivational effects of encouraging feedback from a robot companion or instructor?*

One other important dimension for designing SAR is the embodiment of the system. From a cost-benefit ratio it would be easier to use and deploy Virtual Agent (VA) compared to SAR, because robots are harder to maintain and difficult to deploy. So far, no study exists that shows that embodied SARs enhance exercising performance. Fasola et al. [Fas13] tested different embodiments of a SAR, but found differences only on subjective evaluations of the user. Thus, it raises the question whether there are benefits of having an embodied SAR for exercising:

RESEARCH QUESTION 4 *How does the embodiment of a coaching system change the user's exercising motivation and perception of it?*

The second part of this thesis looks at adaptive processes for SAR. Previous works have investigated the effects of SAR for exercising in tasks (e.g. arm raising) that were suitable for elderly but might be not challenging enough for other populations. Therefore, preferences regarding the exercises of a SAR could provide may depend on the user population. Furthermore, individuals' personalities correlate with physical activity preferences [Rho06]. Thus, the adaptation of the system regarding a user's exercising preference is an essential requirement for future SAR. Therefore, one further question of this thesis is:

RESEARCH QUESTION 5 *How can a SAR learn online a user's exercising preference?*

This question raises concerns about how a system can actively learn a user's preferences without querying too many information or relying on prior knowledge. It needs a suitable learning framework that works online and is intuitively usable for the human. Besides, the embodiment of the recommendation system should not influence the user's agreement with the learned preferences. Finally, this opens the last research question of this thesis:

RESEARCH QUESTION 6 *Should the system or the user be in control of the exercising program?*

The system might not need to learn the user's preferences and adapts by itself. The user could also control the robot and choose the exercises she/he wants to do, which is equivalent to the Level of Automation (LoA) of the system. If the system adapts by itself and offers appropriate exercises, the system's LoA is high, while if the user controls the system the system's LoA is low. Following the theory of Epley et al. [Eplo7], a high level of self-control could lead to unexpected system behaviors for the user and thus could turn into higher associated levels of anthropomorphism of the robot. The increase of anthropomorphization could increase the perceived competence and trust in it, because it may then be a more believable coach and companion than just as a static computer program.

1.3 HYPOTHESES

From the research questions mentioned earlier, I derived four general hypotheses for this thesis.

HYPOTHESIS 1.1 *People exercise longer with a Robot Companion than with a Robot Instructor or exercising individually in isometric abdominal plank exercises.*

A proof of Hypothesis 1.1 verifies the Köhler Effect and can be a foundation for future investigations using social robots as exercising partners. Based on the results of the investigations on Hypothesis 1.1, following investigations on the manipulations of the robot's behavior can be compare the results against these baseline findings. Therefore, this thesis investigated the usage of encouraging feedback and looks at the motivation effect of encouragement. Therefore, I will test the following hypothesis:

HYPOTHESIS 1.2 *People persist longer in abdominal plank exercises if the Robot Companion gives encouraging feedback compared to a Robot Instructor or when exercising individually.*

In contrast to other works related to feedback (see [Fas12; Ham14; Swi15; Lei14]), encouragement feedback is a plain form of feedback that does not require any qualitative or quantitative performance evaluation. Finding evidence for 1.2 can unlock a relatively simple mechanism for motivating people to exercise longer.

As stated in the previous section, the embodiment is a crucial factor for SAR, and current research should further investigate the importance of embodiment on interaction and exercise times. Gao et al. [Gao15] showed in a meta-review that embodied co-present robots mostly have a positive influence on interaction time and persuasion, which is very important for exercising. Thus, I formulate the following hypothesis on embodiment:

HYPOTHESIS 1.3 *People exercise longer with Embodied Robot (ER) partners than with Virtual Agent partners.*

The previous hypotheses looked at the motivational effects of one type of exercising. However, future application of social robots for exercising should incorporate a diverse set of exercises. As stated in the research question section, finding the activity that suits the user's preferences might be an essential feature for future applications. Hence, I will look at how a human perceives an adaptive robot and investigate the following hypothesis:

HYPOTHESIS 1.4 *People perceive an adaptive robot coach as more trustful, competent and motivating than an adaptable robot coach.*

Particularly for companions to increase physical activity it is essential that users perceive SARs as trustworthy and competent because these attributes potentially influence the user's motivation to interact with the system repeatedly. Showing this association helps in guiding future research efforts in machine learning application for adaptive SAR.

To answer the research questions and hypotheses mentioned above this thesis will mainly use Bodyweight Exercises (BWEs) as a test scenario that the following section explains.

1.4 SCENARIO & SYSTEM

Our previous work has focused on the assistive capabilities of a SAR for long-term indoor cycling training [Süs14; Sch15]. This project provided first-hand experiences in designing robotic systems for long-term interaction and sports interventions with space missions as a use case. However, the constraints of the scenario specification (i.e., the assistance of robot on space stations with zero-gravity) limited the robot behavior to the role of an indoor cycling instructor in this scenario. Thus, its primary function was to instruct the user to follow an indoor cycling workout routine and accompanied these instructions with gestures to motivate the participants. In contrast to the previous project, this thesis focuses on a scenario where the embodiment of the robot can be fully exploited and is crucial for exercising co-actively. Hence, this thesis uses a BWE scenario that offers the possibility to test different exercises: dynamic and static exercises, exhausting and relaxing exercises, as well as cardio, strength and stretching exercises. BWEs offer the benefit that they usually do not need a long time to learn and most users have experience with some BWE from physical education classes. Moreover, BWEs do not require any other exercising machines and are suitable (nearly) everywhere.

The BWE scenario, including the control of the robot, the coordination of the interaction, as well as the perceptive and decision abilities, were designed using a framework developed for this thesis [Sch17a]. This framework allows to model socially assistive scenarios using a Domain Specific Language (DSL) and is used throughout this thesis to implement the different scenarios. Due to the limitation of this thesis, the focus will be on the results from HRI experiments, and I will not explain the implemented framework in detail. The following referenced figures are all listed Therefore, the interested reader should look at the related publication or Appendix A. However, a brief introduction is giving in the following.

The basic scenario (see Figure A.1.1) is composed of different so-called `movements` which describe one single exercise. The models of these exercises were implemented based on a previously derived interactive action-based motivation model (see Figure A.2.3). These exercises can be either static exercises (e.g., doing abdominal plank exercises, see Figure A.2.4) or cyclic exercises that are defined by a repeated movement (e.g., squats, see Figure A.2.5). Depending on the configuration of the decision system (see Figure A.1.2) for the different movements, state transitions, and dialog acts are triggered based on the received data using the in-house developed Robot Service Bus (RSB) [Wie11]. This framework allows to create complex coaching scenarios, but these kind of scenarios are not the focus of this thesis.

1.5 DEMARCATION TO OTHER WORKS

At this point, it is important to contrast this thesis to other research approaches that use SAR. The potential use cases are diverse, and various aspects are worth to investigate. Hence, it is crucial to set the limits of this thesis. One fundamental limitation is that this thesis is not going in the research direction of coaching robots. Coaching robots are used to instruct people and provide them with feedback and corrections while the user's is performing a task. These coaching abilities require real-time perceptive capabilities and feedback generation that are not within the scope of this thesis. This thesis investigates the effects of having an exercising partner and not the coaching abilities of a robotic partner. Thus, this thesis contributes to the understanding of the effect when working out with SAR in a simple setting and investigate how to implement a robot that is adaptive towards the user's preferences. In all of these scenarios, a coaching functionality would probably enhance the anthropomorphic effects, and the users' exercising motivation, but at this stage of the research, I am only interested in the isolated effects of the robot's embodiment, social role, encouragement, and adaptivity.

1.6 contributions

In this thesis I address some open challenges and research questions in the area of SAR. The following list states the contributions of my thesis:

1. A verification of the Köhler Effect with Socially Assistive Robot.
2. A verifaction of the motivational effects of encouragement from a SAR on a user's motivation.
3. A comparison of the effects of different embodiments on a user's exercising motivation.
4. Design and experimental testing of a Preference Learning (PL) approach based on Dueling Bandits for HRI scenarios.
5. A comparison how different adaptation strategies affect a user's evaluation of the system and interaction motivation.

1.7 OUTLINE

The rest of the thesis is structured as follows:

- Chapter 2 introduces the related work in SAR and theoretical background on motivation and exercising psychology.
- The first major part of this thesis investigates the motivational effects of exercising with SAR and is divided in four chapters:
  - Chapter 3 presents the results of an initial video Human-Robot Interaction (vHRI) study to gain insights in the user's perception of embodiment and companionship style.
  - Chapter 4 shows the results of a study comparing the motivational effects of exercising with a SAR.
  - Chapter 5 highlights the effects of giving encouraging feedback while exercising.
  - Chapter 6 compares the results from Chapter 4 and Chapter 5 with data from experiments with VA.
- The second major part of this thesis looks at the usage of adaptive SAR and is divided into two chapters:
  - Chapter 7 explains the usage of a preference learning model for HRI and evaluates it in a prototype study.
  - Chapter 8 presents a study on the effects of adaptability on the user's HRI experience.
- Chapter 9 and Chapter 10 discuss the results of this thesis and give a conclusion.

2 SOCIALLY ASSISTIVE ROBOTS

The previous chapter introduced the foundation for this thesis, stated the concern, the research question, hypotheses and briefly sketched published work in this area. This chapter will take a more in-depth look at the previous work on SAR for supporting and coaching people during exercises. However, it will not present every possible related work that is important for the following chapters on the embodiment, feedback or adaptation. The respective sections will discuss the relevant related work. At first, this chapter will give a general introduction to the field of SAR.

Feil-Seifer et al. [Fei05] introduced the term Socially Assistive Robot (SAR) as "the intersection of AR and SIR" [Fei05, p. 465]. Assistive Robotss (ARs) are hands-on robotic systems designed for rehabilitation of patients. Those robots are for example manipulators that rehabilitate post-stroke patients by assisting them in acquiring back their motor skills [Kah01]. Socially Interative Robots (SIR) are interactive systems for the study of social interactions between humans and robots or to study theories of social interaction and social behavior with robots [Fon03]. Thus, the term SAR specifies the intersection between both research fields: Using robots for rehabilitative exercises by the presence and social interaction provided by the robot without physical interaction (see Figure 1).

The benefits of using SARs are two-fold; there is a reduced concern for safety requirements when interacting hands-off with a robot, and the SAR approach resembles the style of a therapist that guides a patient through the rehabilitation process, coaches and encourages patients to use their limbs. Researchers recognize this kind of rehabilitation as a more effective than using assistive-robots, because patients tend to exercise longer and generalize acquired motor skills better [Win03]. Nevertheless, the usage of SAR is not limited to rehabilitation. Feil-Seifer et al. [Fei05] distinguish between different tasks in their definition of SAR. Thus, robots be used in educational tutoring scenarios, physical therapy, daily life assistance for elderly or as tools to learn about emotional expressions for children diagnosed with Ausistic Spectrum Disorder. Since this thesis focuses on the usage of social robots as companions for exercising and increasing physical activity (PA), the following section looks only at relevant work for exercising and coaching.

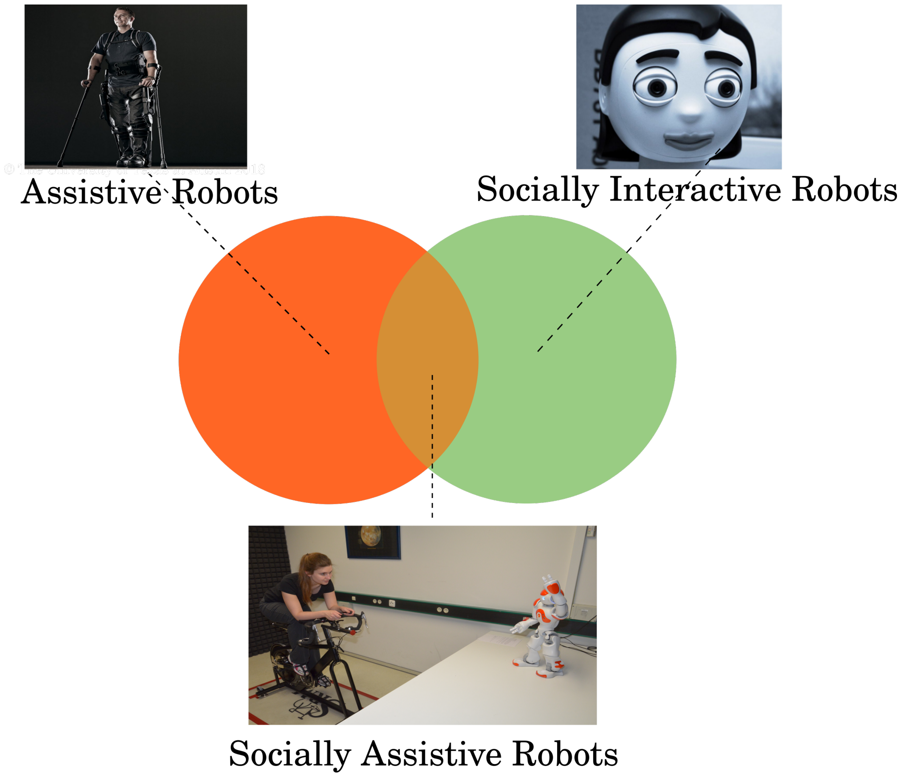


Figure 1: The field of SARs is defined as the intersection between assistive (rehabilitation) robots and socially interactive robots.

2.1 SOCIALLY ASSISTIVE ROBOTICS FOR EXERCISING

Since the introduction of the term SAR, it is increasingly used to describe the research that uses social robots that support people on cognitive, rehabilitative or educational tasks. Advancement in robotics and Artificiall Intelligence (AI) have facilitated an increase of publications in this field (see Figure 2). Part of this rise in publication numbers is because of new software to easily build distributed robotic system (e.g., Robot Operating System (ROS) [Qui09]), as well as cheaper and more widely accessible robot platform (e.g., Nao [Gou08]) with an Application Programming Interfaces (APIs) that allow fast prototyping and testing (e.g., NaoQi Software Development Kit (SDK), Choregraphe [Pot09]). These publications include robotic and computer science aspects, but also belong to an interdisciplinary research community intersecting with e.g., neuroscience, rehabilitation, geriatrics gerontology (see Figure 3).

This sections gives an overview of the current state of research in the field of SAR for exercising or motor rehabilitation. Table B.1 in the appendix gives an overview of the relevant publications from the last years. I analyzed the publications regarding the exercising task, measures used for the interaction, robot type, robot behavior, subjects, study design, and results. Other works concentrate on the design,

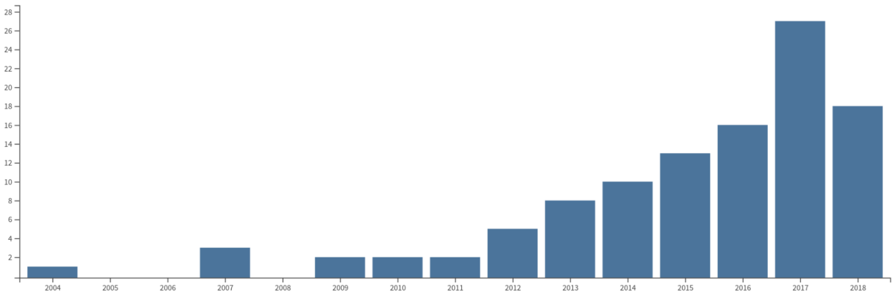


Figure 2: Number of publications per year for the keywords: `socially assistive robot`. Retrieved and created from `https://apps.webofknowledge.com` on June 15th 2018.

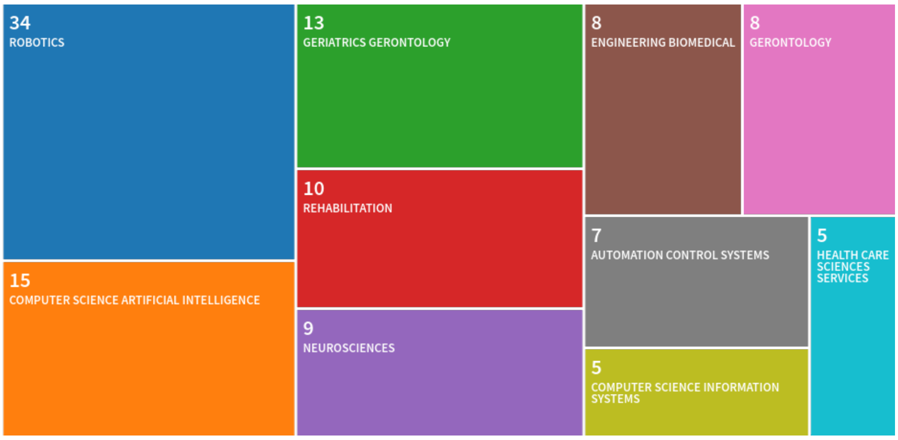


Figure 3: Category tags associated with the keyword `socially assistive robot`. Retrieved and created from `https://apps.webofknowledge.com` on June 15th 2018.

implementation and feasibility of a SAR system (e.g., [Vir13; Gör13]) and others focus more on the effects of SARs investigated in studies (e.g., [Lee14; Fas12; Pow03]). In the following, I will summarize the research findings and research methodologies of these works.

Powers et al. [Pow03] were one of the first to investigate the effects of having a robot as an exercise instructor. In their study, they tested the influence of different robot interaction styles (serious vs. playful) on the user's exercising compliance. They showed that the participants repeat a severe task like exercising more often when paired with a serious robot.

Subsequently, research in SARs was primarily driven by the Interaction Lab from the University of Souther California (USC). Eriksson et al. [Eri05] were one of the first researchers to use a robot with target subjects. They used a SAR to encourage post-stroke patients to use their stroke-affected limb to shelve books. They evaluated their system with six participants of which four were stroke survivors. They tested a system that uses sound effects as feedback, one that uses a synthesized voice and one with pre-recorded human voices. Additionally, they also altered the level of expressiveness of the robot by including robot movements in the human voice condition. Participants saw these conditions randomly. The authors do not provide any information about their questionnaires and interviews but conclude that the patients and physical therapists received the robot well.

Gockley et al. [Goc06] also worked on a hands-off mobile robot to encourage physical therapy compliance for stroke rehabilitation. The robot-assisted the participants while moving wooden pencils from one bin to another, lifting magazines or flipping through newspapers. Participants (n=11) in their study were students with a high understanding of robotics and technology. They tested different engagement and proxemic modes of the robot. Their results showed higher compliance when the participants thought that the robot was engaged. Tapus et al. [Tap07a] investigated user personality matching for SAR by varying the robot's proxemics and vocal features. Their results show that the users prefer a matching personality of the robot. The same group continued the work on SAR and studied the motivational effects of a SAR in different conditions [Fas12; Fas10]. They investigated the positive effects of praise, relational discourse and embodiment. However, these studies were missing a baseline condition which would allow comparing the robot against the user's intrinsic motivation to exercise.

Since then, the field of SAR for exercising or rehabilitation has gained more attraction, and several other research groups published papers on this topic. Gadde et al. [Gad11] worked on an interactive personal trainer for seniors to increase exercising adherence. They presented a feasibility study with 10 participants where the robot gave positive feedback while the user was doing seated arm exercises. However, they do not report any data about motivation or engagement.

Vircikova et al. [Vir13] worked on the usage of Nao as an instructor for spinal disorder rehabilitation. A qualitative analysis of the humanoid-child interaction showed that the children enjoyed the exercising, had no problem to repeat the exercise and wished to continue after the rehabilitation training had finished.

Görer et al. [Gör13] developed a system that learns a set of physical exercises from a professional coach and assists people in performing these gestures. They evaluated their system in one condition with 8 participants by tracking the correctness of the gestures and a subjective questionnaire. Participants reported high scores on immersion and positive affect and low scores on flow. However, it is difficult to interpret the results without having any conditions to compare to or a baseline condition.

Werner et al. [Wer13] studied the usage of a SAR for physical training support with older users. They evaluated the motivation after the demonstration of the training support and found that 70% of the participants think that the robot is 'very much' or 'a lot' motivating. The same amount of participants reported that a human trainer would be a better motivator than the robot. However, they can not conclude on their research question to what extent the system is more motivating than a video version. Fridin et al. [Fri14] investigated the effects of the embodiment of a SAR with experienced and inexperienced preschool children during an exercising task. They used qualitative measurements of eye contact and emotional reaction to investigate the effectiveness. Their results show that experienced children involved in the motor task in both conditions, but interacted less with the virtual agent. However, inexperienced children did not interact with the virtual representation at all. They conclude that embodied robots could be used during an initial phase, but can be replaced with virtual agents afterward.

Lee et al. [Lee14] looked at the effect on motivation regarding the type of robot operation (autonomous vs. tele-operated). The authors conclude that a tele-operated robot increased the competition between the user and thus increased motivation to exercise compared to an autonomous robot. The limitation of this study is the choice and number of exercise (i.e., one exercise: holding arms in front of the body), the sample size (20 participants, within-participants design) and a missing baseline condition where no robot is present.

Swift-Spong et al. [Swi15] compared the effects of self-comparative vs. other-comparative feedback to a control condition with no feedback in a push-button task. They hypothesized that comparative feedback conditions would produce higher self-efficacy and better performance. Moreover, participants will perceive the robot coach more positively in comparative feedback conditions. Though, the authors could not find evidence for any of their hypotheses.

Park et al. [Par16] investigated how social skills performed by a humanoid robot (e.g., mutual gaze, feedback and social distance) can enhance the social interaction in physical training in a study with two conditions (social skills vs. no social skills). They showed that such skills are useful social cues for physical training. However, they showed no link between the training engagement and the effectiveness of the training.

Swift-Spong et al. [Swi16] studied the effects of different backstories of a SAR. They designed a fictional and a real backstory for a robot that is exercising together with an adolescent. Their results show no differences between pre- and post-study on physical activity enjoyment and activity level.

Lotfi et al. [Lot17] introduced the term Exercise Trainer Socially Assistive Robot (ETSAR) as a solution for the demand of having human instructors for the rising elderly population. They included exercises recommended from the National Health Services (NHS) (i.e. sitting, strength, flexibility, and balance). The system instructs the users and gives them real-time feedback on their exercising success. They tested the feedback capabilities of their system but provided no results or explained the user population for their testing. They discussed that testers used their system for ten minutes and that the feedback was that the system's feedback was appropriate and timely.

Guneysu et al. [Gun17] presented another work on SAR for physical exercising for children. They implemented a system that tracks a child's arm movement in real-time and gave corrective feedback to the child during the exercises. They tested their system with 19 children without a control condition. Their results show that the children enjoyed the interaction and rated the robot as a useful exercise coach as an excellent social companion.

This presented list of works is not exhaustive and there probably many relevant publications missing. Also, I constrained the list to publications in the are of SAR with a focus on exercising or motor rehabilitation. Though, also the Human-Computer Interaction (HCI) community is working in the field of agents to rehabilitation and promotion of acrlongPA. Thus, interested readers could have a look at a recent review on Embodied Trainers (ETs) [Men17].

2.1.1 *Summary*

In the following I will summarize some of the main similarities and differences of the aforementioned studies.

Nao is one of the most used robot platforms in the reviewed publications [Gör13; Lew16; Fan16; Gun17; Swi16]. The second most used platoform is USC's robot Bandit [Tap07b; Tap08b; Fas10; Fas12; Fas13; Swi15].

The main exercise that have been used in the past research are different kinds of arm movements [Eri05; Goc06; Tap07b; Tap08b; Fas13; Gun17; Lee14; Gun17]. Others used a broader set of exercises ranging warm-up routines and to strengthening and cardio exercises [Lew16; Swi16]. Some included yoga exercises [Par16] and other used exercises recommended for rehabilitation [Vir13; Lot17].

The past works also show a variability in their used study population. Most subjects were in college-age, followed by elderly people and children. Elderly were studied as subjects in, e.g., [Fas12; Fas13; Lew16; Fan16] children or adolescent have been studied in, e.g., [Vir13; Fri14; Swi16].

However, the most research has been conducted with participants in a college-age level (e.g., [Goc06; Pow03; Eri05; Gör13; Par16; Kas17]).

Most studies used the robot as an instructor or coach, that is presenting some kinds of exercises and asks the participants to repeat them. Therefore, the supportive behavior of the robot was mostly verbal feedback and guidance through the exercises. Hence, one of the most used measurements for the presented systems is the arm position and angles of the user. Thus, researchers infer about the participant's compliance in the interaction by measuring whether the participants follow the instructions. Additionally, those objective measures are accompanied by subjective impressions assessed using questionnaires and post-study interviews.

In general, the works show that the usage of SARs in exercising situations is feasible. The found results report that participants from all study populations enjoy to interact with the robot and also that nurses are enthusiastic about the usage of SARs(e.g., [Vir13; Fan16; Lew16]). Besides these promising results about the usage of robots for exercising applications, there are still some open issues.

2.1.2 *Open Issues*

The growing numbers of publications in the field of SARs for exercising applications is a sign that there is a demand for such kind of research and robots for applications to increase lay peoples' sedentary time. However, this field of research is comparably new to other fields in robotics, and therefore many open questions need to be addressed and answered. From analyzing the works mentioned above, I conclude that there are some research issues that future work should target. Those issues are a lack of studies with a control-based study design to conduct statistical inferential tests about the effictiveness of SAR, a lack of studies looking at the quantitative motivational effects of embodiment/feedback of SAR and a lack of studies with different and strenuous exercises. In the following, I will briefly describe these lacks and relate them to the publications above.

SAMPLE SIZE AND STUDY DESIGN Many of the previous works do not compare different conditions in their studies (e.g. [Eri05; Gad11; Wer13; Vir13; Lot17; Gun17; Gör13]). Results are often presented descriptively, which makes an evaluation and interpretation of the results challenging. Though feasibility studies of designed systems are valuable, also more studies are required that contribute to an understanding of the motivational aspects of using SARs. Besides lacking conditions, there is also a lack of a sufficient number of participants in the evaluation and sometimes missing demographic information. This evidence hints that the knowledge about the effectiveness of using robots as a social exercising partner lacks a proper evaluation with sufficiently naive participants and conditions that allow inferential statistics.

QUANTITATIVE MOTIVATIONAL EFFECTS OF FEEDBACK One of the main advantages of using SAR is that they can give the user instantaneous feedback about the user's exercising quality and progress. However, missing control conditions in the study design also affect the knowledge we have about the value of robot feedback. Regarding the studies that included different conditions in their research, the primary comparison is between the verbal assistance and feedback they get from the robot (e.g. [Fas12; Swi15]). Thus, researchers are often comparing different types of feedback from the robot (e.g., comparative, relational), but are missing a baseline condition to compare the results. Thus, no previous research looked at feedback as an isolated concept and the benefits of feedback are still an open issue. Additionally, other works that include feedback (e.g., [Gad11; Gun17]), are missing a baseline comparison which shows that exercising with a robot is quantitatively excelling to working out alone.

EXERCISE INTENSITY, VARIATION AND TIME Many of the works focus on light to moderate physical activity, which is mostly due to the focus of rehabilitation for children and elderly. In most cases the work concentrated on exercises like arm movements (e.g., [Fri14; Gad11; Gun17; Kas17; Fas13]). While these are essential exercises for post-stroke rehabilitation or as exercises for elderly while seated, these exercises are not sufficient as a regular exercise activity for non-elderly and healthy persons. Few works investigated the usage of robots in intensive exercising activities. Hence, this thesis will examine whether a robot exercising companion is capable of pushing participants to their exercise limits during a full body work out. To study this question, this thesis looks at the motivational effects if the robot is instructing the users to do different exercises, but also how it influences the motivation when the user and the robot are exercising together.

Additionally, few of the works include a set of different exercising activities. Among them is the work of [Swi16]. In their study, they used a cardio exercise, step up training as well as muscle strengthening activities. They compared different background stories of the robot (i.e. real vs. fictional story) during a four-session in-between subject design study. However, from their results, they cannot draw any conclusion about the effectiveness of the workout with the robot based on the background conditions.

Furthermore, the studies do not show whether the participants like to engage with the robot because they also like the exercises. Considering exercise preference is essential this because research from exercising psychology showed that people's personalities result in different preferences for physical activity types. Hence, this thesis looks at how to build a system that is adaptable to the user's choice using their feedback during the interaction. This adaptivity results in a diverse HRI experience which is unique for every user. The introduction of adaptive capabilities brings the challenge of autonomy into this thesis. There are several ways of how to incorporate a system with a variety of exercising opportunities. The user can control the system or the system uses the user's feedback and adapts autonomously. So far, there is no research yet that investigates what effects the autonomy of the exercising companion has on the user.

In summary, the past researchers let's fundamental issues for the usage of SAR in exercising scenarios open. What are the motivation effects of exercising together with a robot? How does feedback from the robot influence my exercising motivation? And, how can the robot adapt to one's preferences? These are some of the questions I will try to answer with my thesis.

Though, one of the most important concepts here, is the term motivation. If the robot is designed to engage the user in an exercising activity, it means that the robot should try to influence the user's motivation. Therefore, the next section will briefly introduce some of the psychological background.

2.2 PSYCHOLOGICAL BACKGROUND

SARs are designed to help people on a task by their presence, guidance or feedback. Thus, previous works have dealt with the concept of motivation in their works. Motivation is defined as the intrinsic determination toward goal attainment [Pl000]. In contrast, the feedback and encouragement from family, friends or professionals are seen as social support [Kin92]. Additionally to those external support measures, approaches to modify people's lifestyle (e.g., sedentary time) should target a person's internal motivation, because it is one of the significant determinants of exercising adherence [Dis80]. Hence, one question for researchers in SAR could be, how to enhance a user's internal motivation by providing social support. Therefore, a brief look at different theories on motivation are presented in the following sections. This list of theories is not inclusive or complete.

2.2.1 *Social Cognitive Theory*

The Social Cognitive Theory proposes that motivation is a cognitive function and that self-efficacy is the primary mediator for behavior change [Ban77]. Bandura [Ban77] define self-efficacy as the belief in one's competence for a given task. Thus, it determines how people think, behave and feel. One predictor for the observation that people start and exercising and then quit early is that people do not believe that they can ever accomplish a task or become sufficiently good at it, which is having a low self-efficacy belief. For an uncomfortable task like starting an exercise regimen and becoming more physically active, it is essential to look at the . People with low believe that they are beyond their skills and lose their confidence in their ability. Thus, motivation is concerned with the initiation and the maintenance of a behavior to increase one's belief about the own capabilities.

Swift-Spong et al. [Swi16] investigated whether self-efficacy increases when a robot provides comparative feedback. However, they could not find evidence for their hypothesis.

2.2.2 *Self-Determination Theory*

Ryan et al. [Rya00] conceptualized the Self-Determination Theory (SDT) and distinguished between "different types of motivation based on the different reasons or goals that give rise to an action" [Rya00, p. 55]. They distinguish between intrinsic motivation and extrinsic motivation.

INTRINSIC MOTIVATION Intrinsic motivation is a process to challenges oneself on new tasks, become aware of one's limitation and skills, to observe and increase one's knowledge [Rya00]. One's interest and enjoyment is the primary drive to engage in such tasks. Thus, it does not rely on any external pressure. One's self-determination, observed improvement, and competence on a task modulate this motivation. One can further divide intrinsic motivation into intrinsic process motivation which means that someone is doing a duty because of enjoying to do the work, and internal self-conception, referring to behavioral change based on personal values and standards.

EXTRINSIC MOTIVATION Extrinsic motivation is rooted in influence from an external source [Rya00]. Most often, extrinsic motivation is induced either by a reward (such as monetary compensation or marks) or by a threat of punishment (e.g., doing extra hours, bad grades, more chores). Regarding the application for coaching and exercising, sports competitions are also an extrinsic motivator since competitors are spurred to win and beat the other participants.

One can further divide extrinsic motivation into integration, identification, introjection and external regulation. People perceive external regulations as controlled and regulated. The motivation to show a particular behavior in these situations is to obtain an external reward or satisfy others. Introjected regulation refers to behaviors to avoid guilt or anxiety. People will act because of a feeling of pressure. In regulation through identification, if a person identifies with the importance of behavior, she or he will accept the regulations associated with it. Ryan et al. [Rya00] define integrated regulation as the most free form of extrinsic motivation. In this form of motivation, people entirely assimilate regulations with self, and one includes external regulations in one's self-evaluation and beliefs.

Most of the works in the SAR domain use the concept of extrinsic motivation, because the interactions primarily used feedback from the robot as an external encouragement for the user (e.g., [Gun17; Par16; Lot17; Gör13; Gad11; Eri05]). Still, at the current stage, it is also likely that the motivation to exercise or interact with a robot is intrinsically motivated because most people have no prior experience in interacting with robots. This motivation and interest, to interact with new technologies, is known as the novelty effect.

2.2.3 *Flow*

Related to intrinsic motivation and self-efficacy belief is the concept of flow [Csi00]. If one is experiencing flow, one is fully immersed and involved in an activity and feels enjoyment in the process of this activity. People that are experiencing flow report to lose one's sense of space and time. To have a mental state of flow requires to have an activity with clear goals, immediate feedback and a balance between the perceived challenge of the task and the own perceived skill in doing the task [Csi00]. This state may also appear in challenging activities such as in sports and exercising fields. Thus, the concept of flow might be an applicable concept well-suited for SAR scenarios in the context of exercising. The system could change the task difficulty to match a person's ability and thus create an optimal challenge. Fasola et al. [Fas13] presented in their work a system that uses the concept of flow by providing a variety of challenging exercise games with different difficulties.

2.2.4 *Group Dynamics*

Motivation is also influenced by being part of a group [Web07]. The Köhler Effect is a phenomenon observable when a person increases his/her efforts on a task as a member of a group compared to when being alone [Ker11; Irw12]. Kohler [Koh26] found that a motivational effect appears in conjunctive tasks where the group success is dependent on the individual effort of each team member (e.g., mountain-climbing, rowing). The less-capable member of the group tends to show extra effort in such tasks. The Köhler Effect roots in social comparisons and the impact of an individual being indispensable to the group. One effort boost can arise from the understanding that others are performing better than oneself which leads that a person sets higher goals for better comparison with others. The other effort enhancement stems from the fact to know that a group is depending on one's performance. This motivation gain is most significant when members' abilities are only moderately different (versus about the same or very different) due mostly to social-comparison [Koh26]. If others are much more capable, then the comparer will stop comparing himself because it seems like an unachievable goal to compete or match the other. None of the reviewed literature explicitly models the robot as a group member in a conjunctive task. In most cases, the robot is designed as an instructor that gives feedback or as a coach that shows exercises and asks the user to imitate the exercises. However, there are exceptions like the imitation game presented in [Fas13] in which the user is presenting some arm movements, and the robot needs to imitate the motions. Nevertheless, none of the works included a scenario where the robot and human are exercising in a team-based fashion.

2.2.5 *Summary*

In conclusion, all of the above mentioned theories could be utilized for the design of Socially Assistive Robot scenarios. Robots could be designed to give users feedback during an exercise. This might contribute to a higher compliance during tasks and motivate them to exercise longer. In turn, this increase in exercising time may also increase one's , which could lead to an increase of intrinsic motivation over time. The systems could chose exercises for the user that aim at creating an optimal challenge so that the user experiences a state of flow. Finally, engineers should design scenarios where the robot and the user are exercising in a conjunctive task, which could further increase exercising motivation. Thus, all of the above mentioned approaches could facilitate the theories behind that increase exercising motivation.

In the course of this study I will first investigate the usage of Socially Assistive Robots in a conjunctive task, thus proving that the Köhler Effect will increase an individuals effort. Following, I will look at the effects of extrinsic motivational encouragement and investigate whether this kind of motivation further increases one's exercising time. Finally, I will look how a system might learn a user's exercising preference, so that the systems learn which exercises a user likes.

3 EFFECTS OF A ROBOT'S EMBODIMENT AND SOCIAL ROLE IN VIDEO HUMAN-ROBOT INTERACTION

As an empirical starting point for the investigations in this thesis I target video Human-Robot Interaction (vHRI) as an easy to use research tool to get first user impressions on a SAR. Using vHRI offers several advantages: It allows to gather many data in a short period, it helps to prototype and refine the research questions, it is easy to use and to deploy. However, it has also some drawbacks: Participants in vHRI studies are watching videos of recorded interactions only . One can not assure that the participants are observing the videos attentively. Hence, it is necessary to backup vHRI studies with real HRI studies afterwards. This chapter investigates whether it is possible to gather user ratings regarding an exercising companion depending on the social role and its embodiment in a vHRI.

NOTE: Parts of this chapter were published as a late-breaking report in S. Schneider et al. "Does the User's Evaluation of a Socially Assistive Robot Change Based on Presence and Companionship Type?" In: *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. ACM. 2017, pp. 277–278.

3.1 INTRODUCTION

Using Active Video Games (AVGs) to promote rehabilitation and physical activity has been a recent trend, and AVGs are used in addition to standard methods of rehabilitation or physical education. The expectations are that people, and especially kids, will be more motivated to exercise if it is perceived more like play than a workout because they "capitalize on children's natural interest in computerized video interaction" [Gao15, p. 465]. However, the long-term effects of exergaming with AVGs are at the same time disputable [Zen17].

As stated in the introduction, Li [Li15] found evidence that the agent's embodiment affects the persuasiveness of the system. Thus, the question of whether AVGs and Virtual Agents (VAs) are sufficient to motivate people during tasks like exercising and rehabilitation is essential to answer. Furthermore, it is necessary to investigate whether there it is important for a Socially Assistive Robot (SAR) to have physical representation. Therefore, it is yet unclear how a lack of embodiment of an agent (e.g., exergames, VA) might influence interaction and training motivation. Is it that being physically present in one's personal space is such a strong social factor which would additionally boost motivation to workout? Then the conclusion would be that a motivation gain is achievable by the deployment of a social artifact (e.g. a social robot) in the surrounding of a person. Still, it is not known which aspects influence that physically present robots might be more motivating. Is it because the users perceive them as more anthropomorphic and thus increase their belief of exercising with human-like partner?

Moreover, deploying robots also brings new challenges. Building robots that are capable of exercising conjunctively with a human partner is a challenging engineering task and an expensive one. Thus, in the context of exercising it is important to consider whether it is sufficient for a SAR to instruct the users and give them feedback (i.e., to be a Robot Instructor (RI)) during the physical activity or to exercise with them conjunctively (i.e., to be a Robot Companion (RC)). Results will lead to different requirements for engineers developing new robot platforms. Do future social robots need to be capable of exercising along with their interaction partner or not? Therefore, I will also have a look at how the social role of a robot influences the user's perception of the robot in this chapter.

As a starting point to answer the questions of this thesis, this chapter presents a video Human-Robot Interaction (vHRI) study that gathers people's evaluation of different embodiment and social roles. It presents a 2 (embodiment) x 2 (social role) between study design with following conditions: a remote-located Robot Instructor (RRI), a remote-located Robot Companion (RRC), a co-located Robot Instructor (CRI) and a co-located Robot Companion (CRC). For each condition, I recorded a short video of a robot interacting with a human during a sports exercise and asked the participants to view the video and rate their perception of the robot using the Godspeed questionnaire [Bar09].

This chapter is structured as follows: The next section gives an overview of current research in the field of physical embodiments of the robots, appearance and social role. Section 3.3 introduces the study design and explains the video material used for the evaluation. Section 3.4 shows the results and the subsequent section discusses the results.

3.2 RELATED WORK ON EMBODIMENT AND SOCIAL ROLE

The role of embodiment of social robots has been investigated by various researchers having different research questions, various types of robots and different methodologies. Also, the influence of the social role of the robot has been investigated. This section summarizes different research results using SAR.

From a perspective of artificial consciousness the embodiment has always been a crucial factor. The Embodied Cognition Theory claims that for the development of artificial consciousness having a real body is a necessary condition. However, what is the actual meaning of embodiment? In general terms, it has been defined as the "basis for mutual perturbation between system and environment" [Dau99a]. Therefore in this work I focus on robots that are made from real world material having motors and actuators so that they possibly can perturb their environment. Hence, robots can be either embodied and co-present or embodied and remote-located (i.e. mediated through a video of a real robot). These types of embodiments are in contrast to virtual agents, which have no physical representation in the real world.

One crucial question is how the presence and embodiment changes the user's perception of the social agent. Is a co-located robot more beneficial than a remote-located real robot or VA? A meta-review on recent papers has tried to answer this question [Li15]. The results show no difference between a VA and a remote-located robot. However, a co-located robot was found to be in favor compared to a VA and a remote-located real robot. Therefore, I exclude in my work a condition with a virtual representation of the robot and focus on a remote-located embodied robot and a co-present embodied robot.

3.2.1 *Embodiment of SAR*

Regarding the embodiment and presence of SAR several studies have been conducted. Leyzberg et al. [Ley12] studied the effects of the presence of a SAR during cognitive tasks. Their results show that the embodiment of the robot improved the learner's gain. Fasola et al. [Fas13] investigated the effects of a VA coach versus a SAR. In a long-term study they investigated how the user engagement changes based on the embodiment of the robot (virtual vs. present). The results show that participants are engaged longer in the training and are more satisfied with the rehabilitation. The difference between a co-located, remote-located and simulated robot has been studied in [Wai07]. The results show that the users find the co-located robot more appealing than the remote-located and simulated robot. Fridin et al. [Fri14] investigated the effects of the embodiment of a SAR with experienced and inexperienced preschool children during an exercising task. Their results show that experienced children involved in the motor task in both conditions, but interacted less with the virtual agent. However, inexperienced children did not interact with the virtual representation at all. They conclude that embodied robots can be used during an initial phase, but can be replaced with virtual agents afterward.

3.2.2 *Social Role*

The social role can be interpreted and modeled in different terms. The role can be constructed as the robot's modeled personality (e.g. serious or playful, extroverted or introverted), as a companion or instructor, or it can have different ascribed gender roles.

[Pow03] studied the user's compliance to follow the instructions of a robot based on its demeanor. They found that if the behavior of the robot matches the seriousness of the task, people are more willing to comply with the robot. The matching of the personality of the robot and the user has been studied in [Lee06]. They concluded that users enjoyed the interaction with a robot when the robots' personality was complementary to their own. The same was found for the user's rating of the robot's intelligence and social attraction. Walters et al. [Wal11] also investigated the effects of an introvert and extrovert robot using a vHRI paradigm. They found that people prefer the extrovert robot and associate traits like intelligence, interest, friendliness, and diversity with it. Kuchenbrandt et al. [Kuc12] investigated the impact of gender typicality on the user's performance during Human-Robot Interaction (HRI). They found that participants were less willing to accept help from a robot in the context of a typically female work domain, independent of robot and participant gender. When instructing on a male task, the female robot was perceived as more competent by the participants. The trustworthiness of a robot based on different behaviors and performances have been studied in a vHRI study in [Bru14]. They showed that the motion fluency and the task performance of the robot have an impact on the trustworthiness.

3.2.3 *Research Questions and Hypotheses*

The related work suggests that there is evidence for positive effects of embodiment. Therefore, I ask whether people perceive robots that are co-located (embodied) on a video differently than when the robot is remote-located (telepresent) on a video. Also, the social role, personality, or gender seem to have different effects on people's perception of a robot. Hence, I propose that people's perception of robots is different depending on the social role. Are people perceiving a robot that is doing a task conjunctively (companion) with a human on a video differently than when the robot is only instructing (instructor)?

On the basis of the previous work from Li [Li15] and Feltz et al. [Fel14], I hypothesize that

HYPOTHESIS 3.1 *A co-located robot will be evaluated higher on all scales compared to the remote-located robot.*

HYPOTHESIS 3.2 *A co-located robot companion will be evaluated higher on the likability and animacy scales compared to all other conditions.*

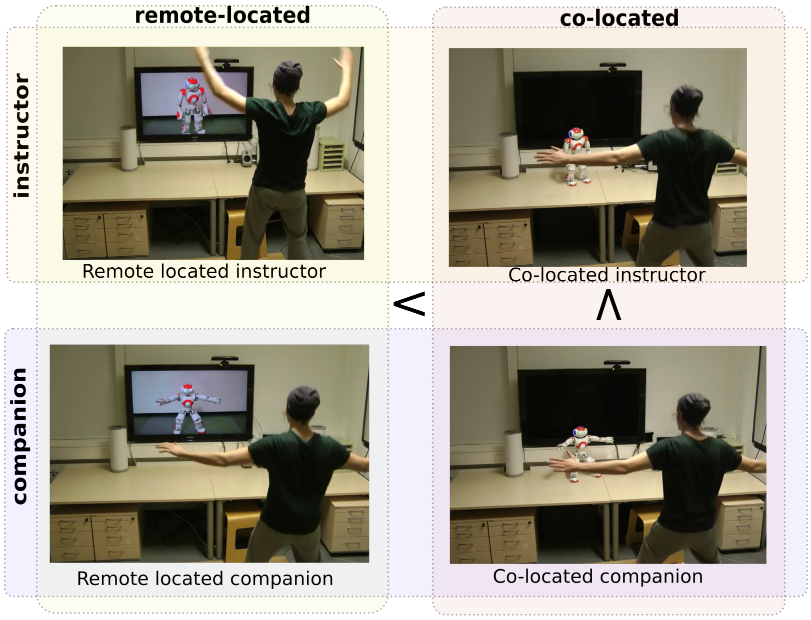


Figure 4: The stimulus material for the different conditions of the video Human-Robot Interaction study. Arrow indicate the direction of the hypotheses.

3.3 STUDY DESIGN

To investigate the hypotheses and research questions, I conducted a vHRI study. Differences in the attitudes people have towards robots based on watching them in a video might also be transferable to actual HRI. A vHRI study has two advantages: They can be conducted as an online study, which is a fast and easy way of prototyping HRI [Bru14]. The usage of this research paradigms is backed by recently successfully works to study social vHRI [Ho10; Bru14; Wal11].

3.3.1 *Video Material and Script*

I recorded four videos that always show the same robot behavior. Only the embodiment of the robot (co-located vs. remote-located) and the robot's social role were manipulated. The videos show the humanoid robot Nao as a robot interaction partner. The motion of the robot was animated using the provided Choregraphe software. Each video starts with a short introduction of the exercises. The exercises were jumping jacks, side lunges, and squats. The robot first introduced each of these exercises, while a human interaction partner was standing in front of the robot observing it. Afterward, the screen faded out and the robot instructed the user to do the exercises either together with the robot or independently respectively to the condition. The video always showed three repetitions of each exercise before fading out and continuing with the next exercise. The order of the exercises was jumping jacks, side lunges, and squats. After the last exercise, the robot thanked the user for the participation and said goodbye. Figure 4 depicts some example screenshots of the interaction. For the remote-located condition, the same behavior of the robot was pre-recorded and displayed on a screen in front of the HRI partner. The language of the robot was in English and subtitles were provided to counteract a low speech recording quality.

3.3.2 *Conditions*

CO-LOCATED ROBOT INSTRUCTOR The co-located Robot Instructor (CRI) represents the condition where the social role is of a Robot Instructor and the embodiment is co-located, i.e. present in the same room.

CO-LOCATED ROBOT COMPANION The co-located Robot Companion (CRC) represents the condition where the social role is of a Robot Companion and the embodiment is co-located, i.e. present in the same room.

REMOTE-LOCATED ROBOT INSTRUCTOR The remote-located Robot Instructor (RRI) represents the condition where the social role is of a Robot Instructor and the embodiment is remote-located, i.e. a virtual/telepresent representation of the robot.

REMOTE-LOCATED ROBOT COMPANION The remote-located Robot Companion (RRC) represents the condition where the social role is of a Robot Companion and the embodiment is remote-located, i.e. a virtual/telepresent representation of the robot.

3.3.3 *Measures*

DEMOGRAPHIC MEASUREMENTS The participant's age, sex, as well as their weekly exercising activity were assessed on a 6-point Likert-scale (e.g. 1: less than one hour; 6: more than eight hours per week).

PRIOR KNOWLEDGE Participants rated how much experience they have with computer, programming, robots, speech-interfaces on a 5-point Likert-scale (e.g. 1: no experience; 5: a lot of experience).

PERCEPTION OF THE PARTNER The Godspeed questionnaire was used to assess the user's perception of the robot [Bar09]. This questionnaire measures the perceived intelligence, likability, anthropomorphism, animacy and safety of robots on a 5-point differential scale (see Appendix D.1 for the used scale). However, I will not use the subscale perceived safety in this thesis, because this thesis is not focussing on safety issues when interacting with robots.

wish for having an assistive system Participants were asked whether they like to a) use the presented system in the future, b) have a system that assists them during sport, c) have a system that motivates them during sport on a 5-point Likert-scale.

role of the interaction partner At last the participants were asked to rate which role they would ascribe to the robot. The range goes from machine, toy, useless technology to partner, trainer, teacher.

3.3.4 *Participants*

Participants were acquired through online social media and on the university campus of Bielefeld. They were randomly assigned to one of the four conditions. In total 90 participants with mean age M=27.28 and SD=9.8 (male: 38, female: 52) took part in this study. However, 6 participants were excluded, because they did not finish the survey.

3.3.5 *Statistics Used in This Thesis*

From this point on in this thesis, I gather data and use them to find answers for my research questions and pieces of evidence for the hypotheses. Thus, I will not only present data in box plots or means (M) and standard deviations (SD), but I also test my hypotheses using inferential statistics. Gathering data from experiments is not only extensive (especially in a one-person project), but also prone to errors. In some cases, the data does not meet the requirements for specific statistical tests. Thus, I am not only drawing from the pool of parametric test, but I will also use a non-parametric test in cases of violation of the assumptions for parametric tests. Additionally, I will report the Cronbach's α for the used scales as a measure for internal consistency [Cro51]. Throughout this thesis, I tested the data for normality assumptions (i.e., I used qq-plots to look at the data, tested for homogeneity of variance with Levene's test and conducted Shapiro-Wilk test (SW-Test) to test for normality [Lev61; Sha65]). In cases where the criteria for parametric tests are satisfied, I used Analysis of Variance (ANOVA) and Welch's two-sample t-tests, in other cases I used Kruskal-Wallis Rank Sum Test (KW-Test) and Wilcoxon Rank Sum Test (WC-Test) [Fis21; Wel47; Kru52; Wil45].

I will complement the general statistics of the test and p-values with a measure of the effect size. Though, I am only presenting the effect size when it is valuable to report them. Therefore, I will, in most cases, not report the effect size when I check the data for a successful manipulation. In all other cases, I will report one of the following effect sizes. I use Cohen's d as a measure for effect size of Welch's two-sample t-test [Coh88]. For the Wilcoxon Rank Sum Test I will report r as a measure of effect size [Ros94]. For ANOVAs I use η_p^2 as a measure of effect size. Additionally, I will also report ω^2 as an effect size for ANOVAs when there is a significant main effect. However, I will not report an effect size measure for manipulation check measurements. For the Kruskal-Wallis Rank Sum Test, I will not report any effect size due to the advice of Field et al. [Fie12]. Omnibus test (e.g., ANOVA, KW-Test) are followed by pair-wise comparisons using t-tests or Wilcox tests with p-value adjustments correcting for the number of comparisons following the method of Bonferroni-Holm [Sie88; Hol79]. Data were analyzed using the statistical programming software R [R C13]. Statistical significance in figures or tables is depicted as follows: * : $p < .05$; ** $p < .01$; *** $p < .001$.

3.4 RESULTS

A one-way ANOVA between the groups for prior knowledge ($F(3,82) = 0.4$, $p = .98$) and exercising per week ($F(3,82) = 0.47$, $p = .70$) showed no significant differences. However, an ANOVA revealed a significant difference for the factor age, $F(3,82) = 1.18$, $p < .18$). Still, pairwise comparisons using t-tests with non-pooled SD did not show any significant difference between the conditions for age (all $p > .1$). Hence, I conclude that the randomization was successful.

I conducted several one-way ANOVAs or KW-Tests to find differences in the perception of the robot based on the Godspeed questionnaire. The internal consistency of the item sets is as follows: intelligence ($\alpha = .81$), anthropomorphism ($\alpha = .76$), animacy ($\alpha = .86$), likability ($\alpha = .9$). Thus, the internal consistency ranges from good to satisfactory. Figure 5 shows the results of the Godspeed questionnaire.

There was no significant effect found on perceived intelligence ($F(3, 82) = .83$, $p = .48$, $\eta_p^2 = .01$), anthropomorphism ($F(3, 82) = .30$, $p = .82$, $\eta_p^2 = .01$), animacy $H(3) = .56$, $p = .90$, and likeability ($F(3,82) = .5$, $p = .68$, $\eta_p^2 = .01$).

Also, two-way ANOVAs for social role and embodiment showed no significant main or interaction effects. There was no significant main effect of the social role on perceived anthropomorphism ($F(1, 82) = .00$, $p = .94$, $\eta_p^2 = .00$) and embodiment ($F(1, 82) = .55$, $p = .45$, $\eta_p^2 = .00$). There was also no interaction effect, $F(1, 82) = .11$, $p = .55$, $\eta_p^2 = .00$. There was no significant main effect of the social role on perceived likability ($F(1, 82) = .38$, $p = .53$, $\eta_p^2 = .00$) and embodiment ($F(1, 82) = .00$, $p = .94$, $\eta_p^2 = .00$), as well as no interaction effect, $F(1, 82) = 1.10$, $p = .30$, $\eta_p^2 = .01$. There was also no significant main effect of the social role on perceived intelligence ($F(1, 82) = .42$, p

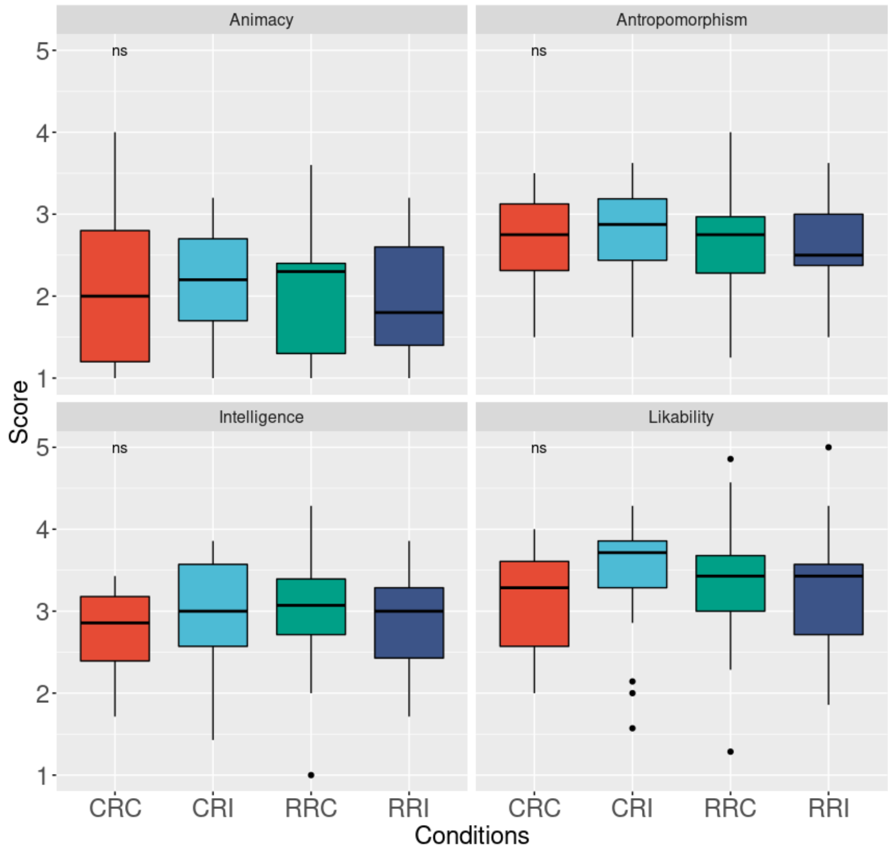


Figure 5: Boxplots depicting the results of the Godspeed questionnaire. ANOVAs and KW-Tests show no significant differences on all subscales between the four conditions remote-located Robot Companion (RRC), remote-located Robot Instructor (RRI), co-located Robot Companion (CRC) and co-located Robot Instructor (CRI) in the vHRI study.

$= .51$, $\eta_p^2 = .00$), embodiment ($F(1, 82) = .09$, $p = .76$, $\eta_p^2 = .00$) and no interaction effect, $F(1, 82) = 1.98$, $p = .16$, $\eta_p^2 = .02$. Finally, there was no significant main effect of the social role on perceived animacy ($H(1) = .03$, $p = .84$) and embodiment ($H(1) = .30$, $p = .58$). There was also no significant interaction effect as determined by pairs of post-hoc wilcox tests.

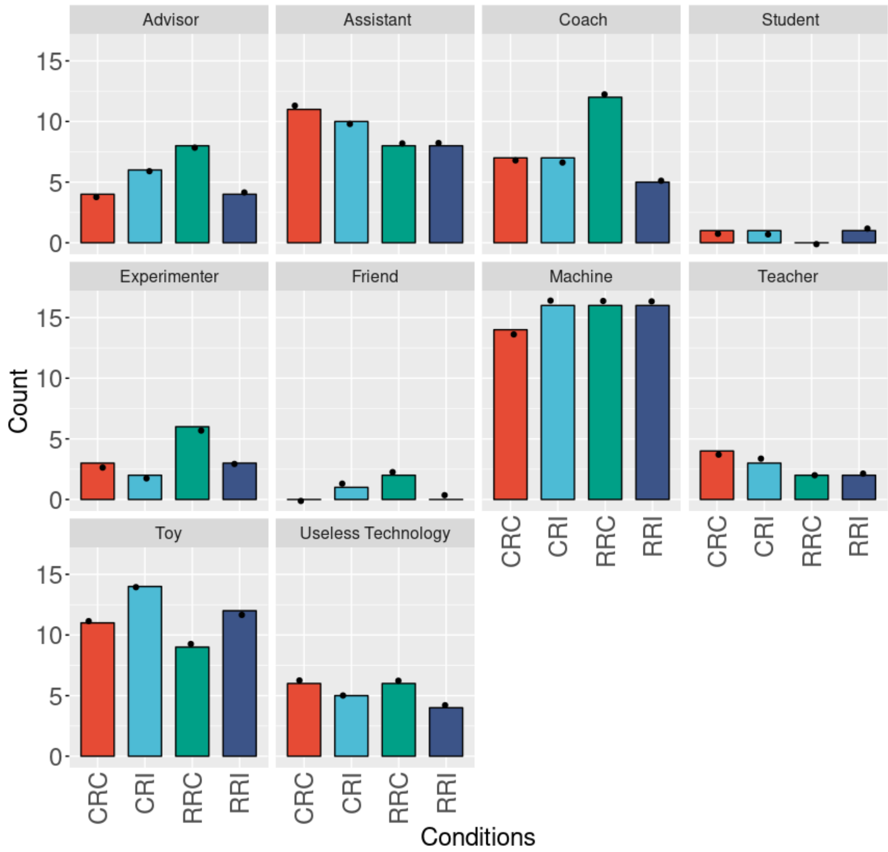


Figure 6: The perceived social role for the different conditions. The ratings were not statistically different across conditions. The robot was mainly perceived as a machine and toy. Particiapnts in the RRC condition associated the agent more often with the role of a coach.

Also no differences in the wish for future assistance were found, $F(3, 82) = .61$, $p = .6$, $\eta_p^2 = .02$. The rating for the perceived social role of the system is shown in Figure 6. A χ^2 analysis revealed no differences between the conditions, all $p > .1$.

3.5 DISCUSSION

The primary objective of this chapter was to investigate whether varying embodiments and social roles of a robot in a vHRI affects people's perception of the robot.

The results do not support any of the presented hypotheses. Neither the co-located robot condition was evaluated higher compared to the remote-located condition, nor the co-located robot companion was evaluated higher on the likability and animacy scales compared to the co-located robot instructor. Regarding the embodiment of the robot, the findings are opposite to studies which have found a positive effect of embodied robots [Ley12; Wai07]. However, those studies focused on the benefits of task success and not on the ratings of the robot. Regarding the social role, the findings are also opposite to studies which have found differences due to the role [Pow03; Lee06; Kuc12]. Also, in contrast to the studies that used the same methodology, this study revealed no differences in the presented conditions [Wal11; Bru14]

Hence, I propose explanations why there is no support for the hypotheses and discuss some confounding factors. The effect of the embodiment and social role of the robot is probably hindered by a) the short length of the video, b) the fact that the size of the robot is subjectively bigger in the telepresent conditions, c) and that the appearance of both robots is too similar.

First, the duration of the video might have been too short to account for any differences in the perception of the robot. The videos have a duration of 2 minutes with five different scenes (introduction, three exercises, farewell). This duration might be too few time for the participants to establish some perception and thought about the observed robot behavior or embodiment.

Second, because a widescreen displayed the robot in the telepresent condition, participants might have experienced the size of the robot differently. Literature shows that size influences the perception of people; taller persons have been found to influence group decisions more [Hua02]. Thus, the size of the robot in the video could affect the ratings. The bigger size of the telepresent robot could confound with differences in the perception based on the embodiment of the robot. Therefore, the robot size could override embodiment effects. Psychological research has also shown that people perceive taller referees as having higher competences and that taller US presidents were more likely to be reelected, and perceived as having more leadership and communication skills [Stu12; Stu13]. Due to the evidence that people respond socially to technology, it might explain why the participants have perceived a taller telepresent robot similar to its smaller co-present counterpart [Ree97].

Third, it is also important to note that I used a recorded video of the real Nao in the remote-located condition. Thus, the realism of both robots might be too close to each other. People are watching a video of a video, and thus the rating between the co-located and remote-located robot condition may not differ. It might explain why the results from vHRI are in contrast to real live HRI experiments. Previous work has investigated the effects of embodiment also using video recordings of a real remote-located robot [Wai07]. Wainer et al. [Wai07] showed that there is a difference between the rating if people interact with a co-located compared to a remote-located robot even though it was the same robot. These results show that participants perceived a co-located robot as more appealing than a remotely present robot and a simulated robot. However, there were no differences between the simulation of the robot or the video presence of a remotely located robot. Hence, the assumption is that it makes no difference to use the simulation of a robot or the video of a remote-located one, which is in line with the findings of a survey on experimental works comparing co-present robots, telepresent robots and virtual agents [Li15]. Nevertheless, the results did not show that there is a difference between the embodiment. Thus, I argue that vHRI experiments are probably not suitable to investigate embodiment, because it exposes participants in a not embodied condition to a video of a video/simulation of a robot interacting with a human. Thus, the participant's attribution regarding the embodiment of the robot might be ambiguous, and the embodiment does not make a difference since the users are not interacting with a real system.

The same argument might apply to the missing difference regarding the role of the robot. Since participants are not interacting with the system, they do not perceive any difference regarding the social role of the robot and thus are not inclined to rate the robot conditions differently. I presume that also a stronger framing in the study design would not have led to different results because those subtle feelings may only be measurable in real HRI.

3.6 CONCLUSION

This chapter presented a vHRI study to investigate the effects of embodiment and social role, which I hypothesized to affect the user's exercising motivation. I looked at whether these features influence a user's evaluation of the robot. However, the results show no differences between the conditions. Hence, I am going to investigate this question in the following chapter in a real HRI study. I will compare how users perceive a robot that is exercising along with the participants to a robot that is only instructing them. Furthermore, I am going to investigate whether this results in any motivational gain for the participants to exercise.

4 MOTIVATIONAL EFFECTS OF EXERCISING WITH ROBOT PARTNERS

The previous chapter presented a study to get user's impressions of an exercising companions. The results of this study give no insights for future work except that this kind of research methodology, namely vHRI, needs to be complemented by real HRI studies. Hence, this chapter presents an investigation of the motivational effects of a robot's social role on exercising duration to find an answer for Hypothesis 1.1 of this thesis.

NOTE: Parts of this chapter have been published in S. Schneider et al. "Exercising with a humanoid companion is more effective than exercising alone." In: *Humanoid Robots (Humanoids), 2016 IEEE-RAS 16th International Conference on*. IEEE. 2016, pp. 495–501

4.1 INTRODUCTION

As stated in the introduction of this thesis (see Chapter 1) it is crucial for robotic design reasons to figure out whether the robot needs to be able to exercise together with a trainee or whether it is sufficient to instruct them to motivate them. The research field of SAR has presented many works on how the presence of the robot influences the motivation and how robots could facilitate working on a task ([Fas10; Ley12; Sch14] and many more). However, no present study explicitly questions whether the mere presence is sufficient for a motivation boost or whether the robot should do a task with the user together. Research on the effects of group dynamics in psychology shows that the least capable group member exhibits a relative motivation gain when performing a task in a group (see 2.2.4). This effect is called the Köhler Effect and was investigated by Feltz et al. [Fel14] for Virtual Agents (VAs). In this chapter, I present a study to replicate this effect with robots and investigates whether a co-actively exercising robot has a stronger influence on a user's motivation than a robot that is not exercising co-actively.

Related to the work of this chapter is a HRI workshop paper that proposes to investigate the Köhler Effect with robot. However, the author never published any subsequent papers regarding this research statement [Swi13]. Thus, I assume that the work presented in this chapter is the first to investigate the Köhler Effect with social robots.

Why is it interesting to investigate this effect with a humanoid robot? The assumption of the Köhler Effect is that the motivation of a less capable team member is influenced by the presence of a more capable team member. For the field of exercising this means that the less physically fit team member will be more motivated when exercising with a more physically fit team member. However, the fitness level and muscle fatigue of a team member are challenging to control, and an adequate training partner might not always be available for rehabilitation or exergaming tasks. Consequently, humanoid robots could be a tool to support physical activity (PA), since machines have no muscle fatigue[1] or motivational issues. Furthermore, relying on a human exercising partner comes with additional problems like time scheduling and health condition of the human exercising partner. Thus, a humanoid exercising partner could bring an additional benefit besides the possible motivational effects.

The secondary objective is to investigate the motivational effects of a Robot Companion (RC) which is exercising together with the partner compared to a Robot Instructor (RI) which is only instructing the user to exercise.

Why is it important to raise this question? As stated earlier, building robots that are capable of exercising along with a user requires much more work on the hardware design and dynamics control. Thus, it would be favorable, from a cost-benefit point of view, not to be forced to build robots that are able of all the complex motions required for exercising.

To tackle these two objectives, the study in this thesis builds upon previous studies that tested the Köhler Effect with VA [Fel14]. Feltz et al. used five different abdominal isometric exercises to identify the motivational effects of varying degrees of a VA partner. The authors chose the abdominal plank exercises because they do not require high motor skills and their study design incorporates different fitness levels of the participants. In the present study, I use the same exercises to test whether a Robot Companion enhances a user's exercising time compared to a Robot Instructor.

This chapter is organized as follows: The next section introduces related work on exercising with artificial agents. Section 4.3 introduces study and system design. Section 4.4 shows the obtained results and in Section 4.5 these results are discussed.

4.2 RELATED WORK AND HYPOTHESES

4.2.1 *Köhler Effect with Virtual Agents*

The Köhler motivation effect in sports interactions was intensively studied by Feltz et al. [Fel14]. The authors investigated different aspects of the Köhler Effect to enhance training engagement and have recently reported a study where they evaluated the Köhler Effect

1 Nevertheless, a robot's motors and batteries wear off over time

with VA [Fel14]. They have tested four different conditions (i.e. an Individually Exercising (IC), Hardly Human Partner (HHP), Nearly Human Partner (NHP), and Human Partner (HP)) during an isometric workout. The participants had to do five isometric exercises (i.e. plank, side plank left/right, plank and lifting left/right leg). Each participant had to do the exercise individually first. After the first round they had a break, and then they either had a partner (HHP, NHP, HP) or did the exercises alone again (IC). Their results show that even if the participants were training together with the HHP, they were holding the exercises longer than in the Individually Exercising condition. Thus, the researchers could show that an artificial remote-located exercising partner is at least better than no partner at all.

4.2.2 *Köhler Effect with Robot*

A literature review revealed one research project that studied the Köhler Effect with robot companions in a preliminary stage [Swi13]. The paper does not present results or detailed study descriptions, and a further investigation showed no follow-up reports on the Köhler Effect with robots. Thus, it seems that there are no other works that tested the Köhler Effect with robot companions. Hence, this study is the first attempt to verify that this effect is also measurable with robot exercising partners. There are nevertheless other researchers that investigated the usage of SAR in exercising task. However, most of them are missing a baseline condition that shows the effectiveness of having a SAR exercising partner (e.g., [Goc06; Fas13; Gun17; Gör13]).

4.2.3 *Hypotheses*

My hypotheses for the presented studies are drawn from the Köhler Effect:

HYPOTHESIS 4.1 *People exercise longer with a co-actively exercising Robot Companion (RC) compared to a Robot Instructor (RI) and to exercising individually.*

HYPOTHESIS 4.2 *People exercise longer in the Robot Instructor (RI) condition than in an individual condition.*

HYPOTHESIS 4.3 *People perceive the robot in the Robot Companion (RC) conditions as more animated, anthropomorphic, likable and intelligent than in the Robot Instructor (RI) condition.*

4.3 STUDY AND SYSTEM DESIGN

The study design replicates the one from Feltz et al. [Fel14] so that a comparison between the results is possible. Nevertheless, due to

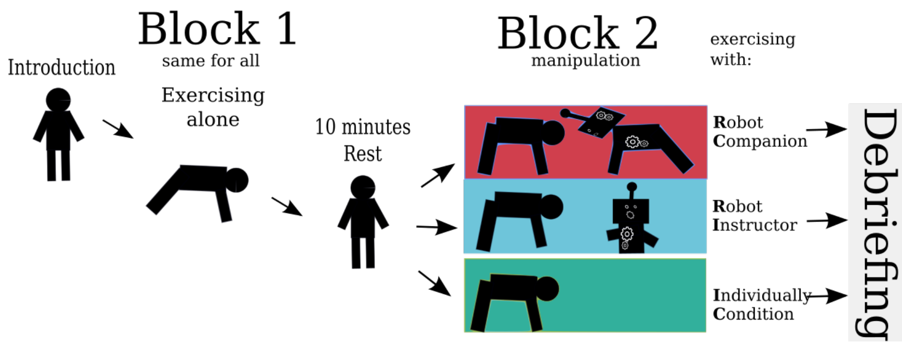


Figure 7: Study design to investigate whether the Köhler Effect can be replicated with humanoid robot companions as exercising partners. Participants exercised in Block 1 individually to obtain a baseline measurement for their exercising abilities on five abdominal plank exercises. The manipulation happend in Block 2. Partipants were randomly assigend to one of three conditions: Individually Exercising (IC), Robot Instructor (RI) and Robot Companion (RC). To reduce biasing effects, participants were neither told that they will have two exercising blocks nor that they will be paired with a robot.

the limitations of the robot, I included changes in the study design. I changed the exercises from forearm planks to full planks due to the robot's limited Degree of Freedom (DoF).

I also neglected to include a condition to study the differences between the embodiment of the robot due to recent evidence for the benefits of robots compared to remote-located or virtually present robots [Li15]. However, I will compare the obtained data from this study with the data from Feltz et al. [Fel14] in Chapter 6.

Additionally, I could not include a Human Partner condition in the study design. Feltz et al. [Fel14] simulated a HP on a video screen due to the impossibility to control that a Human Partner is always performing longer than the study subject, which is an essential factor for the Köhler Effect. I could have simulated a Human Partner using a fake video stream as Feltz et al. did, but this would add the embodiment as another variable in the study design. Therefore, I decided to exclude both a Human Partner condition and a Virtual Agent condition.

4.3.1 *Experimental Design and Participants*

Participants (n=56) were in one of three conditions with 18 participants in the Robot Instructor condition and 18 in the Robot Companion and 20 in the IC condition. Participants were mostly students (29 male, 27 female; mean age M = 25.55 years; standard deviation SD = 6.48) from Bielefeld university acquired by advertisements distributed on the campus. They received seven Euros as monetary compensation. I excluded three participants from the IC. One was an outlier already persisting much less during the first part of the session when the participants were exercising by themselves compared to all other participants, possibly because the instructions were not clear for this participant. I excluded two other participants because they were exercising wrongly. Additionally, data from one participant in the RI from the survey evaluation is missing. In all other cases, no outliers were removed.

4.3.2 *Conditions*

Participants had to do two blocks of five isometric abdominal exercises each (see Figure 8). During the Individually Exercising condition the participants did all exercises two times alone. In the other conditions the participants did the exercises alone first and with the humanoid robot Nao in the second block. During the Robot Instructor condition the robot was announcing the exercises the user had to do, as well as how long the break was. While the users were exercising, the robot was standing in front of the user observing him/her. After the users had finished an exercise, they received general encouraging feedback. In the RC condition, the robot was uttering the same sentences as in the RI condition. However, instead of just standing in front of the participant the robot was exercising together with the user.

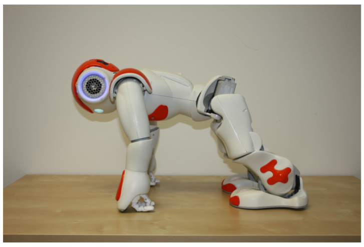
ROBOT COMPANION The Robot Companion (RC) instructs the user, guides her/him through the experiment and, conjunctively exercises with the user.

ROBOT INSTRUCTOR The Robot Instructor (RI) instructs the user and guides her/him through the exercises.

INDIVIDUAL CONDITION Participants exercised individually.

4.3.3 *System Design*

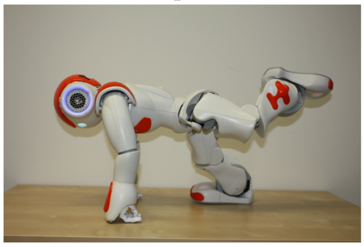
The interaction flow was modelled using the framework described in [Sch17a]. NaoQI SDK version 2.1.14 was used to trigger text-to-speech output and motion on the robot. The exercise animations for the robot were designed in Choregraphe. The animation was exported as Python code for a custom-made tool to synchronize speech and motion on Nao. For visual and auditory perception of the user, Microsoft Kinect and the built-in speech recognition and face detection of NaoQI were used. A simple moving average on the depth user image from Kinect sensor was used to detect whether the user was standing or in the plank position. After the participant has en-



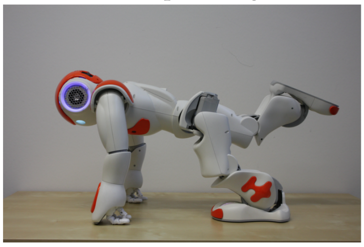
(a) plank

(b) side plank left

(c) side plank right



(d) plank leg raise left



(e) plank leg raise right

Figure 8: The five different abdominal plank exercises used in this study. In contrast to the exercises used in [Fel14], the exercises had to be changed from forearm planks to full planks due to the robots limited DoF. At the point of this study it was not possible to animate the forearm side plank exercises on Nao. Exercises were animated on Nao using the animation toolbox of the Graphical User Interace (GUI) `Choregraphe` provided by the `NaoQi SDK`.

tered the room, the system recorded an average height of participant that the system used as a treshold to decide whether the person is standing or in a plank position. Combined with an entropy decider from the framework mentioned above this approach was sufficient to recognize the user position. The interaction flow during one single exercises is modelled as an state chart (see Figure A.2.7 in the appendix for an example).

4.3.4 *Study Procedure*

Participants arrived at the lab individually, read and signed a consent form which informed them that they would be recorded during the whole time of the experiment. They watched a short video of Nao demonstrating the five exercises. Afterward, they were brought to a fitting room to change clothes and strap on a heart rate belt.

They were instructed to 1. do each exercise as long as they could and to stand up immediately if they were unable to hold the pose any longer, 2. rate their perceived exertion, 3. take a break for 30 seconds, 4. and then to continue with the next exercise.

Participants were guided to the lab and told to start after they had waited for a short time so that the experimenter could check wether the recording was working correctly (see Figure 9 for an overview of the experimental room). Then the participants did each exercise alone in the lab while the experimenter observed them from a different room and took the times of each exercise. The participants completed Block 1 (each exercise once). Afterward, the participants had a ten-minute break in which the experimenter offered them a glass of water. After the break participants were told their average time holding the planks and that they would complete the same set of exercising again (Block 2).

In every condition, the participants were not told that they would do a second block of exercises until they had finished the first block. In the Robot Instructor condition, the experimenter told the participants that they would do the same set of exercises again but that this time a robot would be present which would be instructing them. In the Robot Companion condition, the experimenter told the participants that they would do the same set of exercises again and that this time a robot would be present. This is where the Köhler Effect manipulation happened. The experimenter told participants in the Robot Companion condition that they would exercise as a team from now on and that the team's time was the time of the one who stopped holding the plank first, which created a conjunctive task. The experimenter told the participants the true average time they held the planks, like in the Individually Exercising condition, but gave them false information on how long the robot could persist the exercises. The robot's average plank time was reported by the experimenter to be forty percent higher than the average time of the participants, which created an unfavorable comparison. This discrepancy, in line with previous research, leads to more significant effects and was adopted from the previous study [Fel14]. Again the experimenter did not enter the room together with the participant.

Nao was waiting in the room for the participant to enter. When the system detected that a person was in the room (using the depth sensor) and standing in front of the robot (using the face detection), the robot greeted the user. In both robot conditions, the participant and robot had a short interaction phase. During this phase, the robot told them its name (Nao), hometown (Paris) and hobbies (gardening, reading) and waited for a short time to give the human participant a chance also to share his/her personal information. This information exchange is vital because prior research showed that people treated agents more like humans when there was an initial verbal interaction

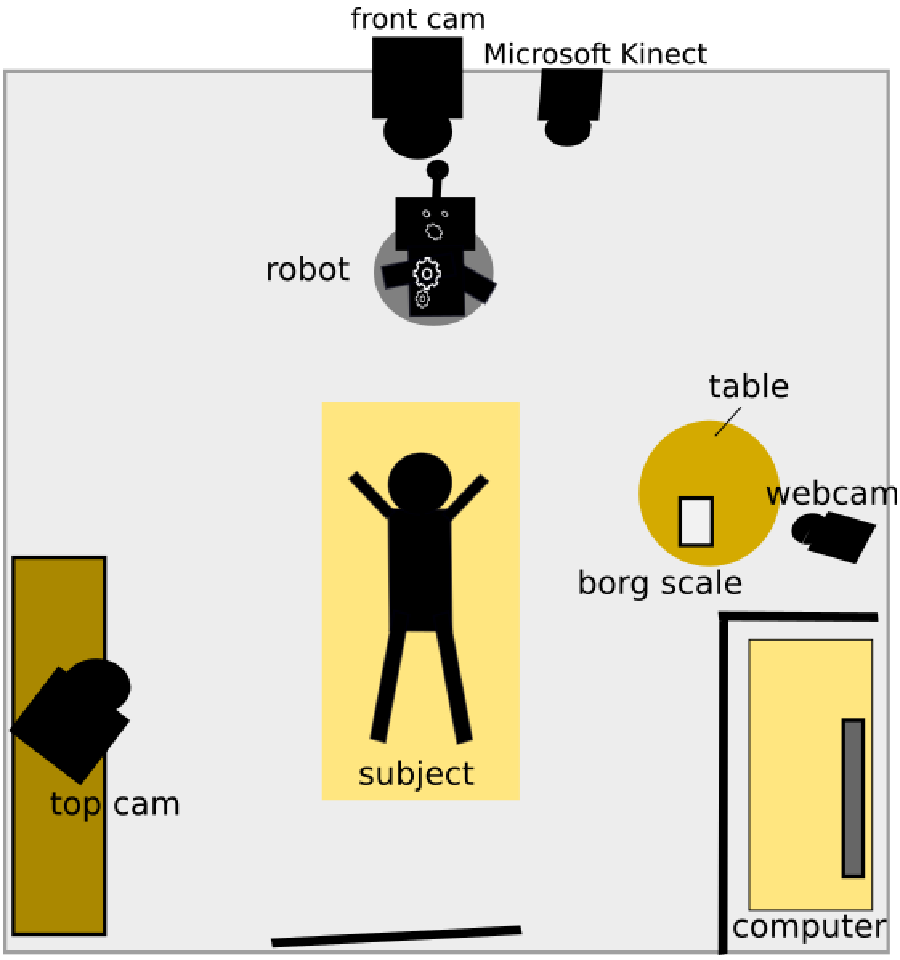


Figure 9: Overview of the experimental room for the studies of this thesis. The robot was positioned so that it will be the first thing participants see when they enter the room. Two cameras were used to record the sessions. Participants were instructed to exercise on a yellow yoga mat. A table was placed in the room, where a paper questionnaire was positioned. This forced participant to stand up immediately and enter their values on the perceived exertion (BORG) scale. A computer was positioned behind partitioning walls where the system was running. Additionally, a kinect was positioned in the room to perceive the user depth image. The webcam was used to monitor the studies.

between them [Bic05]. Then, the robot asked the user whether to start the exercises. After the robot had detected any synonym of 'yes' using the internal speech recognition, Nao instructed the person to go to the plank position and also went to the plank position. In the RI condition, the robot remained standing in front of the user, waiting for the user to go to the ground before announcing to start the first exercise. Then the robot verbally announced the next exercise and instructed to hold the exercise as long as they can. During the exercise, the system gave no feedback to the users, and the system waited until the user finished the exercise. When the participant finished the exercise by standing up or laying down the system triggered a simple motivational phrase (i.e."you did very well"), instructed the user to rate their exertion feeling and told them that there would be a 30 seconds break. After every 30 seconds pause, the robot announced the next exercise and the behavior repeated until the last exercise. When the participant completed the last exercise, the robot thanked for the participation, told the user that he/she was allowed to leave the room, that it needed to rest a bit and powered itself down. After leaving the room, the participant completed a questionnaire, was debriefed and received monetary compensation. The whole procedure took about 45 minutes to one hour. The ethical comittee of Bielefeld University approved this study and the following studies in this thesis.

4.3.5 *Measures*

PERSISTENCE Persistence was the number of seconds a plank was held from the moment participants moved into position until they quit. Block scores were calculated as average of seconds held on all five exercises in one block:

$$\overline{\mathrm{Block}_x} = \frac{1}{n}\sum_{i=1}^{n} t_i$$

where x is the exercising block, t_i is the duration (s) a participant holds the exercise i, and n the number of exercises.

PERCEIVED EXERTION Perceived exertion was measured using the Borg rating scale ([Bor98]). The scale goes from 6 to 20 (6: "no exertion at all", 20:"maximal exertion"). The participants were asked to rate their exertion immediately after each exercise.

PERCEPTION OF THE PARTNER To asses different perception of the robot between the conditions, participants were asked to rate the robot based on the Godspeed questionnaire (5 point-based differential scale, [Bar09]).

ROLE OF THE INTERACTION PARTNER Participants were asked to rate which role they would ascribe to the robot using a multiple-choice input. See Figure 12 for a list of available choices.

PHYSICAL TRAINING ENJOYMENT Participants rated their physical training enjoyment with the Physical Activity Enjoyment Scale (PAES) [Ken91]. The average overall item responses calculate the overall enjoyment score. See Appendix D.3 for the used scale.

INTENTION TO EXERCISE Participants were asked to rate their intention to train tomorrow for at least 30 minutes on a five-point Likert scale.

WISH FOR HAVING AN ASSISTIVE SYSTEM At last, participants had to answer whether thez would like to use a system that supports them during exercises in the future.

4.4 RESULTS

4.4.1 *Manipulation Check*

An KW-Test showed no differences in age, $H(2) = 1.33$, $p = .51$. However, a KW-Test showed a significant effect for weekly physical activity between the three conditions, $H(2) = 6.22$, $p < .05$. Though, multiple comparison test after the KW-Test showed no significant differences between the conditions.

A preliminary ANOVA on Block 1 persistence scores showed no significant effects, $F(2, 49) = .63$, $p = .53$, $\eta_p^2 = .02$. Random assignment was successful in creating no differences in the mean persistence at Block 1. This is important because a significant difference during the first block would indicate different physical fitness between the groups.

4.4.2 *Persistence*

I adopted measurement and analysis from previous studies on the Köhler Effect to be able to draw conclusions and comparisons [Fel14].

As a primary dependent variable, I used the average difference persistence time in seconds between the two blocks ($\overline{\mathrm{Block}_2}$ (s) - $\overline{\mathrm{Block}_1}$ (s)). This approach controls for individual differences in strength and fitness and shows possible changes in persistence. The results obtained for the average block score of Block 2 subtracted with the average block score of Block 1 are shown in Figure 10. A 3 (conditions) x 1 (persistence) ANOVA on the difference scores showed a significant main effect for the conditions, $F(2, 49) = 11.8$, $p < .001$, $\eta_p^2 = .32$, $\omega^2 = .29$.

Participants in the Individually Exercising condition persisted on average 16 seconds less in Block 2 than in Block 1. Prior studies observed this difference due to fatigue or/and boredom of the study participant [Ker11; Irw12]. The results of this condition will now be used to detect any motivation gain in the other conditions. Participants in the Robot Companion condition persisted at almost the same time on average at Block 2 as in Block 1. A pairwise comparison using t-tests with pooled SD and Holm adjustment revealed that the difference of 16 seconds between Individually Exercising and the Ro-

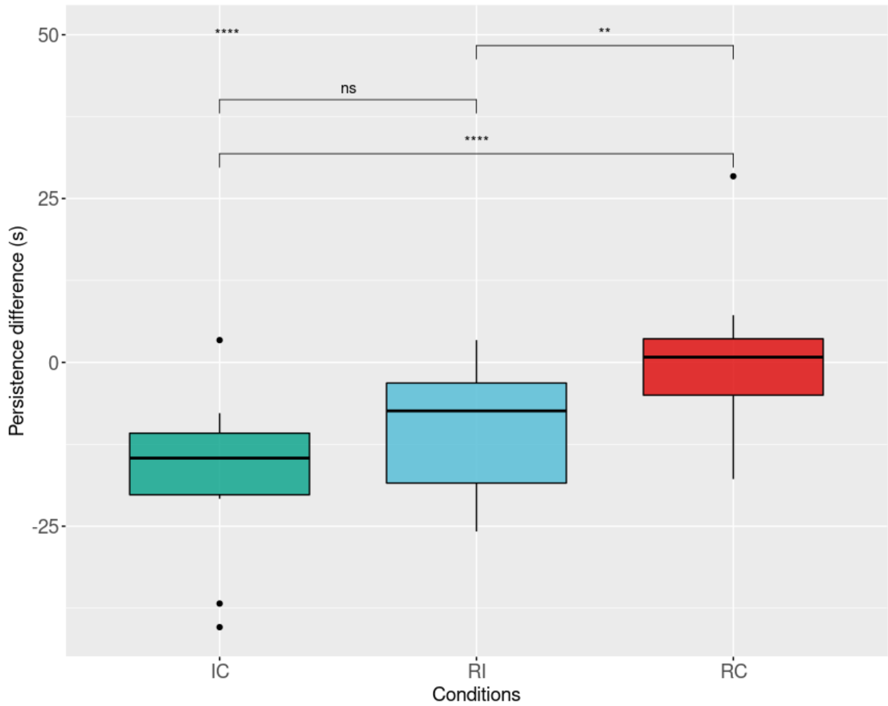


Figure 10: Box plots showing the mean duration (s) exercise persistence difference scores ($\overline{\texttt{Block}_2}$ (s) - $\overline{\texttt{Block}_1}$ (s)) for the conditions Individually Exercising (IC), Robot Instructor (RI) and Robot Companion (RC) for the Köhler Effect study. Results show significant exercising time differences between the RC condition and the RI and IC conditions.

bot Companion condition is significantly different ($p < .0001$). Also the persistence in the Robot Instructor condition was slightly greater (M = -10.29, SD = 9.09) than the Individually Exercising baseline (M = -16.12, SD = 10.35). However, this difference is not significant (p = .07) Furthermore, the persistence between the Robot Companion condition (M = -0.05, SD = 9.83) is also significantly greater than in the Robot Instructor condition (M = -10.29, SD = 9.09, $p < .01$). Hence, for this participant population, a difference in the social role of the robot did affect the magnitude of the observed Köhler Effect.

4.4.3 *Perception of Partner*

The reliability scores of my items are: anthropomorphism (α = .8), animacy (α = .77), likability (α = .88), intelligence (α = .82). Figure 11 shows the results of the Godspeed questionnaire and PAES.

A Welch Two Sample Test indicated that ratings for animacy from users in the Robot Companion were statistically significantly higher than in the Robot Instructor ($t(31.99) = 2.13$, $p < .5$, $d = .73$). A WC-Test with continuity correction indicated that ratings for likability from users in the Robot Companion were statistically significantly higher than in the Robot Instructor, $W_s = 223.5$, $p < .01$, $r = .46$.

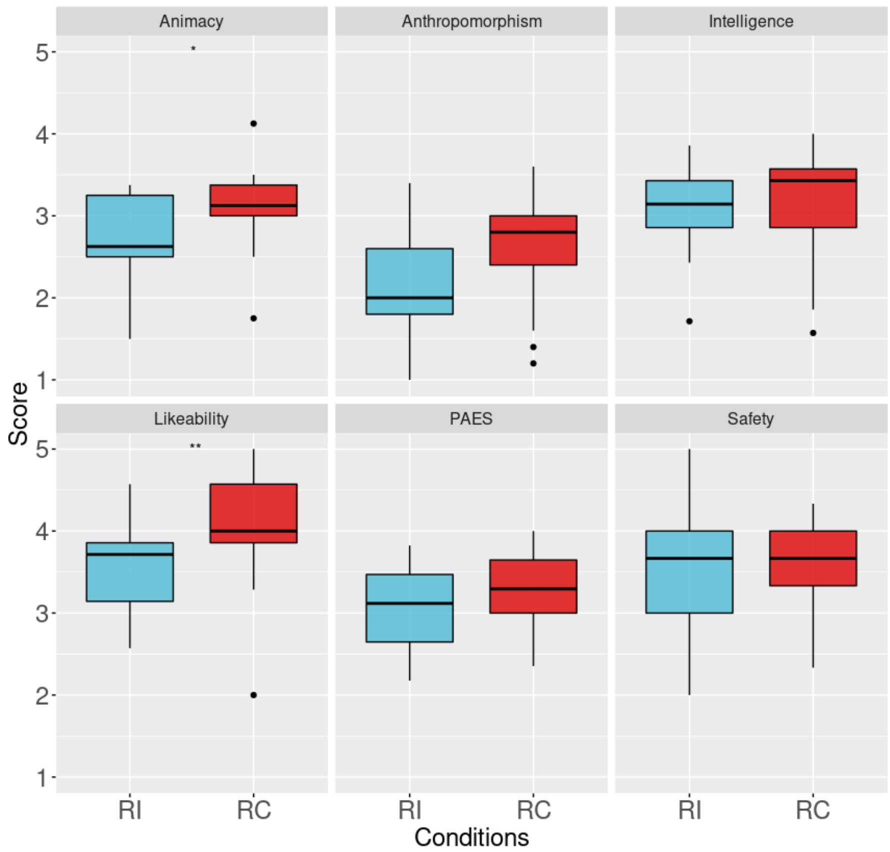


Figure 11: Boxplots depicting the results of the Godspeed questionnaire. ANOVAs show significant differences on the likability and animacy scale between Robot Instructor (RI) and Robot Companion (RC) in the Köhler Effect study.

4.4.4 *Perceived Exertion*

A KW-Test for the perceived average exertion per block showed no significant differences between all conditions, $H(2) = 4.45$, $p = .10$.

4.4.5 *Future Assistance*

A KW-Test for the wish for future assistance ($\alpha = 0.8$) showed no significant differences between the conditions but a tendency, $H(2) = 5.1$, $p = .07$.

4.4.6 *Social Role*

The frequency table for the social role ascription is depicted in Figure 12. The frequencies between the conditions are quite similar. Nevertheless, participants in the Robot Companion condition perceived the system more often as a coach. There was no significant association between the kind of robot condition and the association for a social role when considering the whole frequency table using a Fisher's Exact Test (FET) for Count Data ($p = .449$). However, there was a significant association between the type of robot condition and whether the participants ascribed the robot the role of a coach ($\chi^2(1) = 5.62$, $p < .05$). This seems to represent the fact that, based on the odds ratio, the odds for the ascription of the role as a coach were 0.11 (0.009, 0.72) times higher if participants trained with the Robot Companion than with the Robot Instructor.

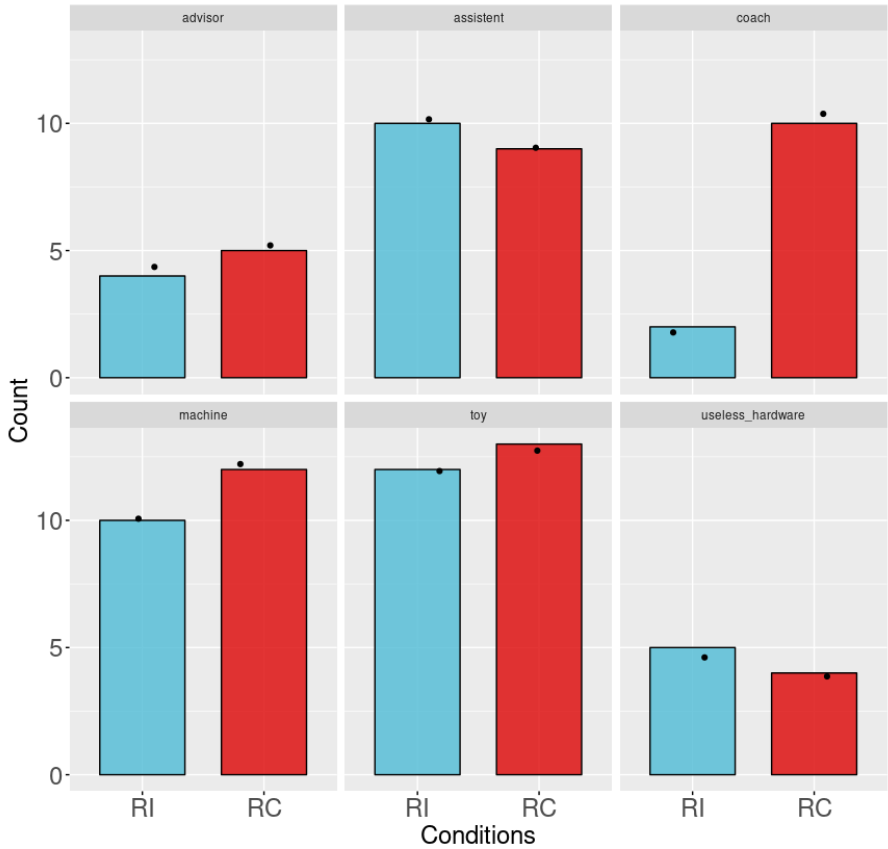


Figure 12: Bar plots showing the perceived social role of the system in the Köhler Effect study. Participants ascribed the Robot Companion (RC) more often the role of an coach compared to the Robot Instructor (RI).

4.4.7 *Further results*

INTENTION TO EXERCISE Significant differences were found on intention to exercise, $H(2) = 6.3$, $p < .05$. However, multiple comparison test after the KW-Test showed no significant differences between the conditions.

ACTIVITY ENJOYMENT As determined by ANOVA, PAES (α = .82) was also not different between the three conditions, $F(2, 49) = 1.69$, p = .2, η_p^2 = .06.

4.4.8 *Free Responses*

Participants could leave a response in the survey. An english translation of the qualitative user feedback related to the robot is listed in the following for each condition. The original German version is in parentheses.

ROBOT INSTRUCTOR CONDITION

RI01 "I felt pretty pressurized, just by the presence of the robot. In his presence, I had great problems concentrating completely on the exercises"
("Ich habe mich ziemlich unter Druck gesetzt gefühlt, allein durch die Anwesenheit des Roboters. Ich hatte in seiner Anwesenheit große Probleme mich vollkommen auf die Übungen zu konzentrieren")

RI02 "The robot could join the task. and distract from the effort. I felt like I had a great performance at the first encounter in the practice, because that was new and I was interested"
("Der Roboter könnte die Aufgabe mitmachen und von der Anstrengung ablenken. Ich hatte das Gefühl ich hätte eine Mordsperformance bei der ersten Begegnung in der Übung, weil das neu war und mich interessiert hat")

RI04 "It should be able to perform and, if necessary, correct exercises and be motivating in the sense of a competitive function (different levels of difficulty)"
("Es sollte in der Lage sein, Übungen vorzumachen und ggfs. zu korrigieren, sowie motivierend im Sinne einer Wettbewerbsfunktion sein (verschiedene Schwierigkeitsstufen)")

RI17 "[It] could motivate something more"
("[Es] könnte etwas mehr motivieren")

ROBOT COMPANION CONDITION

RC03 "More motivation during the exercise"
("Etwas mehr motiv[i]eren in der Übung")

RC12 "If you can not persist it anymore, the robot should motivate you"
("Falls man nicht mehr durchhalten kann, sollte der Roboter einen motivieren")

RC18 "Motivational words during the exercise"
("motivierende Worte während den Übungen")

4.5 DISCUSSION

The primary objective of this chapter was to study whether the Köhler Effect can be demonstrated using robotic partners. The secondary objective was to investigate whether the presence of a Robot Instructor (RI) also has a facilitating effect.

The results support that performing plank exercises with a Robot Companion (RC) boosts one's effort relative to performing those exercises individually or with a Robot Instructor (RI). This result supports the Hypothesis 4.1. Participants significantly increased their effort in exercising even though they were told that their partner was a robot and not a real human. This finding is in line with previous research where researchers found that people often respond socially to technology as if interacting with a human. Nass et al. [Nas96] reported that people perceive computers as teammates and experience the same team dynamics similar to human teams. Hence, this study is the first to validate the Köhler Effect with a Robot Companion and shows the motivational effects a human has when conjunctively working out with a robotic partner.

However, the results do not support Hypothesis 4.2. The presence of the Robot Instructor (RI) did not result in better performance compared to the Individually Exercising (IC) condition. Comments from the participants give a possible explanation. In post-study interviews and open feedback in the survey, participants in the Robot Instructor (RI) condition reported that they were wondering about the usefulness of the robot and that they felt irritated by its presence. In previous research on SAR supporting users in cognitive tasks I came to a similar conclusion [Sch14]. If the usefulness of the SAR is not evident for the user, its mere presence has no facilitating effects. This conclusion contrasts other studies showing that the presence of a robot which is not explicitly exercising together with the user is also useful (e.g., [Pow03; Tap08b]). There are two possible explanations for these differences.

First, compared to other studies, the robot in this study did not give any task-related feedback to the user. It announced the next exercises, reminded the user when the break was over and gave some general encouragement after the exercises. Those utterances were neither individualized nor task-specific.

Second, that the robot was not exercising conjunctively with the user, but just present in the room and observing the user might indicate an audience effect. This effect describes that people perform better on a dull or rehearsed task in the presence of others (humans and robots) and worse on complex, challenging or new tasks compared to their performance when alone [Str02; Rie12].

The results partially support Hypothesis 4.3. The Robot Companion was rated higher on two of four scales of the Godspeed questionnaire compared to the Robot Instructor. It is not surprising that the Robot Companion was rated higher on the animacy scale. However, it is interesting that it was rated more likable than the Robot Instructor due to joint training. Thus, one can hypothesize that higher ratings of likeability increase the Köhler Effect for robotic exercising partners.

At last, there was a significant difference in performance between the conditions, but no differences for exertion, intention, and enjoyment. Participants exercised longer without any feeling to be working harder, enjoying it less or hindering their future exercise plans. These results show that it is possible to extend exercising time without adverse effects. The results are in line with the findings from [Fel14]. Their study showed that a Virtual Agent also could increase the exercising time. The data from both studies will be compared in Chapter 6 to conclude whether a virtual partner is more motivating than a robot exercising partner.

A limitation of the study is the lack of a human partner condition. In previous research, a human partner was simulated using a prerecorded video of a confederate [Fel14]. Since this experiment focussed on the motivating effects while the partner is in the same room, the human partner condition was neglected. There are various reasons for this decision: The positive impact of a human partner has already been investigated. Furthermore, the implementation of a human partner condition where the human is always performing the exercises longer than the study subject is almost impossible to control. During repeated experiments on the same day, it is not possible to ensure that the human partner has no muscle fatigue and is always showing a moderately higher performance than the study subject. A pre-recorded human partner condition would also be feasible. However, it would have introduced a remote-location as another source of variability in the study design. Therefore, this study only used robot conditions. A further limitation is the restrictions to variants of abdominal plank exercises. Thus, it is uncertain whether the found effect also appears in other types of exercising (e.g., strength or cardio exercises).

Future work, which is not part of this thesis, could also investigate the audience effect. A study could explicitly gather two participant populations, one that is experienced in doing the exercises and one that is not. This study would allow making accurate conclusions whether experienced people will perform better in the presence of the Robot Instructor (RI) compared to less experienced people. Additionally, the verification of the Köhler Effect might also have an impact on other scenarios using robots. For example, robots in educational settings could be designed as co-learners. Hood et al. [Hoo15] present a project that goes in such a direction. Additionally, it is worth investigating how the Köhler Effect could be applied to industrial settings were robots are co-workers.

4.6 conclusion

This chapter showed that exercising conjunctively with a Robot Companion increases the motivation to exercise longer compared to an Individually Exercising and a Robot Instructor condition. Thus, to the best of my knowledge, this is the first study to experimentally verify the Köhler Effect with a humanoid robot and supports Hypothesis 1.1 of my thesis. The Köhler Effect can be replicated with robot companions and results in a higher motivational gain to exercise than purely instructing robots.

However, the interactive capabilities of the system were intentionally limited. Thus, participants stated that the usefulness of the system was not evident for them in the Robot Instructor condition. Therefore, the next chapter will address the open question on the motivational effects of receiving encouraging feedback from a SAR during exercising.

5 MOTIVATIONAL EFFECTS OF FEEDBACK FROM ROBOTS

The previous chapter showed that a Robot Companion (RC) could enhance the exercising motivation based on the Köhler Effect. Additional motivational effects might be induced by giving feedback, which is the main topic of this chapter. It answers the question of whether encouraging feedback for the user can further enhance the exercising motivation in short-term interaction.

NOTE Parts of this chapter have been published in S. Schneider et al. "Motivational Effects of Acknowledging Feedback from a Socially Assistive Robot." In: *Social Robotics: 8th International Conference, ICSR 2016, Kansas City, MO, USA, November 1-3, 2016 Proceedings*. Springer International Publishing, 2016, pp. 870–879

5.1 INTRODUCTION

Being a companion is one motivational effect suitable for socially assistive scenarios. However, recently developed SARs not only support people with their presence, but also by using verbal and gestural assistance [Fas12; Ley14a]. Moreover, VAs and SARs, which were used as an alternative to conservative methods to support peopl eon dieting, exercising or cognitive tasks, exploit different findings from psychology like goal-setting, empathy, backstory, and personalization (e.g., [Ley14b; Lei14; Goc05; Bic05; Fas12; Bic01; Kid08b].

One other important aspect that coaches are regularly using is an encouragement. Encouragement is a type of feedback, which is used to appreciate the current state of exercising which motivates a trainee to keep up on a task. It is positive reinforcement and does not compare the current performance to other persons or previous sessions. Previous works studied different types of feedback like positive, comparative or corrective feedback in HRI (e.g., [Ham14; Swi15; Süs14]). However, looking at the studies on feedback from SAR for exercising applications, few published works present comparative results to a baseline condition using a negative control like no robot or no feedback. Therefore, it is difficult to distinguish between various motivational effects of SAR, especially the positive reinforcement like encouragement.

Chapter 4 showed that a Robot Companion could enhance exercising motivation. Nevertheless, if the system is only instructing, users perceived the robot as not useful. Thus, the aim of this study is to analyze the quantitative motivational effects of encouragement from a robot while working out. The result of this chapter might contribute to decision whether a SAR needs to workout together with a person or whether having encouragement is sufficient to motivate a user to exercise longer. My hypotheses are that:

HYPOTHESIS 5.1 *Participants exercise longer when they receive encouraging feedback from a Robot Companion compared to robots that do not give feedback or are not co-actively exercising.*

Backed with evidence from the previous chapter, I argue that encouragement can further enhance the positive effects of a Robot Companion. Therefore, exercising together with a robot paired with direct feedback (i.e., Robot Companion with Feedback (RCF)) should result in a peak of training success.

HYPOTHESIS 5.2 *Participants exercise longer when they receive encouraging feedback from a Robot Instructor compared to a robot that is just instructing them and to an individual condition.*

Also for the the robot instructor that is giving feedback (i.e., Robot Instructor with Feedback (RIF)), I argue that encouragement will increase persistence time. However, the exercising time will be shorter compared to the Robot Companion with Feedback, because the robot is not co-actively working out.

The chapter is organized as follows: The next section gives the reader an overview of existing literature in the field of motivational feedback from SARs. Afterward, I will describe the study design to investigate my hypotheses and the system design I used. Section 5.4 presents the results and Section 5.5 discusses the results.

5.2 RELATED WORK ON FEEDBACK FROM SAR

This section highlights studies about SARs which give (motivational) feedback to the user.

RELATIONAL FEEDBACK The preferences for a relational vs. a non-relational robot-coach have been investigated in [Fas12]. The study shows that users have a preference for relational feedback. The authors define relational feedback as the robot's capability to exploit its social interaction and personalization skills. Thus, the robot always gives the user praise for accurate completion of an exercise, it provides reassurance in case of failures, refers to the user by its name, references to past experiences and uses humor. In the non-relational condition, the robot coach gives instructional feedback but does not employ any relationship building. The authors of this study used an in-between subject design to evaluate the different relational styles of the robot. They did not find any differences in the exercise performance based on the relation or non-relational robot.

PERSONALIZED FEEDBACK In my previous work I have investigated the effects of performance-based feedback for users doing a cognitive task [Sch14]. Personalized feedback from robotic tutors has also been investigated in [Ley14a]. In these scenarios, the robot gave the participants individual recommendations how to perform better on a task depending on the user's ability. Both works support that individualized feedback based on the user's performance can increase the tutor's effectiveness.

COMPARATIVE FEEDBACK Swift-Spong et al. [Swi15] compared the effects of self-comparative vs. other-comparative feedback to a control condition with no feedback in a push-button task. They hypothesized that comparative feedback conditions will produce higher self-efficacy and better performance (compared to nCF). Moreover, participants will perceive the robot coach more positively in comparative feedback conditions (sCF and oCF). Though, the authors could not find evidence for any of their hypotheses.

EVALUATIVE FEEDBACK Positive, negative and neutral feedback of a robot or human instructor has been studied in [Par11]. Participants answered a short quiz and received either positive, negative, or neutral feedback regardless of the true score. Results show that study subjects have a preference for positive feedback from a robot. However, they could not find any feedback preference from a human interaction partner.

EMPATHIC FEEDBACK Leite et al. [Lei14] studied the effects of empathic feedback from a SAR. This empathic feedback is composed of an empathic appraisal (i.e., a facial expression associated with the user's affective state) and a supportive behavior (e.g., information support, esteem support). They used the robot as a chess partner in schools. The robot gave empathic feedback regarding the child's valence and the game state. They tested their system in a long-term study, which is why they did not incorporate a baseline condition in their study design.

ENCOURAGING FEEDBACK Encouraging feedback in group exercising has previously been studied in the exergaming research community with telepresent human partners [Irw13]. They investigated the effects of having a partner with and without encouraging feedback versus a a baseline condition without a partner. Their results show that participants exercise less when they were paired with a more capable partner who is giving encouraging verbal feedback. They conclude that encouragement from the stronger to the weaker team member can mitigate the motivational effects that were found in the Köhler Effect conditions.

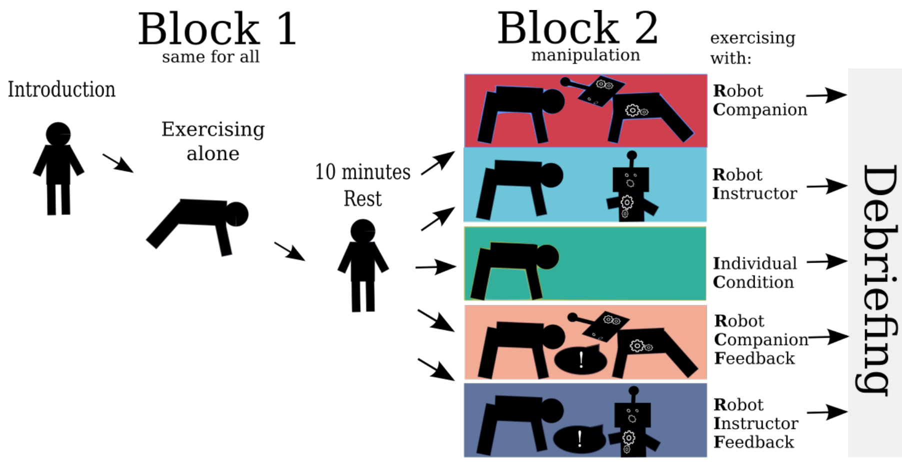


Figure 13: Study design to investigate whether encouragement has an additional effect on exercising motivation. The study design is the same as in Chapter 4. Only two additional conditions were measured: Robot Instructor with Feedback (RIF) and Robot Companion with Feedback (RCF).

The presented works, show the beneficial effects of an agent's feedback for the user in various scenarios. Still, the quantitative motivational effects of feedback from robots in exercising scenarios need further investigation. Thus, this chapter studies how the inclusion of encouraging feedback from a Robot Companion or Robot Instructor influences the exercising time compared to a baseline condition. The results of this study may allow conclusions about the importance of feedback for the user's motivation to persist an exercise as well as how the social role moderates it.

5.3 STUDY DESIGN

The study design in this chapter is the same as the one in Chapter 4 with two additional conditions where the robot is encouraging while the user is exercising. Figure 13 depicts the used study design.

5.3.1 *Experimental Design and Participants*

I used the results from the previous study and collected data for two additional conditions. Thus, the data set included 95 participants for this analysis (44 female , 51 male; average age M = 25.4 years, standard deviation SD = 5.6). 56 participants were from the experiment from the previous chapter plus 39 participants from this study. Participants were mainly students from Bielefeld university acquired by advertisements and distributed in one of five conditions, respectively IC, RI, RC, RIF and RCF. They received seven Euros as monetary compensation. Besides the six participants that had to be excluded in my previous experiment, two participants were excluded from the RCF condition from the survey evaluation because their questionnaire answers were missing. However, it is still possible to analyze their exercising data. Exercises were the same five abdominal plank exercises as in the previous experiment (see Figure 8). Study procedures were also the same as in the previous chapter (see Section 4.3.4).

5.3.2 *Conditions*

This study includes two additional conditions in which the robot is giving encouraging feedback. In these conditions (i.e., RIF and RCF) the robot was giving some verbal encouragement during each exercise e.g., "Keep on! You are doing great!".

robot companion with feedback The Robot Companion with Feedback (RCF) had the same behavior as the Robot Companion (RC) in the previous chapter. Additionally, this robot gave encouraging feedback while the user was exercising.

robot instructor with feedback The Robot Instructor with Feedback (RIF) had the same behavior as the Robot Instructor (RI) in the previous chapter. Additionally, this robot gave encouraging feedback while the user was exercising.

This point in time to trigger the encouraging feedback was generated based on the user's performance in Block 1. The ratio between Block 2 and Block 1 is depicted in Figure 14. The figure shows the proportion of how much of the time from Block 1 the participants could persist the exercises in Block 2 for all condition. This depiction is similar to Figure 10, but instead it shows the percentage of the persistence time between the two blocks. Since most of the participants at least persist 75% of the time in Block 1 also in Block 2, I have chosen this as a threshold for the feedback generation of the robot. This threshold is drawn as the red line in Figure 14. The system was implemented using the same framework as in the previous chapter and presented. The only change is that the interaction state chart now includes an acknowledgment state that triggers the encouraging feedback (see Figure A.2.8 in the appendix).

5.3.3 *Measures*

The same measurements as described in Section 4.3.5 were used. Additionally, this study includes measurements on the negative attitudes towards robots and perception of the partner as teammates.

negative attitudes towards robots Negative attitudes towards robots were measured using the Negative Attributes Towards

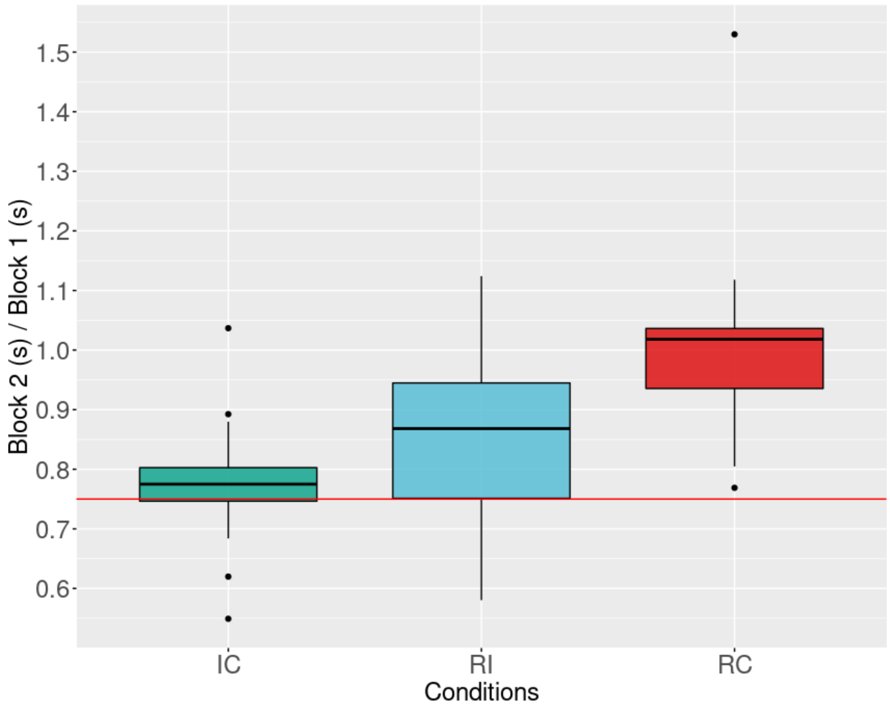


Figure 14: Ratio of individual exercising persistence of Block 2 to Block 1 for conditions: Individually Exercising (IC), Robot Instructor (RI) and Robot Companion (RC). The red line shows the chosen threshold when to generate an encouraging feedback in the present study. This line shows that most participants exercise in Block 2 at least three quarters of their times from Block 1.

Robots Scale (NARS) (e.g., 'I would feel uncomfortable controlling a robot') on a five-point Likert scale [Nom06]. Negative attitudes towards robots could be a confounding factor explaining results obtained on persistence measurements or perception of the robot (see Appendix D.5 for the used scale in this study).

PERCEPTION OF TEAMMATES The participant's perception of the robot as a team mate, was taken from the work of Nass et al. [Nas96] on computers as teammates. This measures include the participant's perceived information quality of the system (e.g., 'the information of the system was useful', see Appendix D.9.1 for the used scale), the cooperation with it(e.g., 'I tried to cooperate with the system', see Appendix D.7 for the used scale), the openness to be influenced by it (e.g., 'I was open to the system's suggestions, see Appendix D.11.1 for the used scale) as well as a general perceived team perception (e.g., 'We were on the same team', see Appendix D.6, for the used scale). All of these items were measured on a five-point Likert-scale [Nas96].

5.4 RESULTS

5.4.1 *Manipulation Check*

One-way ANOVAs and KW-Tests showed no significant effect on the differences in enjoyment ($F(4, 84) = .9$, $p = .46$, $\eta_p^2 = .04$), performance on Block 1 ($F(4, 84) = .8$, $p = .47$, $\eta_p^2 = .04$), perceived exertion ($F(4, 85) = 1.64$, $p = .17$, $\eta_p^2 = .07$), and overall amount of time spent exercising per week ($H(4) = 4.65$, $p = .32$).

A KW-Test showed a significant difference in the intention to exercise for at least 30 minutes on the following day, $H(4) = 11.26$, $p < .5$. Though, post-hoc multiple comparison tests after KW-Test showed no significant differences. Additional, a KW-Test showed a significant difference in age between the conditions, $H(4) = 8.89$, $p < .5$. Although, multiple comparison tests after KW-Test showed no significant differences between the conditions for age.

5.4.2 *Persistence*

As a primary dependent variable I used the average difference persistence time between the two blocks ($\overline{\mathtt{Block}_2}$ (s) - $\overline{\mathtt{Block}_1}$ (s)) again. The results obtained for the average block score of Block 2 subtracted with the average block score of Block 1 are shown in Figure 15. An ANOVA on the difference scores showed a significant main effect of persistence time for the conditions , $F(4, 85) = 8.13$, $p < .001$, $\eta_p^2 = .27$, $\omega^2 = .24$. A pairwise comparison using t-tests with pooled SD and Holm adjustment revealed significant differences between IC and RC ($p < .0001$), IC and RCF ($p < .0001$), IC and RIF ($p < .0001$), RC and RI ($p < .05$), RCF and RI ($p < .05$) and RI and RIF ($p < .05$) (see Figure 15).

5.4.3 *Perception of the Partner*

Results on the perception of the partner are depicted in Figure 16. ANOVAs and KW-Tests showed no significant main effects for the ratings of animacy ($\alpha = .76$, $F(3, 67) = 1.54$, $p = .21$), anthropomorphism ($\alpha = .76$, $H(3) = 4.11$, $p = .25$), intelligence ($\alpha = .76$, $F(3, 67) = 1.95$, $p = .12$). However, significant main effects for perceived likability was determined by an KW-Test, $\alpha = .84$, $H(3) = 4.19$, $p < .01$. Focused comparisons of the mean ranks between groups showed that likability ratings were not significantly different between the RC and RCF ($\mathtt{difference} = 1.0$), RC and RIF ($\mathtt{difference} = 9.3$), RCF and RIF ($\mathtt{difference} = 10.37$) and RI and RIF ($\mathtt{difference} = 10.32$). However, likability scores were significantly different between RC and RI ($\mathtt{difference} = 19.61$), and RCF and RI ($\mathtt{difference} = 20.7$). With a critical difference of 17.96 for RC vs. RCF/RI and RCF vs. RI and a critical difference of 18.67 in all other cases.

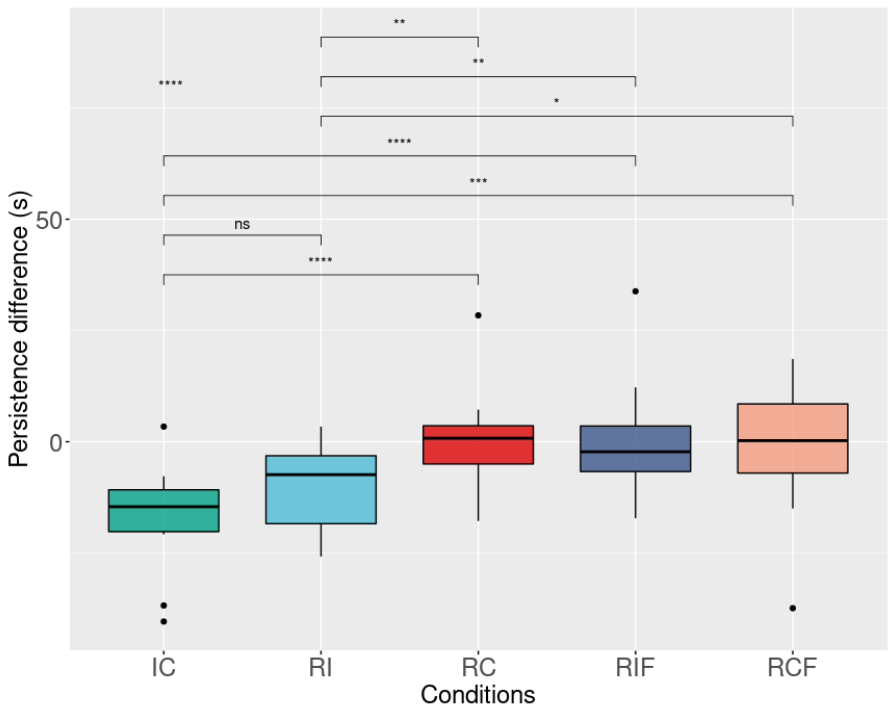


Figure 15: Box plots showing the mean duration (s) exercise performance difference scores ($\overline{\text{Block 2}}$ (s) - $\overline{\text{Block1}}$ (s)) for the conditions Individually Exercising (IC), Robot Instructor (RI), Robot Companion (RC), Robot Instructor with Feedback (RIF) and Robot Companion with Feedback (RCF) for the encouragement study. Additionally to the results depicted in Figure 10, the results show significant exercising time differences between the RCF and RIF conditions and the RI and IC conditions.

5.4.4 *Additional Results*

I analyzed the differences between the ratings for team perception, perceived information, cooperation, openness to influence and NARS between the Robot Instructor with Feedback (RIF) and Robot Companion with Feedback (RCF) conditions as well as the perceived exertion between all conditions. Figure 17 depicts the data of the aforementioned results , but not the perceived exertion.

NEGATIVE ATTITUDES TOWARDS ROBOTS A Welch's two-sample t-test revealed no differences between RCF ($M = 2.57$, $SD = .53$]) and RIF ($M = 2.87$, $SD = .73$) for the NARS scale ($\alpha = .8$), $t(28.78) = 1.42$, $p = .16$, $d = .48$.

OPENNESS TO INFLUENCE A Welch's two-sample t-test showed no significant difference on the openness to influence scale ($\alpha = .84$) between RCF ($M = 3.59$, $SD = .72$) and RIF ($M = 3.39$, $SD = .67$), $t(33.94) = -.82$, $p = .41$, $d = .74$.

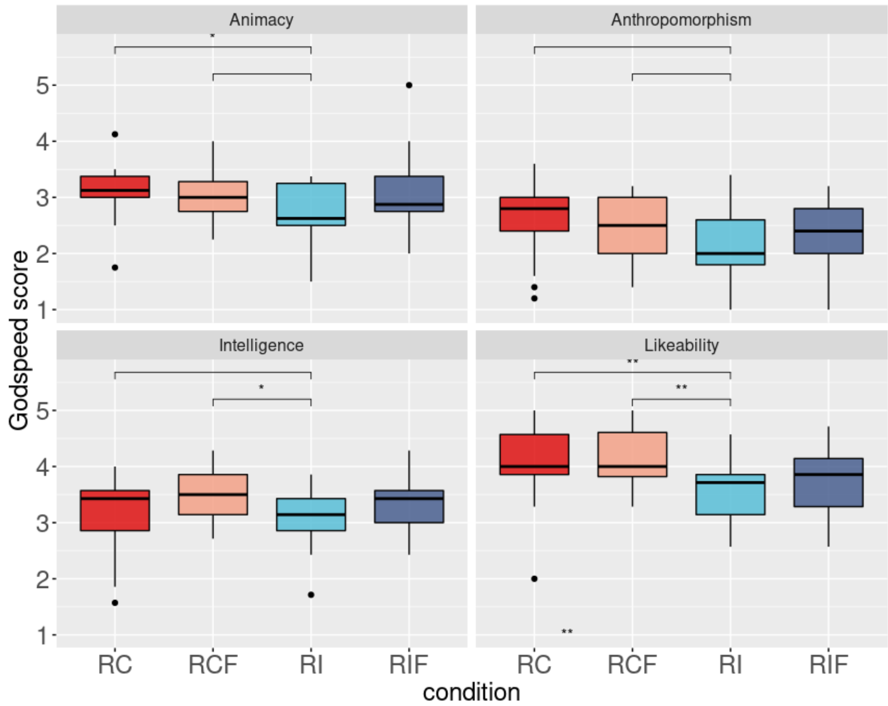


Figure 16: Boxplots depicting the results of the Godspeed questionnaire. ANOVAs show significant differences on the likability scale between the Robot Companion with Feedback (RCF) and Robot Instructor (RI) condition in the encouragement study.

COOPERATION A Welch's two-sample t-test showed no significant differences on the cooperation scale (α = .62) between RCF (M = 3.37, SD = .79) and RIF (M = 3.37, SD = .76), $t(33.78) = -1.2$, $p = .23$, $d = .4$.

TEAM PERCEPTION A Welch's two-sample t-test showed significant differences for the team perception (α = .78) between RCF (M = 3.15, SD = .87) and RIF (M = 2.48, SD = .87), $t(33.57) = -2.28$, $p < .05$, $d = .76$.

INFORMATION QUALITY A Welch's two-sample t-test showed a tendency for the perceived information quality (α = .84) between RCF (M = 3.51, SD = 1.02) and RIF (M = 2.86, SD = .89), $t(33.98) = -2.02$, $p = .05$, $d = .67$.

SELF EFFICACY BELIEFS Self-efficacy beliefs were measured in the present and the previous study to analyze causes for differences in ones exercising times. This analysis is not part of the central thesis. Still, the interested reader can find an analysis in Appendix C.1.

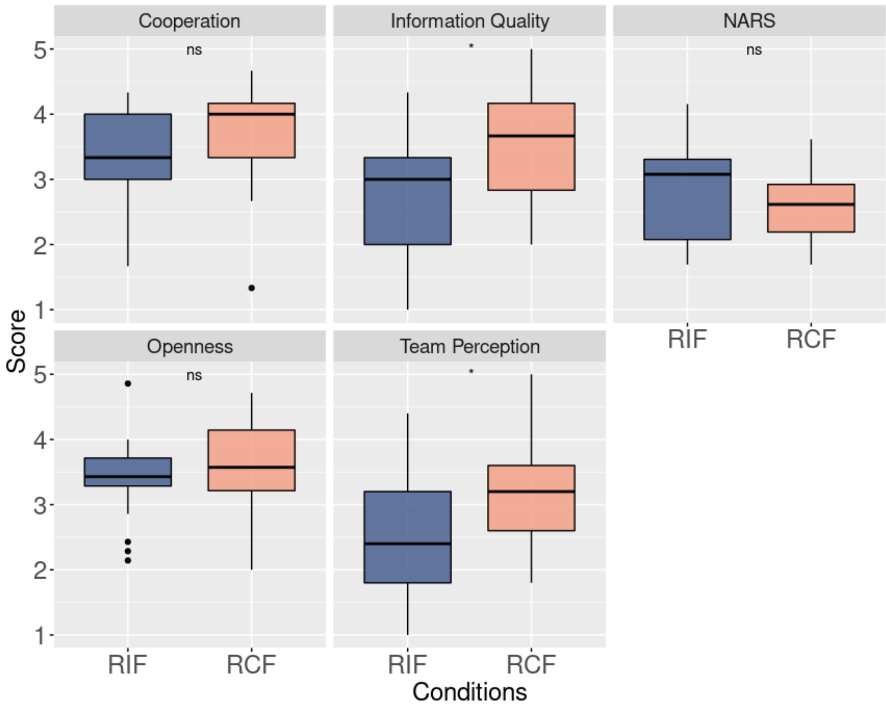


Figure 17: Results and significant differences between the Robot Companion with Feedback (RCF) and Robot Instructor with Feedback (RIF) conditions for team perception, information quality, cooperation, openness to influence and NARS scales from the encouragement study.

5.4.5 *Free Responses*

Participants could leave a response in the survey. An english translation of the qualitative user feedback related to the robot is listed in the following for each condition. The original German version is in parentheses.

ROBOT INSTRUCTOR WITH FEEDBACK

RIF07 “the praise was positive !!!”
(“das Loben war positiv!!!”)

RIF14 “Nao should train. Nao should not be so reserved. Nao should be funny in between.”.
(“Nao soll mittrainieren. Nao sollte nicht ganz so zurückhaltend wirken. Nao sollte zwischendurch lustig sein.”)

RIF17 “I would have expected more support and entertainment from the robot. ”
(“Ich hätte mir von dem Roboter mehr unterstützung und unterhaltung erwartet.”)

RCF08 “Next time using "handstand blocks", that would be gentler on the wrists ”

("Beim nächsten Mal "Handstandklötze" verwenden, das wäre schonender für die Handgelenke")

RCF10 " even more human voice"
("noch menschlichere Stimme")

RCF15 "A proper sports mat would be much more comfortable for the wrists. "
(Eine richtige Sportmatte wäre deutlich angenehmer für die Handgelenke.")

5.5 DISCUSSION

This chapter presented a study on the effects of encouraging feedback from a SAR on the user's exercise performance. It compared encouragement-feedback for a robot exercise instructor and companion (i.e. Robot Companion with Feedback and Robot Instructor with Feedback).

The previous chapter showed that users exercise longer when paired with a Robot Companion compared to a Robot Instructor or exercising individually. In addition to this result, this study further shows that, if the participants receive encouragement from the robot, they also exercised longer when the robot is not co-actively exercising. This result supports Hypothesis 5.2; participants had a significant performance gain in the Robot Instructor with Feedback compared to the Robot Instructor and Individually Exercising condition.

This outcome provides evidence that encouragement from a robot has a positive effect on the user's exercising performance. Thus, this (simple) interactive motivational capability could be exploited by technical systems to enhance a exercising duration and is particularly important in light of the results that participants exercised longer without enjoying the training less or feeling more exerted.

However, there is no evidence for Hypothesis 5.1. Participants did not exercise longer when paired with a Robot Companion with Feedback compared to a Robot Companion. Supposedly, this result is due to a ceiling effect caused by the selection of the exercises. The exercises were full planks, which puts much force on one's wrist. It is possible, that doing these exercises becomes painful over time and that there is a limit reached where participants might still be motivated to exercise, but their pain causes an early stop. The post-study feedback of the participants backs this speculation that people could not persist the exercises any longer. Therefore, more investigations are needed that explore different kinds of exercises and measure how difficult the participants experience the exercises concerning their task performance.

While the study showed no performance gains in the companion conditions, it found that in both companion conditions the users liked the system more. Furthermore, if the robot is exercising along with the subjects, the information quality of the system is rated higher. One can assume that this is an essential aspect for long-term HRI and could lead to more extended training engagement.

The found results are in contrast to the one from [Irw13]. The authors found that the encouraging feedback from a superior partner mitigates the motivational gain of the Köhler Effect. However, the outcome of this study do not show that the human partner has a lack of motivation when exercising with a robot that is giving encouraging feedback. What are the possible explanations? One pragmatic reason is that robots are not humans. Robots might be perceived as human-like but not entirely as humans. Thus, the mitigating effects might not appear, because human interaction partners are not feeling as judged by the 'more capable' robot when it is trying to motivate the human by encouraging her/him. This assumption is somehow in line with the observation from post-study interviews. People say that they prefer to exercise with the robot because they do not feel judged by it.

To investigate the effects of the perceived judgment of the robot, one would need to evaluate what happens if the robot is explicitly saying that it is evaluating the human. Reasonably, the effects could turn over, and the people would not feel motivated anymore.

One limitations of the study presented in this chapter is again that it only used isolated abdominal plank exercises. Second, this study is missing a comparison of the advantage of the robot against other technological devices that could provide feedback (e.g., smartphone-based applications that provide feedback). Though, it is likely that technology which elicits anthropomorphization from the user, leads to higher exercise adherence and performance than devices that are less anthropomorphic. This was studied by Feltz et al. [Fel14] who showed that the degree of human-likeness of an exercising partners influences a user's exercising duration. Nevertheless, future research should target this issue and systematically evaluate how the degree of human-likeness influences the exercising motivation.

Future research should also investigate the effects of the perceived judgment of the robot, one would need to evaluate what happens if the robot is explicitly saying that it is evaluating the human. Reasonably, the effects could turn over, and the people would not feel motivated anymore.

5.6 CONCLUSION

This chapter presented a study on the motivational effects of encouraging feedback from a SAR that is either working out with a user or just instructing the user. The results show that encouragement has a positive effect on exercising duration compared with the instructor social role, but not with the companion. Thus, this chapter provides partial evidence for Hypothesis 1.2 of this thesis. Therefore, a recommendation for researchers in the field of SAR who want to build robots to motivate people to workout would be, that the robot should either exercise along or give encouraging feedback if it is only instructing an exercise.

Using the data from the previous two experiments enables examining whether the embodiment influences the exercising times on abdominal plank exercises. Therefore, I will compare the gathered data against the results from Feltz et al. [Fel14], who investigated the Köhler Effect during abdominal plank exercises with virtual partners. Comparing the results could allow concluding whether there are any effects regarding the embodiment of the exercising buddy. The following chapter will present this analysis.

6 MOTIVATIONAL EFFECTS OF EMBODIMENT

So far, this thesis presented results on the effects of exercising together with a robot as a partner. However, today's technologies allow for changing the representation of the partner easily; smartphone applications, exercising videos on online platforms or exercising with human partners via internet video calls are possible variations. Those technologies could be used to emulate the feeling of working out together with a partner. Thus, this chapter analyzes the difference between a Socially Assistive Robot and a Virtual Agent.

6.1 INTRODUCTION

Using SAR for tasks where no physical interaction is needed raises the question whether an embodiment is necessary. As most robotics research will know, Embodied Robots always introduce problems regarding their physical ability, deployability or their maintenance. Thus, Virtual Agents (VAs) have a substantial advantage over robots: they are easily deployable, do not have physical limitations and need less care than robots. However, does the interactin with Embodied Robots (ERs) and Virtual Agents (VAs) results in the same social and motivational effects?

Li [Li15] tried to answer this question with a recent research survey which shows that in most cases robots are in favor to VAs. However, there are also works that are showing contradicting results [Ros16; Ken15]. Previous work has investigated the effects of a SAR's embodiment for rehabilitative tasks on a user's evaluation of the robot [Fas13]. In addition to this work, this chapter evaluates the impact of embodiment on a quantifiable motivational measure.

The previous studies showed motivational effects of working out co-actively with a robot or receiving encouraging feedback from a robot. Now, this chapter analyzes whether these types of robots are also increasing a person's motivation to exercise compared to virtual partners.

This chapter is organized as follows: The next section reviews previous and related works. Section 6.3 introduces the study design and data acquisition. Section 6.4 presents the results and Section 6.5 discusses the results.

6.2 RELATED WORK ON EMBODIED SOCIALLY ASSISTIVE ROBOTS

The effects of a SAR embodiment have already been studied in various scenarios. For example, Leyzberg et al. [Ley12] investigate the effects of embodiment during on cognitive tasks. Their work shows that an embodied robot facilitates a learning gain compared to a virtual representation of the same agent. Other works looked at the authority of embodied robots [Bai11]. They showed that people are more willing to obey orders from an embodied agent and give them more personal space. In contrary, works on language learning and teaching found no differences concerning learning gains between agent embodiments [Ken15; Ros16].

The results from Li [Li15] show that a physically present robot compared to a telepresent robot had stronger effects regarding various factors (e.g., the participant's response, persuasiveness, faster response times). This research further found that, compared to virtual agents, co-present robots are more convincing, increase user's attention and response speed, are favored, and users show more positive attitudes towards co-present robots. Regarding the differences between a telepresent and virtual represented robot, the author did not find any differences. Thus, the meta-review concludes that co-present robots have a benefit compared to virtual agents or telepresent robots. Still, other studies show contradicting results [The16; Ros16]. Thus, it remains an ongoing question what the benefits of being physically present are and in which tasks they might have an impact.

Regarding the embodiment effects of SARs designed for exercising or rehabilitation tasks, there is one study investigating the impact of the embodiment in a long-term interactions study [Fas13]. The authors compared a physical robot with its virtual counterpart in a longitudinal study with five 20-minute exercising sessions over a period of two-weeks. Their results provide evidence that users perceive a physically embodied robot as more enjoyable, valuable, helpful and socially attractive compared to the virtual robot. However, these are subjective evaluations from the participants that do not show whether embodied SAR have a measurable motivational effects compared to virtual agents. Thus, this chapter contributes to the ongoing efforts in understanding the effects of embodiment and tries to further close the research gap by showing that embodied robots also increase exercising time. Based on the previous research from [Li15] and [Fas13] I propose the following hypothesis:

HYPOTHESIS 6.1 *A robot companion enhances a human's motivation to persist on an exercise compared to a virtual partner*

To test this hypothesis, I combine the data of the two previously done experiments on abdominal plank exercises with virtual agents and robots. I will evaluate this data to find possible motivational effects in persisting the task due to the different embodiments.

6.3 PLANNED DATA ANALYSIS

To investigate whether embodied robots show an advantage in terms of exercising motivation compared to VAs, I analyze the data of the previous studies from Chapter 4 and Chapter 5 and from Feltz et al. [Fel14]. In Feltz et al. [Fel14] conducted a study to compare the motivational effects of exercising with a humanoid virtual partner with a hardly human-like appearance (Hardly Human Partner (HHP)), a nearly human-like appearance (Nearly Human Partner (NHP)) and with a human partner (Human Partner (HP)) compared to a condition in which the subject is always exercising alone (Individually Exercising (IC)). Their results show that even though it is a small effect, exercising with a virtual partner is more motivating than having no partner. The previous chapters, presented a replication of this study, but with the HP and VA replaced by humanoid robot platform Nao. However, due to the robot's limited Degree of Freedom (DoF), the exercises were changed from forearm planks to full planks.

The replication of this study in the previous chapters showed that a co-actively exercising robot companion leads to higher motivation to persist the exercises than exercising alone, but not for the robot that is just instructing the user. Adding encouraging feedback also resulted in greater exercising performance when the robot is instructing but not when the robot is exercising co-actively. This result might appear due to ceiling effects caused by the difficulty of the exercises. The replication of the Köhler Effect study with VAs allows now to compare the obtained persistence data from both studies. these studies with robots and compare my results with the results from exercising with virtual companions and humans. For the reader's comprehension, this section will summarize the conditions, planned data analysis and used measurements.

6.3.1 *Conditions and Experimental Design*

Subject were in one of nine conditions: IC,IC_2, HHP, NHP, HP, RC, RI, RIF, RCF. These conditions are described in the following (see Figure 18 for the conditions and the study design).

6.3.1.1 *Conditions*

HUMAN PARTNER In [Fel14], the Human Partner (HP) was a college-aged partner whose video was prerecorded.

NEARLY HUMAN PARTNER In [Fel14], the Nearly Human Partner (NHP) was the same video as the HP, but with a computerized effect applied to the video.

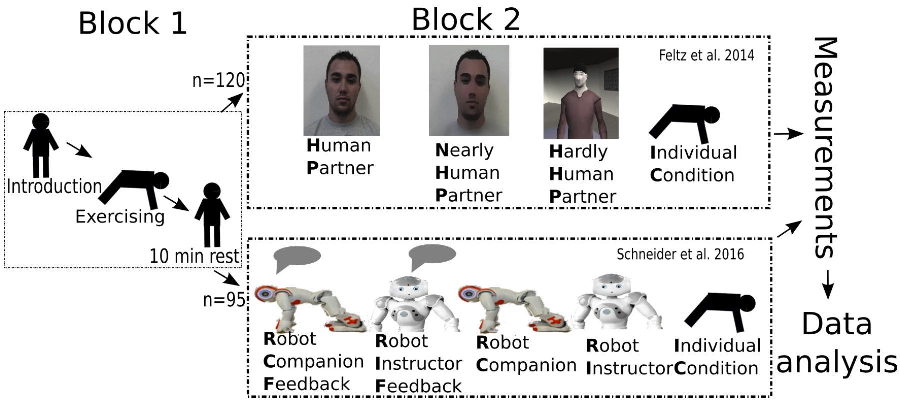


Figure 18: Study procedure, data acquisition and the different representations of the virtual and human partner from Feltz et al. [Fel14].

HARDLY HUMAN PARTNER In [Fel14], the Hardly Human Partner (HHP) were a three-dimensional graphical characters. The character was animated to perform the plank exercises.

ROBOT COMPANION In Chapter 4, the Robot Companion (RC) partner was the humanoid robot platform Nao. Nao's motion was animated using Choregraphe to perform the plank exercises together with the human.

ROBOT INSTRUCTOR In Chapter 4, the Robot Instructor (RI) partner was the same humanoid robot platform. However, instead of exercising co-actively with the human, it simply structures the exercise session.

ROBOT COMPANION WITH FEEDBACK In Chapter 5, the Robot Companion with Feedback (RCF) had the same behavior as the Robot Companion (RC). However, also gave encouraging feedback while exercising.

ROBOT INSTRUCTOR WITH FEEDBACK In Chapter 5, the Robot Instructor with Feedback (RIF) had the same behavior as the Robot Instructor (RI). However, also gave encouraging feedback while exercising.

INDIVIDUAL CONDITIONS Are the baseline conditions were participants exercised a second time individually. Individually Exercising (IC) is the individual condition from [Fel14] and IC_2 is the individual condition from Chapter 4.

In the robot condition studies, participants (n = 95) were randomly assigned to five conditions (IC_2, RC, RI, RIF, RCF). Participants were mostly students (51 male, 44 female; mean age M = 25.4 years; standard deviation SD = 5.6) from Bielefeld university acquired by flyers distributed on the campus. They received seven Euros as monetary compensation. Three participants from the individual condition were excluded. One was an outlier already persisting much less during the first part of the session when the participants were exercising by themselves compared to all other participants. Two other persons were excluded because they were doing the exercises incorrectly. One participant in the Robot Instructor condition had to be excluded from the survey evaluation because the data were missing. Further. two participants from the Robot Companion with Feedback condition were excluded from the survey evaluation because their questionnaire answers were missing. In all other cases, no outliers have been removed.

Feltz et al. [Fel14][1] randomly assigned participants (n = 120) to four exercise conditions (IC, NHP, HHP, HP) with 30 participants in each condition. Participants were undergraduate students (60 females, 60 males; mean age M = 19.41 years; standard deviation SD = 1.52) recruited from a large Midwestern university who completed the experiment for course credit.

The procedure to obtain the data for this analysis was the same for both studies (see Section 4.3.4 or [Fel14]).

6.4 RESULTS

In both studies several different measures were collected. However, not all of them are important for the evaluation in this chapter. Therefore, the analysis will only consider the perception of the partner, activity enjoyment and the persistence on the exercises.

Tests for normality showed that data are not normally distributed. However, a test for homogeneity of variance using a Levene's Test showed a homogeneity of variance. Thus, the data will be analyzed using non-parametric tests (i.e., KW-Test and WC-Test).

6.4.1 *Persistence*

The primary dependent variable was, as in the previous chapter, the average difference persistence time (s) between the two blocks ($\overline{\texttt{Block}_2}$ (s) - $\overline{\texttt{Block}_1}$ (s)). This approach controls for individual differences in strength and fitness and shows possible changes in persistence. At first, a comparison between the exercising times on Block 1 between the two studies has to show that the baseline exercising times are equivalent (see Figure 19).

Though, the exercising time in Block 1 is significantly affected by the studies, $W_s = 3933$, $p < .01$, $r = -.19$. Participants in study [Fel14]

---

1 I would like to thank Feltz et al. for providing their dataset and discussing their work which was supported by grant 1R21HL111916-01A1 from the National Heart, Lung, and Blood Institute

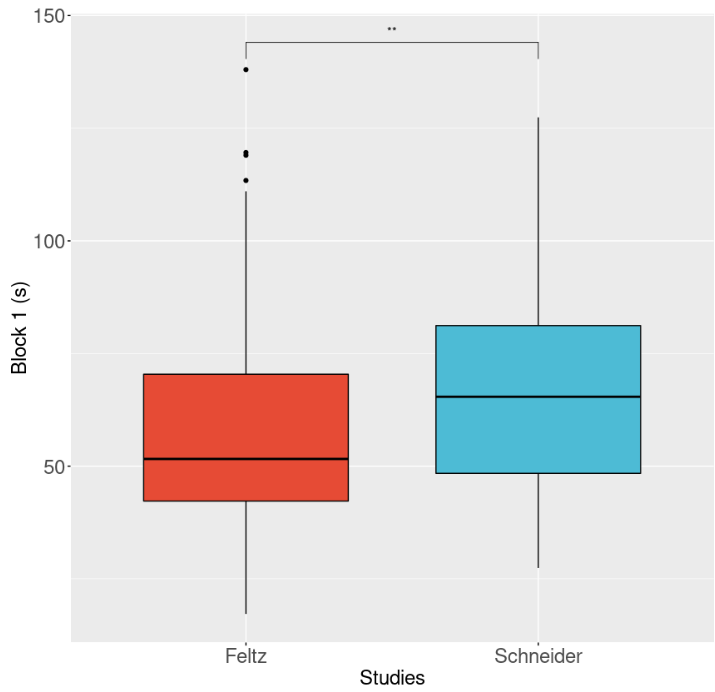


Figure 19: Comparison of the results on average exercising times during Block 1 between the obtained data from Chapter 4 and Chapter 5 and Feltz et al. [Fel14]. Participants exercised significantly longer during Block 1 in the studies from Schneider et al.

exercised on average 7.78 seconds less in Block 1 than subjects in the studies from Chapter 4 and Chapter 5. This difference is possibly due to the changes in the exercise from forearm planks to full planks, which makes the exercises harder to persist but likely more challenging for the user and thus more interesting. Hence, adjustements on the Block 1 measures should control for the difference in exercising times. This value was added to the exercising time of Block 1 for the Nearly Human Partner (NHP), Hardly Human Partner (HHP), Human Partner (HP), and Individually Exercising (IC) conditions.

Figure 20 shows the adjusted results obtained for the average block scores of Block 2 subtracted with the average block score of Block 1. This figure shows the significant difference of the conditions against a base-mean. A KW-Test on the adjusted persistence scores showed a significant main effect for the conditions, $H(8) = 67.93$, $p < .001$. Persistence time in the Human Partner (HP), Robot Companion (RC), Robot Companion with Feedback (RCF) and Robot Instructor with Feedback (RIF) conditions are significantly higher against the base-mean. The IC, IC_2 and Hardly Human Partner (HHP) conditions are significantly lower than the base-mean. A detailed post-hoc analysis after the KW-Test for the persistence time comparing the significant differences between the condition is listed in the appendix in Table C.2.

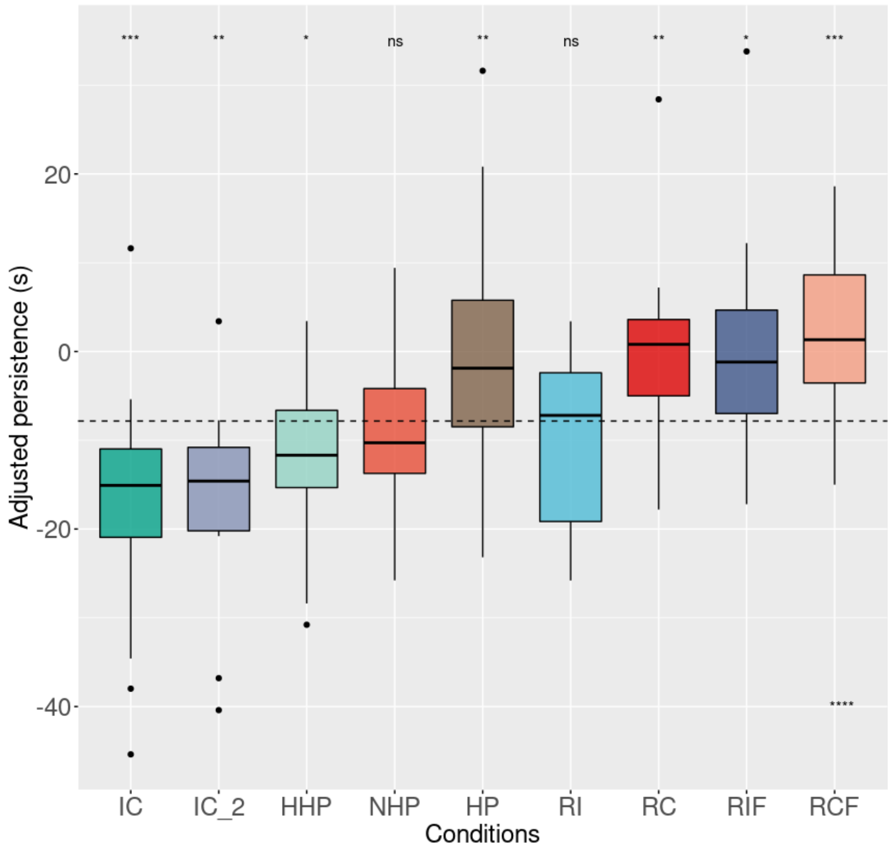


Figure 20: Average adjusted persistence difference (s) between Block 2 and Block 1 for the conditions from Chapter 4 and Chapter 5 and Feltz et al. [Fel14]. Comparison against the mean baseline of all conditions. Results show that participants exercised longer compared to the mean baseline when paired with a Human Partner (HP), Robot Companion (RC), Robot Instructor with Feedback (RIF) and Robot Companion with Feedback (RCF). Participants paired with a Hardly Human Partner (HHP) and participants in the control condition exercised significantly less compared to the mean baseline.

6.4.2 *Physical Activity Enjoyment*

A KW-Test for difference between the conditions on the PAES showed no significant differences, $H(8) = 3.86$, $p = .79$.

6.4.3 *Perception of partner*

The scores on the Godspeed questionnaire are shown in Figure 21. KW-Tests showed significant effects for the perceived animacy of the agents ($H(6) = 17.24$, $p < .01$), anthropomorphism ($H(6) = 21.83$, $p < .01$) and likability ($H(6) = 30.13$, $p < .001$) but not for intelligence ($H(6) = 7.03$, $p = .31$). The Human Partner and Robot Companion are both significantly rated as more animated than then Hardly Human

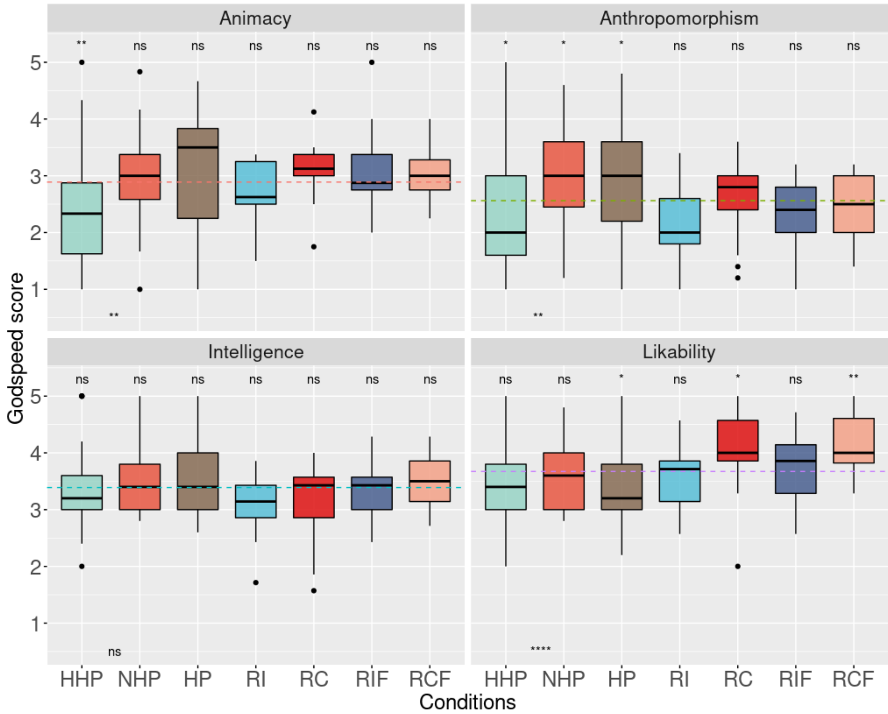


Figure 21: Comparison of the results on the Godspeed questionnaire scales obtbained from Chapter 4, Chapter 5 and Feltz et al. [Fel14]. Comparison against the mean baseline of all conditions show significant differences on the animacy, anthropomorphism and likability scales.

Partner (see Table 1 for the critical differences from the pairwise KW-Tests for the Godspeed scales). The Hardly Human Partner was rated as significantly less anthropomorphic than the Human Partner and the Robot Instructor was rated as significantly less anthropomorphic than the Nearly Human Partner. The Robot Companion and Robot Companion with Feedback were perceived as significantly more likable than the Human Partner and the Nearly Human Partner. Additionally, the Robot Companion with Feedback was rated as significantly more likable than the Nearly Human Partner.

6.5 DISCUSSION

This chapter aims to fill the knowledge gap on the quantifiable motivational effects of exercising with either a co-located robot or with a virtually represented agent. It investigated whether the embodiment of an exercising partner increases the motivation to persist during a conjunctive task. The combined data of the two experiments support Hypothesis 6.1 of this chapter. Participants in the conditions with a robot companion or a robot instructor that gives feedback exercised significantly longer than with a virtual partner on a similar exercising task using the same study design. These results show that the human participants paired with a co-located Robot Companion are more motivated to exercise longer than with a telepresent or virtual representation of the partner. Moreover, the Human Partner does not elicit a stronger motivational effect than the robot. It shows that a robot exercising partner could be at least as motivational as an Human Partner, but more evaluation is needed to assure this.

Therefore, the results need to be further analyzed and the studies replicated due to four differences between them. First, the data have been acquired by two different research groups in different countries. The found effect can be due to cultural differences or subtle difference in the study conduction. Even though I replicated the study as close as possible, it is hard to guarantee that everything went the same as the other researchers did.

Second and most importantly, the study design presented in Chapter 4 introduced a slight change in the exercises. The change from forearm plank exercises to full plank exercises results in different exercising times on Block 1. It is likely that the forearm plank exercises in the study by Feltz et al. [Fel14] were not challenging enough for the participants and thus they stopped the exercising due to boredom and not because of muscle fatigue. This change makes an objective comparison between the data of the two experiments difficult. The proposed solution was to adjust the exercising time on Block 1 of [Fel14] by adding the average difference on this block between the two studies. This approach is reasonable and helps to get an initial view on the motivational effects due to the partner's embodiment, but it still needs to be verified with the same exercises across all conditions.

Third, the virtual representation of the partner was not the same as the used robot. To be sure that the differences are not due to the representation, the study needs replication with a virtual representation of the same robotic platform.

Table 1: Multiple comparison test after Kruskal-Wallis for Godspeed item. This table only shows significant differences.

Godspeed item	comparison	observed difference	critical difference
Animacy	HHP-HP	39.74	35.60
	HHP-RC	41.81	40.58
Anthropomorphism	HHP-HP	36.74	35.06
	NHP-RI	44.21	42.34
Likeability	HHP-RC	42.62	40.85
	HP-RC	50.17	41.40
	HP-RCF	53.70	39.45
	HHP-RCF	46.16	38.87
	NHP-RCF	40.58	40.12

At last, the Human Partner condition was not a co-located partner as in the robot conditions. This difference in co-location could be an explanation why the persistence in the Human Partner conditions was not significantly higher than in the robot conditions. A TV displayed the Human Partner, and the experimenters told the participants that the Human Partner is in a different room connected via a webcam. This difference in the between the human and robot conditions might also influence the results and shows that future research should target this issue. However, it seems to be almost impossible to conduct such an experiment with a co-located human, since the partner has to be always more capable than the participants to implement the Köhler effect. The need for a more trained exercising partner is a hard requirement that seems to be too challenging to fulfill.

The Godspeed questionnaire ratings showed that participants perceived the animacy and anthropomorphism in the robot conditions and the NHP and HP conditions differently. Notably, is the difference in likability between the conditions. The participants not only rated the robot companions as more likable than the virtual partners but also more compared to the human. This difference in perceived likability is an intriguing quantifiable backup for the feedback from participants during post-study interviews in the studies from the previous chapters. Many participants said that they would prefer to exercise with a robot partner than with a human. They argued that the robot is not evaluating or judging them while exercising and thus would feel more comfortable with a robotic partner. This participant feedback supports a future application of SARs as a rehabilitation and exercising tools for people with social anxieties. Thus robots could facilitate the motivational effects of exercising in groups for such a user population.

6.6 CONCLUSION

The question of an agent's embodiment is a crucial question regarding maintenance, cost-benefit ratio, and deployability. Using robots for socially assistive tasks will only be beneficial if they prove to have an advantage compared to other agent representations. Regarding the usage of SAR as exercising partners, I wanted to provide further evidence that a SAR will enhance a user's motivation to exercise and thus potentially increase the physical activity. However, the presented evidence needs further approval with long-term interaction studies and unified benchmarks. Therefore, the research community needs to identify possible exercises to measure the motivational effects and use standardized robot platforms and virtual agent suitable for replicable experiments.

This chapter closes the first part of this thesis on the motivational effects of exercising with SARs. The last three chapters investigated the effects of using robots as exercising partners and showed that they do increase a user's persistence time for an abdominal plank exercise. This exercise is one of many exercises that a robot could do together with the user. Thus, the rest of this thesis looks at ways on how a system could adapt to a user's exercising preference to personalize the HRI experience.

7 ADAPTATION AND PREFERENCE LEARNING

The last three chapters focused on the motivational effects of a SAR in a static abdominal plank exercise. The following part of this thesis looks at the usage of adaptation and personalization for HRI. Choosing the right algorithm for adaptation is challenging and needs careful consideration when designing HRI scenarios. Not every algorithm might be suitable for an online interaction. Hence, this chapter will examine which kind of algorithms might be suitable to create adaptable social robots for sports assistance.

NOTE Parts of this chapter are published in S. Schneider et al. "Exploring Embodiment and Dueling Bandits for Preference Adaptation in Human-Robot Interaction." In: *Proceedings of the 26th IEEE International Symposium on Robot and Human Interactive Communication*. 2017

7.1 INTRODUCTION

The past chapters motivated the work of this thesis by applications where robots assist users during conventional rehabilitation, health care or learning programs (i.e., stroke-rehabilitation [Tap07a], dieting [Kid08a] or teaching [Ley14a]). Though, the main focus is on the usage of robots in a simple exercising scenario and the motivational effects of exercising together with a robot. Looking at the goal of increasing PA level, one sees that this goal requires an extended commitment of the user. Reaching rehabilitation, teaching and health care goals is rarely possible through a single session intervention. Hence, tools such as robots have to apply methods that engage a user in long-term interaction. Furthermore, they eventually have to provide personalized interaction because every person has unique desires and preferences. While highly specialized physicians, therapists or coaches are trained to provide individualized personal interaction for each person, robots are still far from such capabilities. Thus, robots can only function as tools that trained personnel could use as an additional therapeutic measure. However, researchers are working on the required steps to implement social robots in long-term use cases [Lei13]. A review of different works concludes that four major building blocks for robots need to be addressed to engage users in long-term interaction: behavior, adaptation, empathy and design [Lei13]. While all of these aspects are important for engaging users in long-term intervention, the focus of this and the following chapter is on the adaptation aspect.

The problem of adaptation and personalisation in HRI has already been targeted by several researchers [Tap07a; Mit05; Lei12]. In these works, the robot adapts its verbal and non-verbal behavior to the user's preferences to increase its acceptance. Though, it is unknown what kind of adaptation might be necessary to extend a user's commitment to interact with a SARs. Mainly, the purpose of a SAR is to assist users on a task. These tasks can have, e.g., different difficulties, categories, activities, or durations. While parameters like the personality of the robot, proxemics or feedback types were already studied in previous works, the preference for different tasks or task categories has not received much attention yet. Since different kinds of task instances might lead to the same rehabilitation-, learning- or coaching goal, techniques from Preference Learning (PL) could be used to learn a user's task preferences over time.

Preference Learning is widely spread in the domains of recommendation systems in online platforms [De 10]. Among these algorithms to optimize search results or provide customized advertisements Mulit-armed Bandits (MABs) are a popular instance to solve the learning problem (e.g., UCB [Aue02]). By showing a user different kinds of advertisement or search results, the algorithm learns a user's preferences through explicit (e.g., ratings) and implicit feedback (e.g., clicking behavior). The primary concern of this chapter is whether this kind of online algorithm is suitable to learn a user's preference in HRI for a socially assistive tasks. More precisely, it will look at a particular kind of Mulit-armed Bandit (MAB) for Preference Learning (PL), i.e., the k-armed dueling bandit could be applied in an interaction [Bus14]. In contrast to standard bandit learning techniques, this approach does not require a numerical reward function but uses a user's comparative feedback, which has been shown to be more reliable.

The focus of the past chapters was on social assistance during exercising and sportive activities (see Chapter 4). They presented a scenario that used only one kind of exercise (i.e., abdominal plank exercise). However, it is possible to implement different exercises on a robot companion which might suit the preferences of different users better. Then the goal would be to learn a user's exercise category preference. This opens the issue of choosing a set of exercises, which is why I only consider a set of categories that are suitable for a robot to accompany or instruct a user in the near future. The chosen exercising categories are: strength, cardio, endurance, stretching and relaxation/meditation. Using this approach, this chapter will work on the question whether a dueling bandit algorithm is suitable for HRI and can learn a user's preferences.

Though, using robots that suggest exercises comes with an objection: the embodiment of the agent could influence the users so that they more agree with the robot's proposed preference ranking. Why might this be a problem? The literature survey mentioned in previous chapters showed "that a co-present robot is more persuasive, receives more attention and is perceived more positively than a VA even when the behavior of the robot was identical to the behavior of the VA and when both agents had similar appearance" [Li15, p. 465]. Thus, the user might stick more to social norms when interacting with the embodied robot and agrees more with the suggested exercise preference. Hence, this chapter investigates whether the embodiment of the learning agent influences a user's perceived likability, intelligence, and persuasiveness during a PL task.

I draw the hypothesis that:

HYPOTHESIS 7.1 *An embodied agent will increase the user's agreement with the learned preferences, the perceived intelligence and likability compared to a virtual representation or no agent representation.*

The remaining chapter is organized as follows: Section 7.2 gives an overview of related work in the field of adaptation and personalization in HRI. Section 7.3 describes the PL framework. Section 7.4 explains the study design. Section 7.5 presents the results which are discussed in the last section.

7.2 RELATED WORK ON ADAPTATION AND PERSONALIZATION IN HRI

There are two trends for personalization and adaptation in HRI. One trend is to adapt the robot's behavior based on interactive machine learning techniques. These approaches mostly utilize Reinforcement Learning (RL) with user feedback and sensor data (e.g., [Tsi16; Lei12; Tap07a; Bar15; Mit05]). Other approaches create user models to adapt the robot's assistance and behavior (e.g., [Sek13; Ley14a]) or rely on techniques from recommendation systems like Collaborative Filtering (CF) [Lim13]. One of the significant applications of personalization in HRI is concerned with the adaptation of the robot's social behavior to match the user's personality or desires. In these cases behavior adaptation is often based on personality matching to adjust interaction parameters like proxemics, speed, vocal content, robot's appearance or dialog topics (e.g.,[Tap07a; Bar15; Lee12; Rit17]). The goal of these adaptation techniques is to enhance the user's acceptance of the robot which should increase the user's commitment to interact with the system in the long run. Other works include approaches like RL, MAB or Bayesian Network (BN) to adjust session parameters or generate supportive and emphatic behaviors (e.g.,[Tsi16; Lei12; Ley14a; Cha12; Hem17]). In these scenarios, personalization targets the user's learning gains, therapy success or enjoyment during games. Table 2 gives an overview of different research directions in the field of HRI.

All approaches show that an adapted robot behavior is preferred by the user and leads to better learning outcomes and a higher robot acceptance. However, most of the works include some implicit direct feedback from the user (e.g., sensor data), require the user to fill out a questionnaire, or use a Wizard of Oz (WoZ) to personalize the robot behavior. Furthermore, many RL approaches need to have a numerical feedback to learn a user-adapted policy. This approach can be a bottleneck of the implementation because direct feedback is not available or it is based on the engineers understanding of how to represent the numerical feedback. In some applications, it might be difficult to determine a numerical reward function, or it might be challenging how to obtain the actual reward. Hence, this work extends the literature by evaluating how Preference Learning, in this case bandit learning, can be used to personalize the human's HRI experience. Therefore, I draw from research that extended MAB learning scenario to a dueling bandit learning scenario [Bus14]. In those scenarios, the agent learns the user's preference by presenting the user two items. Qualitative preference feedback of the user then represents the feedback. Based on this approach the agent can learn the user's preference for a given set of items without the need of having a numerical reward.

Table 2: Research on adaptation and personalization in HRI.

Work	Method	Variables	Adaptation goals
[Mit05]	RL	body signals	interaction distance, gaze, motion speed and timing
[Tap07a]	RL	personality traits, nu. of performed exercises	interaction distances, speed, and vocal content
[Lei12]	MAB	user's valence	emphatic behavior
[Cha12]	RL	speech, user state, activity state	providing instructions, empathy or help
[Lee12]	WoZ	snack choice patterns, usage patterns, robot's prior behavior	personalized speech topics
[Lim13]	CF	semantic knowledge, event episodic knowledge and emotion	individual student's motivation to prevent negative emotions
[Ley14a]	BN	puzzle state	personalized tutoring
[Bar15]	MAB	numerical user reward	robot's light animation
[Tsi16]	RL	performance, session state	adjust time of movement, move to next exercise or encourage user
[Gor16]	RL	engagement and valence facial expression	maximize long-term learning
[Hem17]	RL	gaze behavior, speech, game state	memory game assistance
[Rit17]	RL	social signlas	robot's personality

7.3 PREFERENCE LEARNING FRAMEWORK

This section will briefly introduce Preference Learning as a formal problem from the perspective of MABs.

PL is a subfield of machine learning that aims to learn predictive models from previously observed information (i.e., preference information) [Für11]. In supervised learning, a data set of labeled items with preference information is used to predict preferences for new items or all the other items from a data set. In general, the task for preference learning is concerned with the problem of learning to rank.[1] There are many different approaches for Preference Learning. It can be solved using supervised learning, unsupervised learning and also Reinforcement Learning. Since there exists no particular data set I could use for supervised or unsupervised learning, it is challenging to build a model that can predict preferences from previously observed information. Therefore, I am focusing on how the system can learn an initial preference relation for a given item set without any prior information (i.e., the cold start problem). Thus, I am trying to solve the PL problem using online methods from MAB algorithms or more precisely Dueling Bandit algorithms.

7.3.1 *Dueling Bandits: Problem Statement*

The dueling bandit problem consists of $K (K \geqslant 2)$ arms, where at each time step $t > 0$ a pair of arms $(\alpha_t^{(1)}, \alpha_t^{(2)})$ is drawn and presented to a user. A noisy comparison result w_t is obtained, where $w_t = 1$ if a user prefers $\alpha_t^{(1)}$ to $\alpha_t^{(2)}$, and $w_t = 2$ otherwise. The distribution of the outcomes is presented by a preference matrix $P = [p_{ij}]_{K \times K}$, where p_{ij} is the probability that a user prefers arm i over arm j (e.g.,$p_{ij} = P\{i \succ j\}, i, j = 1, 2, .., K$).).

The goal of the PL task is, given a set of different actions (e.g., different sport categories), to find the user's preference order for these categories by providing the user two α_i and α_j and update the user preferences based on the selection of the preference between $\alpha_i \succ \alpha_j$ or $\alpha_i \prec \alpha_j$.

Thus, the challenge is to find the user's preference by running an algorithm that balances the exploration (gaining new information) and the exploitation (utilizing the obtained information). In this thesis, I am using the Double Thompson Sampling (DTS) algorithm presented in [Wu16]. An algorithm is sketched in Algorithm 1. Since there are several implementations to solve the dueling bandit problem I need to answer the question of why I have chosen this specific kind of algorithm.

1 Learning to rank can be further divided into three main problems regarding the types of information observed: a) label ranking, b) instance ranking, c) object ranking. For a detailed description of these different main problems please refer to [Für11].

Two reasons mainly drive this decision, the state of the art algorithms at the time of this study were DTS, Relative Minimum Empirical Divergence (RMED) and its successor ECW-RMED [Kom15; Kom16]. Both perform reasonably well regarding their asymptotic behavior. However, at this point I am not interested in the long-term run of these algorithms but in the initial phase. If one takes a look at the first steps of these algorithms, one can see a significant difference between them that likely influence the HRI experience. RMED and ECW-RMED both have an initial phase where all possible pairs are repeatedly drawn for some time. From an algorithmic perspective this is reasonable, but looking at it from the viewpoint of the interaction, this would lead to systematic comparisons that could result in boredom and even annoyance when the interaction partner is seemingly interrogating the user for her/his preferences. Thus, I assume that the DTS algorithm is more useful for HRI (especially for the initial contact between the trainee and the robot coach), because it does not rely on a systematic comparison of all possible pairs.

7.3.2 *System and Algorithm Implementation*

Figure 22 gives an overview of the learning framework. At each time step, the algorithm selects two candidates from the preference matrix (Step 1). In my implementation,[2] I used the DTS algorithm to select the two candidates [Wu16]. However, I neglected the exploitation phase, because I am only interested in the exploration phase where the algorithm obtains new information. Based on the selected categories, two specific exercises are selected randomly from an exercise database.[3] This database holds six different exercises for each sports category. Following, these exercises are presented as text on display. In the robot conditions, the text is additionally accompanied by speech and gestures from a virtual or real Nao (Step 2). Subsequently, the user can give relative preference feedback by selecting the preferred exercise (Step 3). This feedback is then used to update the preference matrix accordingly (Step 4). After twenty iterations the system gives the user a ranking about the so far learned preference matrix. The sports category which wins against most other categories is presented as first followed by the other categories in descending order by their number of wins.

7.4 STUDY DESIGN

As stated in the introduction, this chapter investigates two aspects: the feasibility of PL algorithm for HRI, and whether the embodiment

2 I reimplemented and adapted the algorithm for my purpose in python 2.7 based on code provided by Wu et al. [Wu16]

3 https://www.mongodb.com/,visited on 3/23/2017

Algorithm 1 DTS for Copeland Dueling Bandits as presented in[Wu16]

Init: $\mathbf{B} \leftarrow \mathbf{0}_{K\times K}$; // B_{ij} is the number of time-slots that the user prefers arm i to j.
for t=1 to T **do**
// Phase 1: Choose the first candidate $\alpha^{(1)}$
  $U := [u_{ij}], L := [l_{ij}]$, where $u_{ij} = \frac{B_{ij}}{B_{ij}+B_{ji}} + \sqrt{\frac{\alpha \log t}{B_{ij}+B_{ji}}}, l_{ij} = \frac{B_{ij}}{B_{ij}+B_{ji}} - \sqrt{\frac{\alpha \log t}{B_{ij}+B_{ji}}}$, if $i \neq j$, and $u_{ii} = l_{ii} = \frac{1}{2} \forall i$; $//\frac{x}{0} := 1$ for any x
  $\hat{\zeta} \leftarrow \frac{1}{K-1}\sum_{j\neq i} 1(u_{ij} > 1/2)$; //Upper bound of the normalized Copeland score
  $C \leftarrow \{i : \hat{\zeta}_i = max_j \hat{\zeta}_j\}$;
  for $i, j = 1, ..., K$ with $i < j$ **do**
    Sample $\Theta^{(1)}_{ij} \sim Beta(B_{ij}+1, B_{ji}+1)$
    $\Theta^{(1)}_{ji} \leftarrow 1 - \Theta^{(1)}_{ij}$
  end for
  $\alpha^{(1)} \leftarrow \underset{i\in C}{\text{argmax}} \sum_{j\neq i} 1(\Theta^{(1)}_{ij} > 1/2)$ //Choosing from C to eliminate likeley non-winner arms; ties are broken randomly
// Phase 2: Choose the second candidate $\alpha^{(2)}$
  Sample $\Theta^{(2)}_{i\alpha^{(1)}} \sim Beta(B_{i\alpha^{(1)}}+1, B_{\alpha^{(1)}i}+1)$ for all $i \neq \alpha^{(1)}$, and let $\Theta^{(2)}_{\alpha^{(1)},\alpha^{(1)}} = 1/2$
  $\alpha^{(2)} \leftarrow \underset{i:l_{i\alpha^{(1)}} \leqslant 1/2}{\text{argmax}} \Theta^{(2)}_{i\alpha^{(1)}}$ //Choosing only from uncertain pairs.
// Compare and Update
  Compaire pair $(\alpha^{(1)}, \alpha^{(2)})$ and observe the results w;
  Update **B**: $B_{\alpha^{(1)},\alpha^{(2)}} \leftarrow B_{\alpha^{(1)},\alpha^{(2)}} + 1$ if $w = 1$, or $B_{\alpha^{(2)},\alpha^{(1)}} \leftarrow B_{\alpha^{(2)},\alpha^{(1)}} + 1$ if $w = 2$
end for

of a robot affects the perceived intelligence and likability of the robot during a PLs task. To study these aspects, all participants interacted with the same algorithm running in the background. However, the embodiment of the system differed across conditions (see Figure 23). Participants were randomly assigned to one of the following conditions: computer only, virtual Nao, real Nao. The computer only condition included a Graphical User Interace (GUI) with buttons and a text area. The text area displayed an introductory text, exercise comparisons, explanations regarding the exercises and finally the learned preference ranking. The user can select her/his preferred exercise by pressing the according button. In the robot conditions, either a virtual Nao (displayed using Choregraphe) was presented on the computer display or a real Nao was standing next to the computer. Besides this manipulation, the system behavior was the same for all conditions.

NaoQi API was used for text-to-speech synthesis and gesture generation for the virtual as well as the real robot.

7.4.1 *Study Procedure*

Each participant arrived individually at the lab and had to read and sign a consent form. The experimenter told the participant that she/he would interact with a system that will learn their exercise preferences by displaying different names of exercises and that she/he can select the one she/he is favoring. If the name of an exercise is unknown to the participant, she/he can get more information from the system regarding the category the exercise belongs to (i.e.,"push-up is a strengthening exercise", "running belongs to endurance sports", and so on). After the instructions, the experimenter guided the participant to the experimental room and told that she/he should exit the lab after the interaction has finished. This is where the manipulation happened. In the room was either only the computer, the computer with a virtual Nao or a real Nao present. The experimenter did not explain anything else regarding the virtual or real robot. During the study, the system iterated through twenty exercise comparisons and in the end presented the learned exercise preferences. Afterward, the

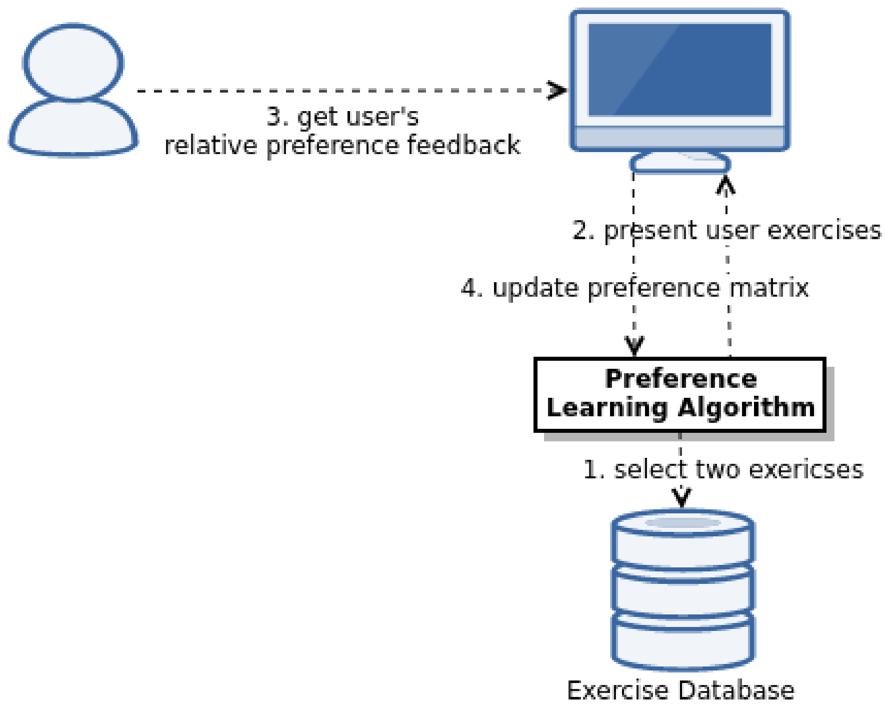
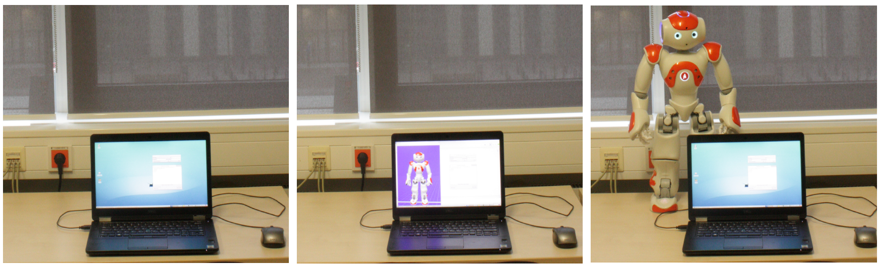


Figure 22: Overview of the system interaction flow for the preference learning study. 1) The algorithm selects two exercise from the database, 2) The exercises are displayed on the screen, 3) The user can give preference feedack, 4) The algorithm updates the learned preference matrix.



(a) GUI only (b) Virtual Nao (c) Real Nao

Figure 23: The three conditions for the study on preference learning and agent's emodiment. The preference learning agent is either represented as a a) Graphical User Interace (GUI), b) virtual Nao or c) or an real Nao. In the virtual and real Nao condition, the GUI was also available to interact with the system.

participant left the room and answered a survey. Finally, the participant received a monetary compensation (4 Euro) and was debriefed.

7.4.2 *Participants*

53 subjects from our campus participated in this study. They were equally distributed between the three conditions (computer: 18, virtual: 18, robot: 17; 18 male and 34 female). In each condition were six male participants. The average age was $M = 25.34$ with $SD = 5.47$.

7.4.3 *Measurements*

General demographic measurements were used as in the previous session, as well as perception of the agent using the Godspeed questionnaire, and the openness and information quality. Additionally, the following measurements were used:

PERSONALITY Participant's personality was assessed using the Neo-FFI-30 personality scale [Kör08]. I used all five sub-scales Neuroticism, Extraversion, Openness, Agreeableness and Conscientiousness. See Appendix D.10 for the used scale.

SYSTEM USABILITY System's usability was measured by the System Usability Scale (SUS) with ten items on a 5-point Likert [Bro96]. See Appendix D.4 for the used scale.

INTRINSIC MOTIVATION AND INTERACTION QUALITY Intrinsic motivation was assessed using a short German version of the Intrinsic Motivation Inventory proposed by [Dec06] (see Appendix D.12). Furthermore, I asked the participants to rate the quality of the interaction on a 5-point Likert scale ranging from 'the interaction with the system was difficulty' to 'the interaction with the system was easy'. The scales are listed in Appendix D.13.

learned preference quality To gain insights into the perceived PL satisfaction, participant's satisfaction with the learned preference were measured on a four-item 5-point Likert scale (e.g.,"The system has learned my preferences"). Additionally, if the participants were not satisfied with the learned preference, they could provide their preference order which I will use later for the system evaluation.

preference ranking error To asses the quality of the obtained preference rankings, two ranking error functions were used: D_{PE} which is the position error distance and D_{DR} which is the discounted error. Given a set of items $X = x_1, ..., x_c$ to rank, r as the user's target preference ranking and $\hat{r}$ as the learned preference ranking, both r and $\hat{r}$ are functions from $X \to \mathbb{N}$ which return the rank of an item x, The position error is defined as follows

$$D_{PE}(r, \hat{r}) = \hat{r}(argmin_{x \in X} r(x)) - 1 \quad (1)$$

The idea of this distance measure is that the target item (i.e., the highest ranked item from r) should appear as high as possible in the learned preference ranking $\hat{r}$. Thus, this distance gives the number of wrong items that are predicted before the target item. The discounted error is defined as follows

$$D_{DR}(r, \hat{r}) = \sum_{i=1}^{c} w_i \cdot d_{x_i}(\hat{r}, r) \quad (2)$$

where $w_i = \frac{1}{\log(r(x_i)+1)}$. This distance measure gives higher ranked items from r a higher weight for the distance error between the rankings.

7.5 RESULTS

7.5.1 *Manipulation Check*

Several one-way ANOVAs or KW-Tests showed no differences for hours spent for sport per week ($H(2) = 1.08$, $p = .58$), age ($F(2,50) = .63$, $p = .63$, $\eta_p^2 = .01$), previous experience with interactive systems or robots ($F(2,50) = 2.2$, $p = .12$, $\eta_p^2 = .08$, $\omega^2 = 0.04$), neuroticism ($\alpha = .79$, $H(2) = 1.32$, $p = .51$), openness ($\alpha = .72$, $H(2) = 0.00$, $p = .99$), agreeableness ($\alpha = .48$, $F(2,50) = 1.34$, $p = .27$, $\eta_p^2 = .05$, $\omega^2 = .11$), extroversion ($\alpha = .62$,$F(2,50) = .74$, $p = .48$, $\eta_p^2 = .02$, $\omega^2 = .1$) and conscientiousness ($\alpha = .76$, $H(2) = 0.22$, $p = .9$) between the different conditions. Thus, the randomization seems to be successful.

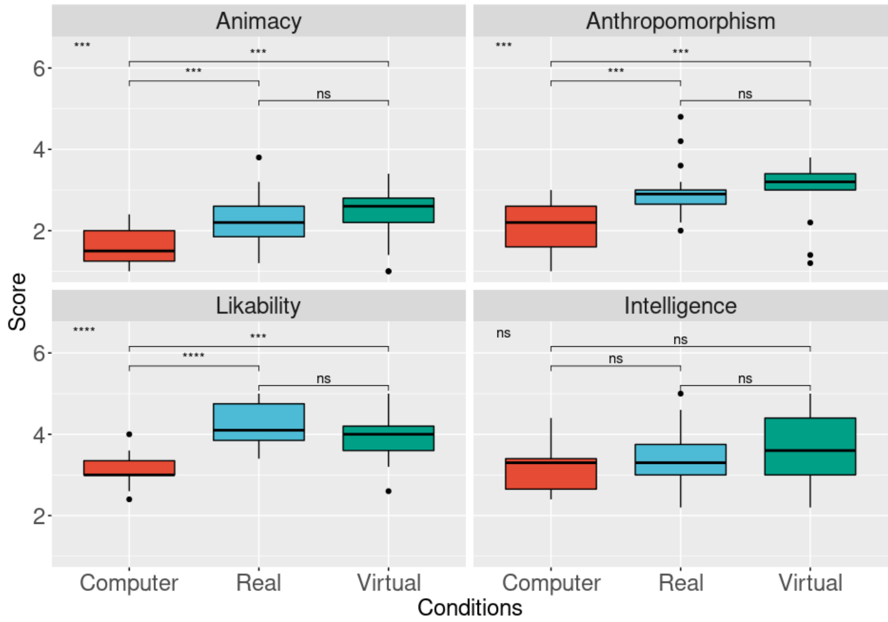


Figure 24: Boxplot showing the Godspeed Questionnaire ratings for the study on the embodiment of a preference learning agent.

7.5.2 *Godspeed questionnaire*

One-way ANOVAs showed significant effects for the perceived animacy ($\alpha = .79$, $F(2, 50) = 9.27$, $p<.001$, $\eta_p^2 = .27$, $\omega^2 = .48$), anthropomorphism ($\alpha = .83$, $F(2, 50) = 10.31$, $p < .001$, $\eta_p^2 = .29$, $\omega^2 = .51$), likability ($\alpha = .89$, $F(2, 50) = 21.04$, $p < .001$, $\eta_p^2 = .45$, $\omega^2 = .65$), but not for perceived intelligence ($\alpha = .84$, $F(2, 50) = 1.22$, $p = .32$, $\eta_p^2 = .04$, $\omega^2 = .09$).

Results from post-hoc pairwise comparisons using t-test with pooled SD and Bonferroni correction for the different items are listed in Table C.4 in the appendix. Table C.3 shows the mean values and standard deviations. I found no significant differences between the real and virtual condition for animacy, anthropomorphism and likability. Not surprisingly, I found significantly different ratings between the computer condition and the other conditions for animacy, anthropomorphism, likability. The computer was rated significantly less on all the Godspeed scales, except for intelligence (see Figure 24 for all significant differences between the conditions).

7.5.3 *System Usability, Intrinsic Motivation and Interaction Quality*

SYSTEM USABILITY SCALE A one-way ANOVA for the System Usability Scale (SUS) ($\alpha = .85$) revealed significant difference across the conditions, $F(2, 50) = 4.59$, $p < .05$, $\eta_p^2 = .15$, $\omega^2 = .34$. Pairwise comparisons using t-tests with pooled SD and Bonferroni correction revealed significant differences between the computer and the virtual agent condition ($p < .05$), but not between the computer and the real robot ($p = .53$) and the real and virtual robot($p = .3$). See Table C.3 in the appendix for the mean values and standard deviation).

INTRINSIC MOTIVATION A KW-Test showed that intrinsic motivation ($\alpha = .84$) was significantly affected by the conditions, $H(2) = 8.66$, $p = .014$. A post-hoc test with focused comparisons of the mean ranks between conditions showed that intrinsic motivation were not significantly different in the virtual condition ($difference = 1.06$) and the computer condition compared to the virtual condition ($difference = 12.40$) with a critical difference of 12.50 for both comparisons. However, the intrinsic motivation were significantly higher in the robot condition compared to the computer condition ($difference = 13.47$) with a $critical\ difference = 12.32$.

INTERACTION Perceived interaction was not signficantly different between the conditions as determined by a KW-Test, $H(2) = .73$, $p = .7.1$

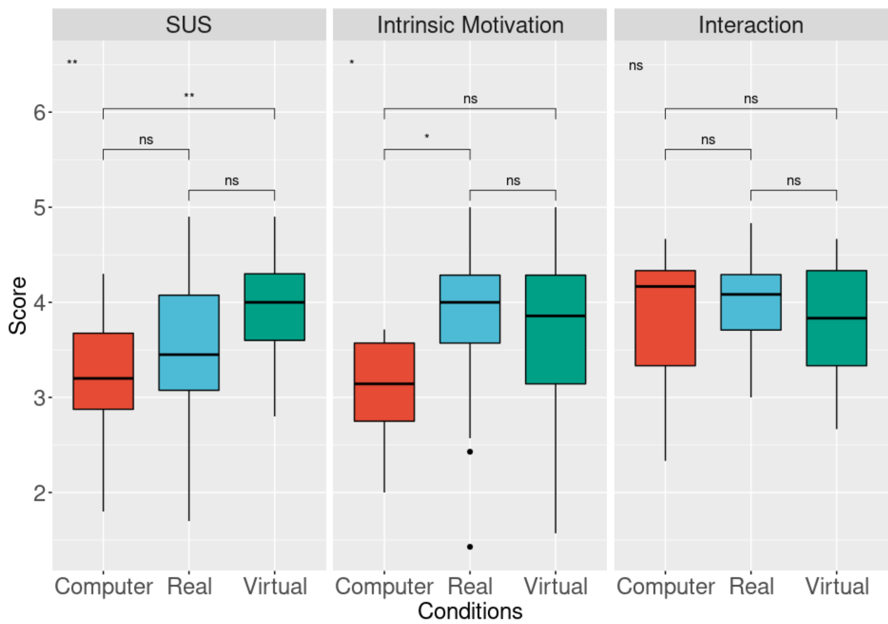


Figure 25: Boxplot showing the user ratings for the System Usability Scale (SUS) and intrinsic motivation scale between the GUI (C), virtual robot (V) and embodied robot (R) conditions for the preference learning study.

7.5.4 *Preference Learning Evaluation*

SUBJECTIVE RATINGS ON INFORMATION QUALITY AND OPENNESS TO INFLUENCE A one-way ANOVA for the perceived information quality ($\alpha = .76$,$F(2, 50) = 1.93$, $p = .15$, $\eta_p^2 = .07$, $\omega^2 = .18$) and

Table 3: Frequency distribution of first and second ranked sport preferences

	Exercises				
Condition	Stretching	Cardio	Endurance	Strength	Relaxation
computer	10	7	11	4	5
virtual	10	7	11	4	5
robot	5	4	11	8	4

comparison quality (α = .78, $F(2,50)$ = .45, p = .63, η_p^2 = .01, ω^2 = .14) showed no significant differences across the conditions. The perception of the learned preference quality (α = .9) did not differ significantly across the conditions, $F(2,50)$ = 2.24, p = .12, η_p^2 = .08, ω^2 = .21). Finally, the openness to influence (α = .88) was also not influenced by the embodiment, $F(2,50)$ = 0.02, p = .98, η_p^2 = .00, ω^2 = .2.

PREFERENCE RANKING ERROR Frequencies for the learned sport preferences are summarized in Table 3. A Fisher's Exact Test (FET) revealed no statistical significance ($p = 0.79$, FET). The ranking errors D_{PE} and D_{DE} are depicted in Figure 26. A KW-Test revealed no significant differences for D_{PE} (P = .55) and D_{DE} (P = .32) between the conditions. To measure the effectiveness of the PL algorithm, I simulated a *random* condition where the ranking is selected by a randomized algorithm. I used the obtained ranking preferences from this study as target criteria, computed the position and discounted error accordingly. Including this random condition, I receive significant differences for D_{PE} ($H(3)$ = 44.09 , p < .0001) and D_{DE} ($H(3)$ = 44.99 , p < .0001). See Table C.5 in the appendix for post-hoc analysis and critical differences. The randomized algorithm significantly ranks worse on the collected data set than the used DTS implementation in all conditions.

7.6 DISCUSSION

This chapter explored the effects of the system's embodiment on the user's evaluation of a Preference Learning system as well as the suitability of a dueling bandit framework for personalization in HRI. Thus, it provides a contribution to the ongoing research on the effects of embodiment. Even though there exists work that reports that embodied robots are found to be more persuasive, enjoyable and entertaining [Li15], the debate is still ongoing by works that show contradicting results (e.g., [Ros16]).

The conducted study examined how the embodiment of the system influences the user's perception of a PL system. The reported results from Section 7.5 do not support the hypothesis 7.1. The users rated

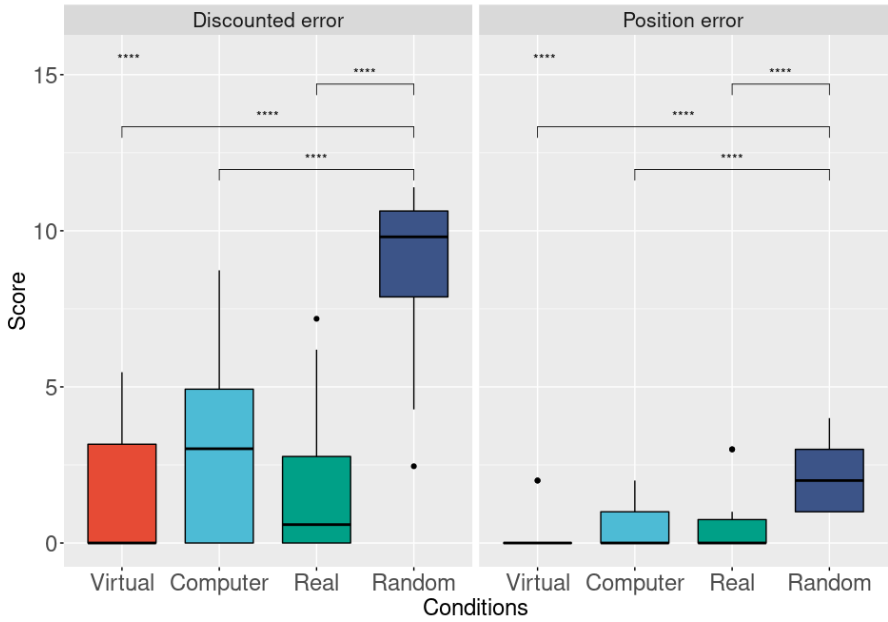


Figure 26: Boxplot showing the preference ranking errors (discounted and position error) for the GUI, virtual robot and embodied robot compared to a simulated condition that selected the exercises randomly.

real and the virtual robot alike on the Godspeed questionnaire. Thus, this results are in contrast to the conclusion from [Li15]. Also, the ratings for the System Usability Scale (SUS) and the Intrinsic Motivation Scale did not significantly differ between the virtual and the real robot. Though, the results show that the embodiment of the system (both virtual and real robot) significantly increased the participant's likability of it compared to the computer only condition. Furthermore, the embodiment increased the user's intrinsic motivation in the real robot condition compared to the computer condition. Regarding the user's preference ranking satisfaction, the embodiment also did not influence the subjective evaluation of the ranking quality. Hence, participants in all conditions were equally satisfied with the suggested preference ranking. This indicates that the perceived quality of the ranking is independent of the embodiment which is a desirable outcome for a preference learning task. The user should agree with a ranking due to the underlying algorithm and not to the embodiment of the agent.

However, there are also several reasons that could hinder an effect of the presence of the robot. The real and virtual robots were an additional interface of the GUI. Hence, the real robot might not have been such a salient cue as expected compared to the virtual one, because participants looked more on the screen (where the virtual one was displayed in the according condition) than on the robot. Thus, the effects might have been different if an external monitor would display a virtual robot with the same height as the real Nao. However, other researchers comparing the embodiment of a robot in a socially assistive task also manipulated the presence and kept a GUI alongside the robot (e.g., [Ley12]). Since the users did not evaluate the intelligence of the robot differently across the conditions, I assume that the perceived intelligence is not influenced by its embodiment but by the underlying algorithm. Thus, more research on the influence of embodiment and algorithmic design on the perceived intelligence is needed.

Moreover, this study investigated the effectiveness of the PL framework for HRI. First of all, the results indicate that the users were satisfied with the system's suggested preference ranking. Their agreeableness with the learned preferences is relatively high and the calculated ranking errors low. To the best of my knowledge, this is the first work exploring the dueling bandit learning approach in HRI. The results indicate that dueling bandit learning might be a suitable framework for personalizing HRI experiences without cognitively overloading the user, needing a numerical reward function or taking much time for the learning process.

In this study, I have also assessed the user's personality. Research from psychology shows a correlation between personality and exercising preference [Rho06]. This correlation might be useful to overcome the problem of a cold start for a Preference Learning algorithm. In future research, this data set could be used to evaluate how adaptation might be accelerated using the personality trait information of a person.

7.7 CONCLUSION

This chapter tested whether dueling banding learning works in a HRI situation and whether the system's embodiment influence the experience for the user. It provides support that such a preference learning approach might be suitable for future applications. However, I could not find support for the Hypothesis 7.1. The virtual and real robot were evaluated equally by the users and the embodiment did not affect the perceived intelligence and preference ranking satisfaction. Thus, it indicates that the usage of an embodied robot in a preference learning task in an applied scenario might be appropriate and will be the main topic of the following chapter.

8 ADAPTATION AND ADAPTABILITY

The last chapter introduced a prototype HRI scenario where a robot learns the user's preferences. Today's recommendation systems usually suggest a user new items based on Collaborative Filtering (CF) or Content-based Filtering (CBF)[1] [Pre10; Mo000]. Though, techniques like Collaborative Filtering require an extensive database to predict a user profile. While it might be possible to have some form of Collaborative Filtering for social robots in the future, to the date of this thesis there is no such a system available. Additionally, such systems also often suffer from the cold start problem. Hence, for this thesis I target an approach that has no prior user knowledge and learns a user's preference.

The past chapter reviewed the requirements and tested the suitability of such a learning phase in a test scenario. However, the scenario was designed to investigate whether the embodiment of the system is influencing the user's acceptance of the system's preference ranking and to test the usability of such a cold start learning phase in HRI. This chapter will introduce an experimental design to look at the effects of an initial contact with an adaptive system as an exercising companion. I will describe how the chosen personalization method is (to an extent) associated with the degree of control of a system. From this distinction of the level of control of a system (concerning adaptivity) I will introduce a study that investigates how the degree of user control influences the trust in and alliance with the system. Thus, this chapter investigates Hypothesis 1.4 of this thesis.

8.1 INTRODUCTION

Future scenarios of socials robots envision a highly personalizable system that is flexible and adapts itself to the user's preferences, and at best knows what the user wants without the need to program the desired behavior. Having adaptive robots is an essential requirement for long-term interaction with social robots [Iol13]. Since it is not possible to anticipate every potential user and pre-program the system for their needs, robots will potentially need to have capabilities to adjust to different users. A robot might enhance the interaction experience by adjusting behaviors that match the personality of the user (e.g., [And15]). While adaptation to enhance the interaction were successfully implemented in web-based applications (e.g., recommender systems on Amazon, Google, eBay), it is still a challenging question

1 There are alos hyrid approaches

for a social robot that is not attached to a user database that enables techniques like CF. Thus, social robots are faced with the cold start problem, which requires the system to gather initial user data. Hence, having an adaptive system comes with some difficulties:

First, querying the user for information in real time HRI might cost much more than in online applications were a short click to a link is sufficient. Cakmak et al. [Cak10] showed that a constant stream of questions in a Learning by Demonstration task is annoying for the users.

Second, it might cause concerns for the HRI experience, when the robot starts to make autonomous personalization decisions. Accordingly, it could be sufficient for the human to adjust the required behavior instead of having the robot adapt by itself. This different types of possible customization strategies would influence the autonomy of the system, which might influence the interaction experience in different ways.

Based on the theory of anthropomorphization, an autonomous adaptive system could create an unexpected experience for the user [Epl07]. This unexpected experience could increase a user's associated degree of anthropomorphization of the robot. Furthermore, this higher degree could increase the credibility of the system and might influence the associated relationship with it or trust in it. In contrast, a system that is controlled and adjusted by the user should increase the match between the robot's behavior and the user's expectation, which might reduce the anthropomorphic effects. The investigation of these two aspects is the core of this chapter. I try to find an answer to the question: What effects have different types of personalizable robots on the user's acceptance, relationship and motivation to interact with the system?

To investigate the effects of the LoA of the system, this chapters presents a study that compares the effects of having an adaptive robot and an adaptable robot.

The difference between adaptive and adaptable robots will be explained in Section 8.2 along with the concepts of autonomy and relationship, which might be important variables when looking at the adaptivity of a system. Section 8.3 introduces the system design and Section 8.4 explains the study design to test the effects of a robot's different personalization mechanisms. Section 8.5 presents the results of the study, which are discussed in Section 8.6. Finally, Section 8.7 gives a conclusion of this chapter.

8.2 ADAPTATION, CONTROL AND RELATIONSHIP

As mentioned earlier, this chapter is looking at a variety of different concepts (i.e., adaptation, control and relationship/trust) that were not introduced yet, but might be important for the interaction with social robots in the future. Grasping these concepts is challenging because they have different meanings and definitions in different disciplines (e.g., philosophy, psychology, economics, biology). However, contrasting them and defining will be helpful to understand this chapter and the impact of the results. Nevertheless, the sketching of these concepts cannot be exhaustive due to the broad range of disciplines using these terms. Therefore, I will concentrate and look on the terms from a computer science and psychology perspective.

8.2.1 *Adaptation: adaptivity vs. adaptability*

In computer science adaptation is the process of adjusting the behavior of an interactive system to individuals using information about them. Even though computer software or robots are running through many software design cycles, it is hard to anticipate the requirements for every possible user. The goal of the adaptive process is to minimize the discrepancy between the user needs and system behavior after the deployment.

This adaptive process can either be automatically initiated by the system, in this case the system is adaptive, or users can adjust the system by themselves, in this case the system is adaptable.

Adaptation can be based on different user profiles (e.g., age, gender, personality), different times (e.g., morning/evening, days of the week, summer/winter) or other user characteristics (e.g., mood, expertise over time). The previous chapter already introduced some of the related work in the area of adaptation and personalization in HRI. Therefore, these works are not listed here again. However, there is a lack of knowledge in the research community because few works compared different possibilities to match a robot's behavior to the user needs.

Most studies investigated the implementation of an adaptive process [Tap08b; Lei11; Tsi16; Mit08]. Although some have compared adaptive robots with experimental baseline conditions (e.g., [Ley14b]), to the best of my knowledge, no works looked at the effects of robot-initiated personalization or human-initiated personalization. Though no works like this exist yet, it is reasonable to argue that the users could be in control and adjust the robot behavior to their preferences too. Leyzberg et al. [Ley14b], for example, investigated the effects of a robot that gives personalized lessons to the user. These tutorials are selected by the robot's decision. However, instead of doing this automatically, the user could have also asked the robot for a specific lesson. Both strategies might lead to user interaction satisfaction, but the underlying difference in decision making is fundamental. One can interpret the different strategies as either more transparent or as more competent. Generally, the question of whether to build an adaptive or adaptable system raises the concern of who is in control and how does it affect the interaction experience. The issue of who is in control is associated with the Level of Automation (LoA) of the system.

8.2.2 *Level of Automation*

Asking for the different Level of Automation of a system leads to looking at an agent's capability to act and react based on information on their own without any other external control instances. An autonomous agent acts based on the information it receives from its sensors, knows in which state it is and makes a decision accordingly which is associated with an agent's action [Rus16, ch. 2].

The Level of Automation of an agent is, depending on the task and environment of the agent, altered by introducing humans in the agent's control loop. It becomes essential were robots are carrying out delicate tasks (e.g., Lethal Autonomous Weapons (LAWs)). There are various frameworks that can be used to classify the Level of Automation of a system (e.g., [She78; End95]). However, most recently, Beer et al. [Bee14] have proposed a framework to classify the Level of Automation for HRI.

In general, systems can be categorized as a) human-in-the-loop systems where the human has to approve a control decision by the autonomous agents, b) human-on-the-loop where the human is informed about decision but the agent would carry out a decision if the human operator is not interfering or c) human-off-the-loop, where a human cannot interfere with the agent's decisions.[2]

The relevance to consider different Level of Automation are apparent in sensible domains such as military operations or medical applications (e.g., surgery or medicine dispenser), but (yet) less apparent in socially assistive domains (rehabilitation or teaching). Nevertheless, also social situations will require to understand whether a social robot should act on its own, semi-self controlled or in full human control. Therefore, it will be crucial to understand the effects of different Level of Automation on the interaction experience. In the course of this chapter, I am interested in the effects of whether the robot exercising companion is in control to choose the next exercises or the users can decide which exercises they want to do. The question of whether the Level of Automation is appropriate and which effects it will have on the interaction will be related to the relationship and trust between the users and the SAR [Bee14].

8.2.3 *Relationship and Trust*

If a robot makes autonomous decisions, it is a significant issue whether humans trust the robot's capabilities [Fre07]. In cases where systems have more control and decide on their own, either directly or

2 Earliest examples of hands-off-the-loop agents are land and naval mines.

controlled by the feedback of users, the relationship between the human and robot might influence the trust users put in the system's decision. Marriam-Webster [Mar18] defines trust as "the assured reliance on the character, ability, strength or truth of someone or something" [Mar18].

This quite vague definition is more specified in the community of HCI where trust is defined as "the extent to which a user is confident in and willing to act on the basis of, the recommendations, actions, and decisions of an artificially intelligent decision aid" [McA95, p. 25]. As Madsen et al. [Mad00] state, this definition "encompasses both the user's confidence in the system and their willingness to act on the system's decision and advice" [Mad00, p. 1]. Thus, it already incorporates a notion of user trust regarding the willingness to take a system's recommendations into account.

To understand how trust influences HRI, Hancock et al. [Han11] reviewed different applications were trust is an important factor when robots and humans are working together in a team. They state that it is an essential aspect of industry, space or warfare applications. Though, one can argue that trust will also be crucial to understanding for social tasks due the rise of SAR for rehabilitative, therapeutic or educational tasks. The authors found several factors influencing trust in HRI, which are related to the human, the robot, and the environment. However, the robot related factors were the most important ones in their meta-review. They found that important factors influencing the associated trust are the human's perception of the system's behavior, adaptability, competence, and performance [Han11]. Considering how types of personalization change the Level of Automation and how this might alter the perceived trust, I question how the manipulation of the Level of Automation (for example how the system adapts or can be adapted) influences the associated competence and the perceived trust in the system.

Rau et al. [Rau13] investigated the influence of a social robot's Level of Automation on the user's trust in a robot and the influence on decision making. They manipulated the robot's Level of Automation by either giving the human the possibility to make a team decision and the robot could suggest a different decision (low autonomy) or the robot makes the team decision and the human can either reject or accept this decision (high autonomy). They hypothesized that a highly autonomous robot would increase the associated trust. Their results show the influence of an autonomous robot on human's decision making, but in contrast to the hypothesis, people rated that they trust the low autonomous robot more.

Other works investigated how perceived anthropomorphization influenced perceived trust [Way14]. Waytz et al. [Way14] found that the degree of anthropomorphization of an autonomous vehicle is associated with higher trust in its competence. This indicates that the perceived level of competence might also influence the associated trust, However, there is, to the best of my knowledge by this date of this thesis, no other works that investigated the influence of a social robot's Level of Automation on the perceived trust in addition to the work of Rau et al.

8.2.4 *Hypotheses*

Based on the reviewed literature, this chapter investigates the effects of preference adaptation with different Level of Automation in HRI. I propose the following hypotheses:

HYPOTHESIS 8.1 *Users perceive an adaptive robot as more competent than an adaptable robot.*

Due to the robot's initiative and control of the interaction people will be likely to associate the robot with higher competence. Since users do not have to control the robot on their own, the robot creates the impression of proactively deciding on its own. I hypothesize that this different level of perceived competence is associated with the perceived trust or relationship with the agent.

HYPOTHESIS 8.2 *The relationship to an adaptive robot is rated better than to an adaptable robot.*

This hypothesis is mainly based on the assumption that the users will more likely trust a system that is perceived as competent. Even though research from Rau et al. [Rau13] did not show any significant effects on perceived trust depending on the Level of Automation, I still hypothesize that the Level of Automation will affect the associated trust. It is likely that Rau et al. [Rau13] did not find an effect on the trust because the robot was only a marginal partner that was not important for the task. Instead the scenario of this chapter the robot is not just a member of the team but also an instructor and coach for different exercises. Therefore, the trust and alliance will be an important feature for the relationship between the user and the robot. Furthermore, since I hypothesize that the participants in the conditions will perceive both the competence and trust differently, I also hypothesize that:

HYPOTHESIS 8.3 *The associated trust between the conditions is significantly mediated by the perceived competence of the system.*

This hypothesis is based on the results from Hancock et al. [Han11], that trust is associated with the human's perceived competence of the system.

Additionally, low trust is often associated with the misuse or disuse of an autonomous robot [Bee14]. Previous works hypothesized that if the people do not trust a robot they stop using it. This trust in the competence of an interaction partner to achieve a desired goal is also highly critical between a client and a therapist [Hor89]. Perceived higher compentence increases the trust in the relationship to achieve a common goal. Thus, if people do not feel the competence in the relationship to achieve a common goal, they do not trust the therapist and are more likely to stop the therapy or intervention.

Thus, I also hypothesize that:

HYPOTHESIS 8.4 *An adaptive robot increases the participant's motivation to engage in a second interaction compared to an adaptable robot.*

8.3 SYSTEM DESIGN

To investigate these hypotheses, I have implemented a distributed system. The system is composed of a database of different exercises for Nao, a session manager executing the exercises with Nao, a simple computer vision system using the Kinect sensor and a preference learning algorithm.

8.3.1 *Exercise Database*

As previously explained, the exercising preference is very individual from person to person. Thus, for the aim of this study I develop a system that provides a variety of different exercises. I have chosen 25 exercises in total from 5 different categories: strength, stretch, cardio, Taichi, and meditation. This set of exercises tackles one of the open issues mentioned in Section 2.1.2: A wider set of exercises for SAR. Previous work often looked at a single type of exericse like arm movements. The approach of using a spectrum of different exercises might show that people can perform various exercises together with a robot.

Table 4 presents the list of the chosen exercises. They have been selected based on a variety of criteria: a) the possibility to animate and execute them on Nao (i.e., Nao cannot jump.) b) the difficulty that users can perform them (i.e., exercises should not be too challenging for the participants) c) the full body weight workout: The exercises should challenge the full embodiment of the robot (i.e., laying down, balancing, standing).

All of them have been animated on Nao using Choregraphe. The robot's instructions for these exercises and the required joint configurations have been configured using the framework presented in [Sch17a]. Nevertheless, I did not want to include coaching capabilities in the system. Hence the system is only configured to instruct the user and not to give them assistive feedback on their execution while exercising nor give them encouraging feedback like in Chapter 5.

Table 4: Used exercises for the presented study.

Strength	Stretch	Cardio	Meditation	Taiji Drills
Push up	Neck	Jumping Jacks	The boat	Golden rooster
Squats	Triceps	Front Lunge	9 breathes	Rainbow
Crunches	Hip	Side Lunge	Deep relaxation	Punch
Superman	Quadriceps	Boxing	Inner light	Parting kick
Bridge	Side	Mountain Climbers	The peace sign	Lifting water

8.3.2 *Session Manager*

The session manager is also implemented using the presented framework from [Sch17a]. The system waits for a user to be present in the room. Depending on the distance, it asks the participant to come closer. Afterward, in the adaptive condition the algorithm selects two exercises from the database and Nao instructs the user to do the exercises. Following, it asks the participant which exercises she or he prefers (preference feedback). Initially, I used the internal speech recognition of Nao. However, prototype experiments showed that the speech recognition capabilities are below an acceptable recognition rate, which is why I manually inserted the user's feedback using a WoZ style.

When Nao performs the exercises, it moves away from the initial position. I have implemented a simple marker based localization strategy, however the robot needed too long to localize in the room and move to the correct position. Since it is a significant disturbance for the HRI experience, I also have implemented a WoZ controller to move the robot to the correct position manually after each exercise. The primary interaction flow for the preference learning conditions is as follows: Based on the current user's preference database, the algorithm selects two exercises (the algorithm is the same as the one presented in the previous chapter, see section 7.3), then the session manager runs the exercises. Afterward, the robot asks the user which of the exercises she or he prefers. The wizard listens to the user's feedback using an installed microphone in the experimental room and feeds the user's input back to the session manager. The robot acknowledges the decision by repeating the chosen exercise. The preference learning algorithm updates the user's preference database and selects the next exercises based on the current user preference.

8.4 STUDY DESIGN

I conducted a study with a between-subject design (adaptive robot vs. adaptable robot) where participants were randomly assigned to one of two conditions.

8.4.1 *Conditions*

ADAPTIVITY The robot in the adaptivity condition used the algorithm described in Section 7.3. At each time step, the system selected two exercises based on the algorithm and executed them consecutively with the user. Afterward, the user was asked to give a preference statement regarding the exercises. This behavior repeated for 14 exercises (or seven iterations). After the 14 exercises, the system asks whether the user wants to continue exercising for two more exercises or quit the experiment. After the two additional exercises, the interaction was finished by the robot. It stated the user's learned preferences and thanked for the participation. I limited the additional exercises to two exercises, due to battery concerns and overheating of joints.

ADAPTABILITY The robot in the adaptability condition did not use any preference learning algorithm and did not select the next exercises autonomously. The robot verbally listed the possible exercising categories in a randomized order and the user could choose the exercise category she or he wants to experience. Thus, the user was in control of the exercise session and could choose the exercise category she or he prefers.

8.4.2 *Participants*

Participants ($n = 40$; average age $M = 26.02$, $SD = 5.48$, 13 female and 7 male in the adaptivity condition; 12 female and 8 male in the adaptability condition) were mostly university students that were acquired by information on the campus and social media. The majority of the participants were naive robot user and had no background in computer engineering or programming.

8.4.3 *Procedure*

Participants arrived at the lab individually. First, they had to sign a consent form. Then, the experimenter led the participants to a room where they can change their clothes. Later, they had to do a pre-questionnaire asking for their self-efficacy beliefs in doing exercises like strength, meditation, stretching, cardio or taichi and got instructions for the next steps of the study. They were told to enter the lab and follow the instructions of the system. Until this point, the participants did not know that they will be interacting with a robotic system. I neglected this prior information to not bias the participants or raise false beliefs. Then the participants entered the lab without the experimenter. The interaction happened for approximately 40 minutes, and the experimenter monitored the experiment from a control room. After the interaction finished, participants had to answer a questionnaire and had a short interview. Finally, they were debriefed and received 8 Euros for their participation.

8.4.4 *Measurements*

In this study, some measurements were the same as before. I assessed the Negative Attributes Towards Robots Scale (α = .8), PAES (α = .91), SUS (α = .84), team perception (α = .52), openness (α = .94) and cooperation (α = .4) as described in Chapter 5. Additionally, I used the measurements listed below.

PERCEPTION OF THE PARTNER Additionally to the Godspeed questionnaire, participants were asked to rate the perception of the robot on the new Robotic Social Attribute Scale (RoSAS). This scale includes the perceived warmth (α = .85), competence (α = .77) and discomfort (α = .76) on a 9 point-based Likert-scale [Car17]. See Appendix D.2 for the used scale.

MOTIVATION To have an additional measure to see whether people are interested in exercising a second time with the robot, I let the participants opt-in for voluntarily exercising with the robot again without monetary compensation. Participants were asked at the end of the questionnaire to enter their email address if they want to exercise again.

WORKING ALLIANCE The Working Alliance Inventory (WAI) (α = .91) is a measure commonly used in helping relationships to assess trust and belief in a common goal of helping that a therapist, clinician or coach has for another [Hor89]. See Appendix D.8 for the used scale. This measure has recently been used in HCI and HRI studies [Bic05; Kid08b] for assessing the relationship and trust between the human and a SAR. The scale is divided into three subscales assessing the bond (α = .9), the goal (α = .84), and the task (α = .84).

8.5 RESULTS

8.5.1 *Manipulation Check*

The data was checked for differences in the participant's previous experience with technology, their average weekly exercising activity and the attitudes towards robots. Previous experience (W_s = 162, p = .14), exercising activity (W_s = 237, p = .45), as well as NARS ($t(37.7)$ = 1.77, p =.08) were not significantly different between the conditions.

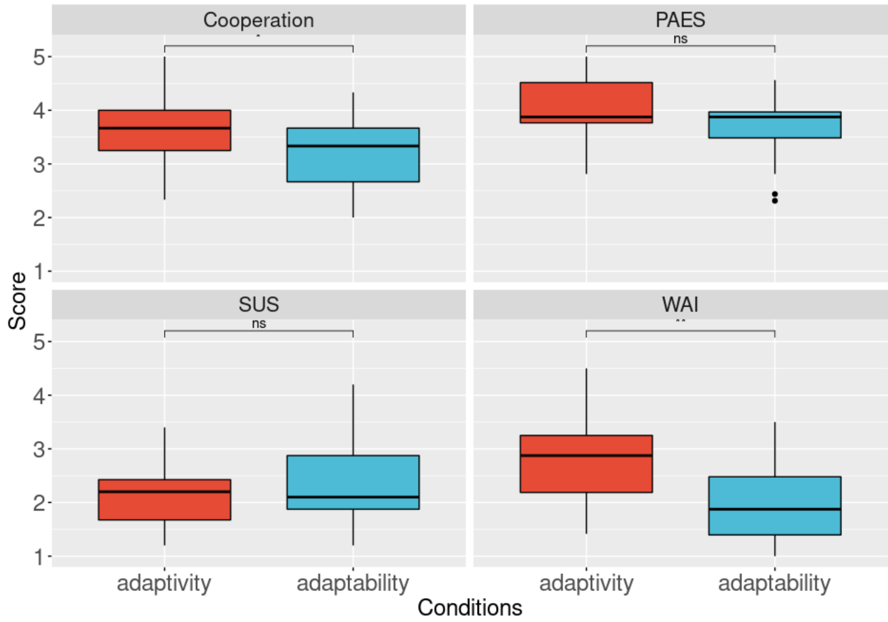


Figure 27: Boxplot showing the user ratings for the cooperation, Physical Activity Enjoyment Scale (PAES), System Usability Scale (SUS) and Working Alliance Inventory (WAI) scales for the adaptability adaptivity conditions in the adptation study.

Cooperation

A Welch's twosample t-test showed a significant difference between the conditions for cooperation, $t(37.91) = -2.43$, $p < .05$, $d = -.79$. The adaptive system has been rated as significantly more cooperative ($M = 3.7$, $SD = .66$) than the adaptable system ($M = 3.2$, $SD = .63$). However, this result can only be carefully considered due to the low internal consistency of the cooperation scale ($\alpha < .4$).

System Usability Scale

A Welch's two sample t-test for differences on the SUS scale showed no significant differences, $t(37.615) = .92$, $p = .36$.

Openness

A WC-Test test showed no significant differences between the two conditions on the openness to influence scale, $W_s = 206$, $p = .92$, $r = -.10$.

Physical Activity Enjoyment Scale

A WC-Test test showed no significant difference between the two condition on the PAES scale, W_s=168.5, $p = .28$. PAES was not higher in

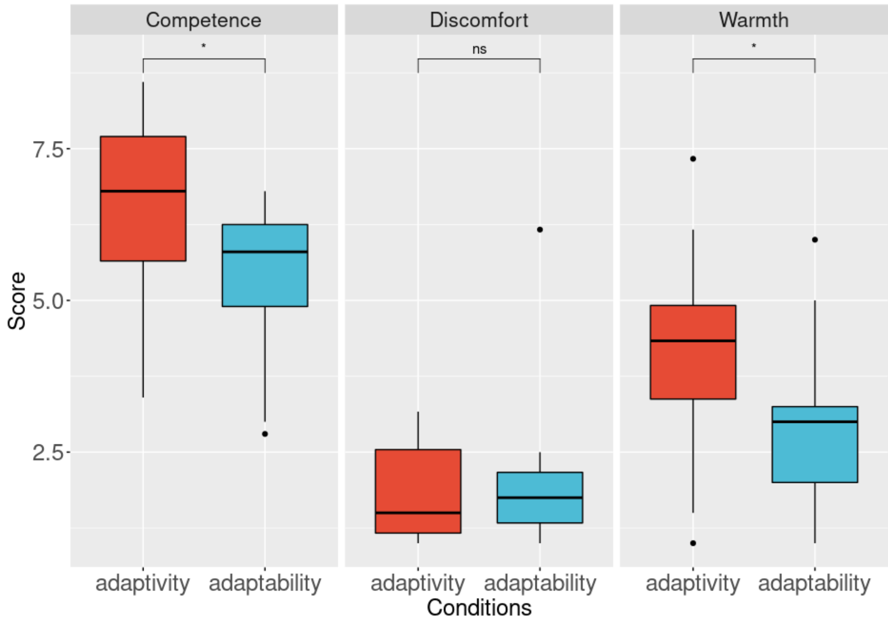


Figure 28: Boxplot showing the user ratings for the Robotic Social Attribute Scale for the adaptability and adaptivity conditions in the adaptation study.

the adaptive condition (M = 4.0, SD = .53) compared to the adaptable condition (M = 3.6, SD = .63).

Robotic Social Attribute Scale

The results for RoSAS are plotted in Figure 28. The detailled analysis is listed in the following.

WARMTH A Welch's two-sample t-test showed a significant effect regarding the warmth subscale of the ROSAS scale, $t(36.22) = -2.47$, $p < .05$, $d = -.82$, $r = .38$. The adaptive system is perceived as warmer (M = 4.08, SD = 1.62) than the adaptable system (M = 2.93, SD = 1.29).

COMPETENCE Regarding the perceived competence of the systems, a Welch Two Sample-test showed a significant difference between the conditions, $t(34.54) = -2.49$, $p < .05$, $d = -.85$, $r = .39$. The adaptive system is perceived as more competent (M = 6.55, SD = 1.67) than the adaptable system (M = 5.4, SD = 1.2).

DISCOMFORT There were no significant difference between the conditions using a Wilcox-ranked sum test, $W_s = 210$, $p = .79$, $r = .27$.

Team Perception

There was no significant difference regarding the team perception between the adaptive condition ($M = 3.38$, $SD = .57$) and the adaptable condition ($M = 3.09$, $SD = .7$), $t(38.0) = -1.48$, $p = .14$.

Wish to Repeat Interaction

The ratio for participant's wish to voluntarily repeat the interaction is depicted in Figure 29. Pearson's Chi-squared test showed that participants opted more often to exercise again with the adaptive condition ($\chi^2 = 4.8$, d.f. = 1, $p < .05$). This effect is however not persistent when using Yates' continuity correction ($\chi^2 = 3.3$, d.f. = 1, $p = .067$).

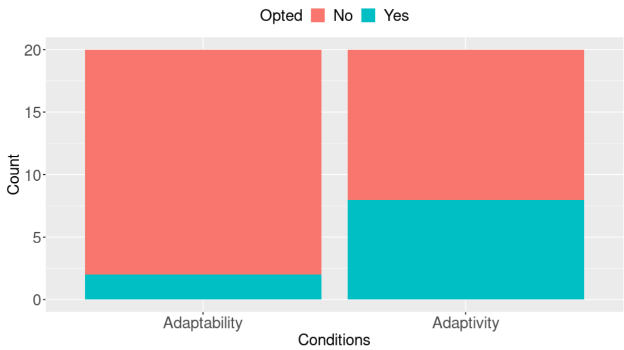


Figure 29: Counts for participants that opted to voluntarily exercise again for each condition of the adaptation study.

Godspeed questionnaire

The results on the Godspeed questionnaire did not differ significantly between the conditions for any of the subscales. Table 5 reports the results of this analysis.

Table 5: Multiple comparison test after Kruskal-Wallis for ratings on the Godspeed questionnaire for the adaptivity study.

Godspeed item	adaptivity M, SD	adaptability M, SD	test results
Animacy	3.09, .62	2.98, .6	$t(37.98) = -.56$, $p = .57$, $d = .18$
Anthropomorphism	2.62, .87	2.35, .62	$t(34.29) = -1.12$, $p = .26$, $d = .36$
Intelligence	3.63, .63	3.32, .54	$t(34.29) = -1.13$, $p = .26$, $d = .53$
Likability	4.73, .37	4.52, .58	$W_s = 166$, $p = .33$, $r = -.15$

Working Alliance Inventory

Results of the Working Alliance Inventory are depicted in Figure 27. A Welch Two Sample t-test revealed significant difference between the conditions, $t(36.89) = -2.99$, $p < .01$, $d = -.99$, $r = .44$. The adaptive system has been rated significantly higher on the alliance inventory ($M = 2.8$, $SD = .93$) than the adaptable system ($M = 1.99$, $SD = .76$).

MEDIATION ANALYSIS To assess whether the condition's effect on overall trust was statistically mediated by perceived competence, I used non-parametric bootstrapping method based on the method from Preacher et al. [Pre08] and coded condition as adaptability = 0, adaptivity = 1. This analysis confirmed that perceived competence statistically mediated the relationship between adaptive condition and overall trust in the robot (Average Causal Mediation Effects (ACME) = .48, $p < .05$, 95% CI = .1 to .91; 10,000 resamples; see Figure 30) with no direct effect of autonomy of the system (Average Direct Effect (ADE) = .31, $p = .09$) and significant total effect ($p < .001$).

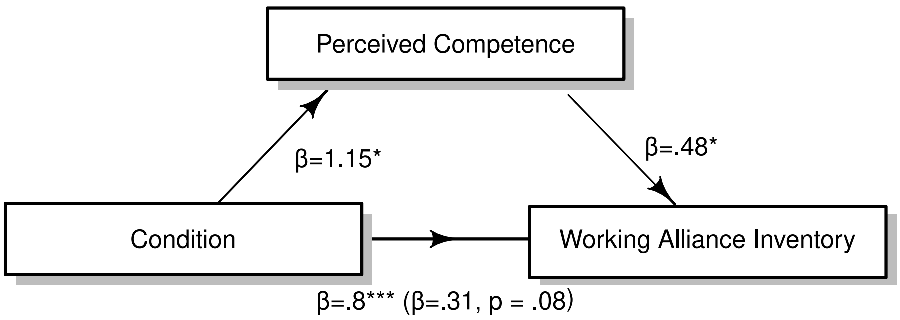


Figure 30: Standardized regression coefficients for the relationship between conditions and user's relationship with the robot as mediated by the user's perceived competence of the robot. The standardized regression coefficient between the conditions and the Working Alliance Inventory, controlling for perceived competence, is in parentheses.

8.6 DISCUSSION

This chapter investigated how a system's type of personalization mechanism alters the user's perception of it. It presented an empirical design to investigate the effects of different personalization techniques on alliance and competence of a SAR. The robot was either controlled by the preference feedback of the user or entirely controlled by the user in an exercising scenario.

I hypothesized that different a Level of Automation alters the user's perceived competence of the robot and relationship with it. The results present evidence that the robot is perceived as more competent, which can be seen in a significant difference between the conditions on the RoSAS subscale. This evidence supports Hypothesis 8.1: An adaptive robot is perceived as more competent as an adaptable robot.

The result regarding the perceived relationship to the robot also supports the hypothesis Hypothesis 8.2: Participants had a stronger alliance with the adaptive robot measured by the WAI. This result supports the hypothesis that Rau et al. [Rau13] had, but could not find evidence to support it. The proposed mediation model provides evidence why the different conditions affected the perceived alliance. The Level of Automation increased the perceived competence of the system which in turn increased the alliance to it.

Other researchers showed that anthropomorphism alters the trust in an autonomous vehicle [Way14]. They manipulated the agency of a vehicle which alters the perceived trust and found evidence that people have higher trust in the vehicles competences when it is perceived as more anthropomorphic. However, they have not measured the perceived competence as an independent mediator in their study. Thus it remains an open question whether the manipulation of the anthropomorphism alters the perceived competence in the system and therefore changes the associated trust.

However, this study could not show that the different Level of Automation alter the perceived anthropomorphism, as measured using the Godspeed questionnare. This lack of difference is probably because I manipulated the Level of Automation and not explicityl the anthropomorphism of the robot. The study used the same robot in both conditions as well as the same speech output. Thus it seems like the difference in Level of Automation doe not affect the user's perceived anthropomorphism. Therefore, more studies are required to investigate the effects of anthropomorphism and perceived competence on the associated trust and alliance with a social robot. Moreover, it is interesting to note that the Godspeed questionnaire showed no differences between the conditions while the RoSAS does. Reasons for this are discussed in Section 9.2.2.

One limitation of the interpretation of the results above is the short interaction time during the study. Trust and alliance are commonly build up over repeated interactions between two people. Therefore, the results on the effects of trust need to be interpreted with caution. Additionally, the scale used in this experiment is primarily designed for measuring the trust in the client-therapist relationships. Therefore, the results might be different, if I have used a scale that is more focused on the trust in the technical competence of the system. Still, the trust in the relationship is an essential part for long-term HRI and especially for use cases where the human and robot partner are working towards a long-term goal like increasing physical activity. Therefore, I would suggest to alter the competence of the system explicitly and additionally evaluate the trust in the system's technical competence in future studies.

Finally, I could find partial evidence for Hypothesis 8.4. Participants in the adaptivity condition opted more often to voluntarily exercise a second time. This result is probably due to the interest in a system that tries to personalize the interaction by itself. It might raise curiosity and participants are interested to see what other exercises the system can offer or whether the system can effectively learn the user preference. However, this result is only marginally significant after applying a continuity correction. To be sure whether this effect is genuinely significant higher sample size is needed.

8.7 CONCLUSION

As to the best of my knowledge, this is one of the first studies in the field of HRI to investigate the effects of different Level of Automation (LoA) types of a robot on the ascribed competence and perceived relationship with the agent. This study is also one of the first to investigate the usage of a dueling bandit preference learning in online interaction with robots in the context of exercising scenarios. I conclude from this study that these kinds of preference learning might be suitable for personalization in HRI, because they work with preference feedback from the user and do not rely on a numerical reward function. However, this study only investigated the initial exploration phase of the algorithm. Whether the algorithm efficiently adapts over time needs to be evaluated in a long-term study. Additionally, I collected a significant amount of video material that captures the HRI. The video could be analyzed in the future to gain more insights on how people adapt to robots.

This chapter concludes the empirical investigations in this thesis. It tried to find an answer to one of the hypothesis of this thesis. The evidence found in the presented study partially supports Hypothesis 1.3. Participants perceived an adaptive robot as more competent and trustworthy. However, I can not conclude whether an adaptive robot is also significantly more motivating than an adaptable robot. Future investigation needs to target this open issue.

9 DISCUSSION

The final chapter of this thesis revises the research steps, gives an overview of the main findings and discusses the meaning of the results. It will discuss why the results are useful for other researchers and how they could contribute to a better understanding of social robots as exercising companions. However, this chapter will also consider the limitations of this thesis.

9.1 THESIS SUMMARY

This thesis explored and investigated the usage of SAR as exercising partners. In the introduction, I stated that a lack of physical activity is one of the leading global risk factors for various diseases and that SAR might be a useful tool for people to increase their physical activity levels. The usage of SAR for rehabilitative scenarios was already presented over a decade ago [Fei05]. In contrast to previous works, this thesis focused on an empirical foundation for using social robots as exercising companions or coaches, by using a control-group based study design methodology to investigate the motivational effects of exercising with robots. It took a breadth-first approach on four important features for robotic exercising companions: social role,feedback, embodiment and adaptation.

First, I investigated the motivational effects of exercising with robots as partners (see Chapter 3 and 4). Second, I looked at the effects of the embodiment of the robot (see Chapter 3, 6 and 7). Third, I looked at the motivational effects of encouraging feedback for the interaction partner (see Chapter 5). As a last factor, I investigated the effects of having SAR with adaptive capabilities and their influence on the interaction experience (see Chapter 7 and 8). I identified these four factors as important for future research directions, because those will have implicatons for the long-term applications of SARs and future research [Iol13].

9.1.1 *Rational for thesis investigation*

The social role is essential for robot design and engineering reasons. If a Robot Companion is eliciting higher motivational gains than an Robot Instructor, then engineers working on exercising scenarios should build robots that will be capable of exercising with humans together. Hence, robots will need to be able to jump and balance humanlike, so that they can be equal exercising partners. However, based on the theory of social facilitation effect it might also be sufficient if the robot is not exercising with a human, but just instructing or observing them [Str02].

Independent of the social role, trainees can benefit from the verbal encouragement of a peer. Indeed, in cases where there is an unfavorable social comparison, as in the Köhler Effect, encouraging feedback from a superior exercising partner can also have adverse effects [Irw13]. Therefore, it is pivotal to understand what effect verbal encouragement of an exercising partner can have on the motivation.

Carrying on, for robot design reasons it is essential to look at the embodiment too. Designing and maintaining a robot is more expensive than a Virtual Agent. Therefore, it is crucial to understand whether using a robot is beneficial.

Finally, investigating the effects of an agent that tries to find out a user's exercising preference by exploration is vital for future application where programmers can not anticipate every potential user of a system. Thus, knowing how exploration algorithms are perceived in HRI is also critical for other applications of social robots.

This thesis presented different HRI scenarios and systems, where a robot works out with a user co-actively, to find answers to these questions. I started with a prototyping study using video Human-Robot Interaction to investigate the question on embodiment and social role, but the results did not confirm my hypotheses. Hence, I introduced a real interaction study where the robot is an exercise companion or instructor. The study design was adopted from Feltz et al. [Fel14] so that a comparison between the two collected dataset on exercising with robots and computer-generated partners is possible. In this study, I varied the social role (instructor vs. companion) and the presence and absence of encouraging feedback. Subsequently, I implemented a system that tries to learn the users exercise preferences based on the Dueling Bandit Problem. I investigated the effects of the embodiment of such a Preference Learning agent and the effects on the interaction experience. To further investigate the effects of such a preference learning agent, I tested it in a real exercising scenario and compared an adaptive versus an adaptable system.

9.1.2 *Result Summary*

Putting the results of this thesis in one sentence, one *could* say that it showed that an embodied SAR, which is exercising co-actively, giving encouraging feedback and adapts to its user will lead to the highest acceptance of the user and should be the gold standard for future SAR implementation. Unfortunately, saying this would oversimplify the results of this thesis. Therefore, I will break down the results into smaller chunks and ask what the single building blocks are.

KÖHLER EFFECT The results from Chapter 4 suggest that the Köhler Effect can also be replicated with humanoid robots. People in this study exercised longer when paired with a moderately superior Robot Companion compared to people exercising alone or with a Robot Instructor. Not only were exercising times longer, but also ratings for likability of the system. The result supports the Hypothesis 1.1 of this thesis However, there was no significant difference in persistence time between exercising individually and being instructed by the robot. The participants perceived the presence of the Robot Instructor as disturbing and unclear. Thus, I could not find an evidence for the social facilitation effect effect. However, this needs to be verified in an experiment that controls for the different participants' exercising levels. Usually, this effect should appear when people already have a high () in doing a task; if not the effect is reversed. However, people that are already confident in exercising probably do not need an assistive system that motivates them.

ENCOURAGING FEEDBACK The study in Chapter 5 was building up on the Köhler Effect study from Chapter 4 and included encouraging feedback while the user was exercising. The results show that encouraging feedback leads to higher exercising times compared to a robot instructor that is not giving feedback or exercising alone. However, there was no additional exercising boost when the robot was exercising togehter *and* giving feedback compared to when the robot is only exercising together with the user. In these conditions, the exercising time was not significantly different. Therefore hypothesis Hypothesis 1.2 of this thesis partly supported. Participants exercised longer when they were receiving encouragement, but not under all conditions.

EMBODIMENT Chapter 6 compared the results from my studies on the Köhler Effect and the results from Feltz et al. [Fel14] on the same effect but with computer-generated partners. The analysis showed that there is a significant difference between the conditions with a robot companion and the software generated partner. The results of this analysis partly support Hypothesis 1.3 of this thesis. Participants exercise longer with an embodied SAR. However, this conclusion needs cautious interpretation. Even though both studies used the same study design, participants were from different countries. Additionally, I needed to change the exercises slightly, which might have resulted in different baseline measurements of the participants between the cohorts.

Besides this results on the exercising times, the ratings for likability on the Godspeed questionnaire were higher for the robot exercising companions compared to the human partner. This, supports feedback from some participants telling that they would rather prefer to exercise with a robot than with a human.

ADAPTATION As the last factor, I have looked at the effects of adaptation. In Chapter 7 and 8, I have investigated the usage of PL learning algorithms that explore the user's exercise preferences. The results indicate that Dueling Bandits might be suitable for HRI. Especially, the comparison between an adaptive and an adaptable system showed that the users ascribe the adaptive system a higher competence and also state that their perceived alliance with the system is higher. Furthermore, also higher ratings for the perceived warmth of the system and the motivation to voluntarily exercise a second time with it supports Hypothesis 1.4 of this thesis.

9.2 INTERPRETATION

Looking at the different results of this thesis one can see that there is a trend that the dimensions of social role, embodiment, encouragement, and adaptation alter the user's motivation and evaluation of the system. However, in the presented studies these dimensions were only evaluated on an on-off basis, but possible co-variations along these dimensions seem to be reasonable to investigate in the future. The following section discusses these aspects.

9.2.1 *Exercising Motivation*

Having a robot exercising companion or a robot instructor that is giving encouragements resulted in significantly longer persistence times. The interpretation of this results is rather straightforward: When participants were paired with one of the agents, their persistence time difference between two exercising blocks was smaller compared to the control-group conditions, thus their motivation to exercises was higher. As introduced at the beginning of this thesis, motivation includes aspects of starting, maintaining and repeating a behavior. I limited the scope of this thesis to the maintaining aspect of motivation. Thus, the results show that pairing users with a robot that gives encouragement or one which is exercising co-actively increases the maintaining aspect of motivation.

Still, what is responsible for this difference in motivation and how can it be explained on a coarser scale? To some extent it could be explained by a novelty effect. Participants could exercise longer due to the novel stimulus of seeing a robot. However, if this would be the only effect influencing the exercising motivation of the participants, then there should also be a significant effect when the robot is only instructing the participants. Since the results do not show this effect, one can conclude that there might be additional effects influencing the exercising time besides the novelty effect.

As an illustration, one could boil it down to some simple equations:

$$
\begin{aligned}
t_{IC,1} &= mot_{intr} \\
t_{IC,2} &= mot_{intr} - \epsilon \\
\Delta t_{IC} &= t_{IC,2} - t_{IC,1}
\end{aligned}
$$

$t_{IC,1}$ is the participant's time to persist an exercise on the first block. That is, in the case where nothing else could manipulate the motivation, the intrinsic motivation mot_{intr} of the participant that influences the persistence time. This is a huge simplification because also a participant's personality, the average weekly exercising time, the mood, the health and many other variables influence the exercising motivation. The exercising time on the second block is $t_{IC,2}$, which is again mainly the intrinsic motivation substracted by an factor of ϵ that could be the exhaustion or boredness.

When a robot instructor is introduced, the novelty of the system could influence the persistence time between the two blocks :

$$\Delta t_{RI} = \Delta t_{IC} + novelty$$

In situations where the robot is playing the role of a moderately superior exercising companion, one could suspect that the Köhler Effect additionally enhances the exercising time.

$$\Delta t_{RC} = \Delta t_{RI} + \text{köhler}$$

Moreover, the exercising time should be additionally enhanced if we introduce encouragement for the exercising partner as an additional factor.

$$
\begin{aligned}
\Delta t_{RIF} &= \Delta t_{RI} + encouragement \\
\Delta t_{RCF} &= \Delta t_{RC} + encouragement
\end{aligned}
$$

If only the novelty effect could be attributed to the enhanced exercising time in the conditions with an exercising companion, then I should have also found an effect so that $\Delta t_{RI} > \Delta t_{IC}$. However, even though the exercising time was on average slightly higher in the condition with a Robot Instructor (RI), the effect was not significant. Therefore, some other factors might be responsible for inhibiting the exercising time when instructed by a robot. One possible explanation could be the social facilitation effect, or audience effect, which is the effect that people behave differently in the presence of others. People tend to perform better on easy or well-rehearsed tasks in presence of other compared to their performance when they are alone. Contrary, they perform worse in the presence of others on complex or less familiar tasks. Hence the equation might look as follows:

$$\Delta t_{RI} = \Delta t_{IC} + novelty + / - audience\ effect$$

Therefore, one could argue, for the case of the Robot Companion, that the contribution of novelty effects was higher than the inhibition of the audience effect, because the robot was doing something unexpected and novel when it exercised with the user. Thus, it is difficult to distinguish whether the difference in exercising time is only due to the novelty effect or the Köhler Effect. However, the obtained findings from the experiments are similar to the results from other researchers working on the Köhler Effect which rule out novelty as an explanation for the enhanced exercising time. Nevertheless, to be sure about this effect and the importance of the novelty effect and Köhler Effect, long-term investigations are mandatory.

Though, how do the equations hold for the other conditions? The studies of this thesis give evidence that including encouragement enhances exercising time (e.g., $\Delta t_{RIF} > \Delta t_{RI}$). However, simply adding up factors like encouragement to increase the persistence time does not hold if we look at the robot companion. If the addition of encouragement and the addition of the Köhler Effect each lead to an increase of persistence time, then including both factors should increase the exercising time even more. Thus, $\Delta t_{RCF} > \Delta t_{RIF} > \Delta t_{RC}$ should be the observed outcome. However, the conducted studies could not verify it. As discussed in Chapter 5, this might hint at a ceiling effect due to the difficulty of the exercises. Doing full planks for a long period, as required in Chapter 4, hurts the wrist, and thus it might not be possible to exercise longer even though one would still have motivation.

However, many pieces of evidence in sports science show that people first feel cognitively exhausted before the actual muscles are exhausted [Hil24; Noa05]. The brain will create this sense of fatigue that has very little to do with the actual muscle's ability to continue to work. People could, in theory, exercise longer even though they feel that their muscles are in pain. Here, one can still see a difference because stopping an exercise due to muscle pain or due to pain in the joints might be something different. Nevertheless, it opens room for further speculations, namely, that motivation is nothing that simply can be added up factor by factor. There might be an upper bound where an increase in motivation is not achievable.

This needs to be verified in future studies with exercises that allow measuring muscle fatigue. Additionally, it is also interesting to investigate how the found motivational effects for conjunctive physical task can be extended to mental taks (e.g., puzzle solving).

9.2.2 *Subjective perception of the partner*

Besides the exercising time as measurement in Chapter 4 and Chapter 5, most studies measured the perception of the partner using the Godspeed questionnaire. The results show variabilities in the ratings for perceived animacy and likability between the conditions, but not for anthropomorphism and intelligence. This section looks at the results of the Godspeed questionnaire across the different studies and discusses the found variabilities.

ANTHROPOMORPHISM The anthropomorphism sub-scale is mainly assessing the perceived human-likeliness. Therefore, it is not surprising that there is no difference in perceived anthropomorphism between the conditions in each study, because it was always the same robot and the human-likeliness of it was not explicitly manipulated in the studies. Moreover, I designed the interaction so that there were only few variations between the conditions. Therefore, in-build face tracking, background behavior, breathing and animated speech were always turned on and the same across conditions. Hence, it might be important to investigate how the perceived anthropomorphism influences exercising and interaction motivation.

INTELLIGENCE The perceived intelligence was not different in the conditions. Neither manipulating the social role, feedback nor embodiment affected the perceived intelligence of the system. Results from Chapter 4, 5, 7 and 8 showed that the perceived intelligence was not significantly different between the conditions and that the intelligence was rated around the middle on a 5-point differential scale. An explanation for this result might be a response bias where a respondent may stick to the middle-point item [Cro49]. This bias could indicate that the intelligence of the system was not a salient cue for the participants.

However, results on the RoSAS showed, in contradiction to the Godspeed questionnaire results, a significant difference between the adaptive and adaptable robot on the competence scale in Chapter 8. Reasons for this are that the competence might not measure the same as intelligence and that the RoSAS is more sensitive. The RoSAS is assessing a 9-point likert scale while the Godspeed questionnaire uses a 5-point differential scale. This difference in scale resolution couold also be a reason why there were no differences in perceived intelligence in the studies in Chapter 4.

ANIMACY It is not very surprising that the ratings for animacy are different when a robot is exercising co-actively and not just standing in a room. I found this result in the study from Chapter 4, but not in Chapter 5. One can see, that there is a trend to stick to the middle for the Robot Companion condition and that the ratings for the Robot Instructor condition slightly diverge from the middle point of the scale. While this result is significant for the case where I only compared two conditions, the results look different if I compare two additional conditions in the study on motivational feedback. Testing more conditions requires to adjust the alpha level, which is why the difference is not significant anymore. Thus, a higher sample size is needd.

Animacy is also not significantly different when comparing a virtual Nao or a real Nao and an adaptive or adaptable robot. In all cases, ratings for animacy stick to the middle item which also indicates a response bias.

LIKABILITY The only sub-scale that repeatedly showed differences between conditions for each study is the perceived likability scale. Results from Chapter 4 showed that the social role influences the perceived likability. Therefore, one could hypothesize that the likability of the system might influence the increase in exercising time. However, this argument does not hold, because the Robot Instructor with Feedback also led to higher exercising times, but was rated significantly less likable than the companion robots. Even though the effects of the higher ratings on the likability sub-scale might not be influence short-term interactions, it could increase the long-term motivation to exercise with the robot.

The data analysis of Chapter 6 showed significantly higher ratings on the likability scale for the roboto companions compared to computer generated partners and human partners. Thus, not only social role influences the perceived likability of a system but also the embodiment. However, one has to bear in mind that Nao, the used robot, is intentionally designed to be cute. Thus, it is possible that the increase of perceived likability is not due to the embodiment but the appereance of the partner. Considering this results shows that also when the partner is not liked that much, as for the human partner condition, participants exercise longer compared exercising alone. Thus, it could imply that an increase in motivation is not due to the likability of the partner. Though, further studies need to verify this.

At last, the likability was not evaluated differently on the Godspeed scale in the adaptation study. For both robots, likability ratings were at the highest value of the scale. Reasons for this result are diverse. Ratings for likability might be influenced either by a social norm bias or by a ceiling effect. If the reasons for the difference are due to a social norm bias, then similar ratings should have been observed in the previous studies. However, this is not the case, which hints at a ceiling effect. It might also explain why the ratings for the RoSAS scale showed a significant difference on the warmth sub-scale, because the RoSAS scale has a wider item response or better captures the concept of perceived warmth or likability. This argument might be plausible, because researcher criticized the Godspeed questionnaire for several reasons (e.g., items are confounded with positive and negative affect, items do not load as expected on the five scale dimensions, items do not correspond to the underlying constructs, high correlation between dimensions [Car17]).

9.3 CONTRIBUTION

After discussing the results of this thesis across the different chapters, it is time to answer the question of why these findings are significant and who can use them.

PHYSICAL ACTIVITY At the beginning of this thesis, I postulated that there is a considerable demand for motivating people to increase their physical activity levels and that SARs might be an appropriate tool for achieving this goal. However, to deploy robots as tools for clinical experts, therapists or coaches one has to guarantee that similar motivational effects appear when interacting with robots as when exercising with human partners or having a human coach. There exists evidence that people stick to the same social norms when interacting with media and technology [Ree97] and that robots provide an interface that relies on the tendency to be anthropomorphized [Epl07]. It was so far not evident that the motivational effects like the Köhler Effect are also measurable when having a robot exercising companion. This thesis filled this research gap by showing that the Köhler Effect can also be replicated with SARs and shows that robots could be used to enhance the maintaining aspect of motivation. This proof has the implication that health practitioners can think about use cases were clients could benefit from an exercising partner, but no exercising partner is available. In such situations, robots could be beneficial. Compared to a human exercising partner a robot does not have a conflicting schedule, a lack of motivation or could get sick. The robot could break, or the battery could die but replacing one of these would be more comfortable than replacing a human.

ROBOT REQUIREMENTS The second implication is the need for robot engineers to build, at best, robot hardware that is robust and flexible. Robots for exercising purposes will need to be able to jump and run to be exercising partners that can workout together with humans. This request sounds like a hard engineering challenge, but can be relaxed by the finding that similar motivational exercising effects occur when a robotic instructor is giving encouraging feedback. Until robots can do a dynamic full body weight workout, it is sufficient for them to give encouraging feedback to motivate the user and participate in the exercises they can do. Nevertheless, the long-term effects of encouraging feedback versus co-actively exercising still needs to be verified. The finding that the information quality is perceived higher when the robot is exercising together, even though both were giving the same information, suggests that robots that do a task together with a human might increase the associated competence and trust in the system, which is likely to affect the long-term interaction.

ROBOTS AS PARTNERS As a third contribution, I suggest that in some cases it might be even better to pair a trainee or client with a robot instead of a human partner. There is a tendency that people do not feel judged or evaluated by a robot. Post-study interviews showed that some people would prefer to exercise with a robot partner than with a human or a group of humans because they feel evaluated, which puts them under stress and induces uncomfortable feelings. Additionally, the evaluation in Chapter 6 showed that a robot companion is rated higher on the likability scale even compared to a human partner. People that are facing a lack of physical activity are likely to have low self-efficacy belief, and thus they probably do not seek an alliance with exercising teams or coaches which in turn leads to no improvement, and so the vicious circle continues. Therefore, people with either social anxieties or a low self-esteem/efficacy beliefs might benefit from a robot exercising partner that acts as a motivator during an initial training phase of a rehabilitation or exercising program especially. Likely, the robot will not be required after people have gained a baseline activity level where they feel competent to engage in exercising activity with other people. Still, this is an open question, which future research needs to address, but could relate to the Social Cognitive Theory (SCT).

Additionally, the findings of the Köhler Effect contributes to other domains where people already interact with robots or are expected to interact with in the future. The found motivational effect suggests that it would be beneficial to model the task and the interaction as a cooperative scenario, where both the human and the robot's behavior contribute to the task outcome. This result might have application in, for example, industrial, medical or service tasks where humans and robots will interact and cooperate in the future. The mutual care paradigm is one research that points in a similar direction [Lam14]. Not only the service robot cares for the elderly at home, but also the user cares for the robot. Thus, creating a cooperative scenario where the task outcome (maybe one could call it: number of days of successfully living together independently) is depending on both interaction partners might also increase the motivation to interact with the robot for extended periods of time.

EMBODIMENT The results on persistence time from the analysis in Chapter 6 showed that people persist longer on the exercising task when paired with a robot companion. This result implies that researchers and practitioners that are working in the field of technology-assisted healthcare applications or Active Video Game should bear in mind that using a robot might lead to better results. Still, one has to carefully consider the cost-benefit trade-off between deploying robots or virtual agents. Even though Virtual Agents might not lead to the same motivational exercising effects as robots, they still increase persistence time compared to having no partner. Nevertheless, the results could guide future research funding to build inexpensive and easy to deploy robots for the reasons mentioned above.

PERSONALIZATION At last, the results from the studies on adaptation and personalization from Chapter 7 and 8 might be useful for other researchers working on adaptive robots. Building adaptive robots that can adjust to a user's preferences is an important topic for social robotics and will remain a challenging research direction in the next years. The problem of adaptation is diverse, and robots could adapt to its user in various ways. Proxemic, speech, emotion or personality are all important aspects in terms of which a robot could adapt its behavior towards the user's preferences. In contrast to other adaptation goals, this thesis investigated the adaptation on a task level and tested a principle from preference learning in HRI. This idea of dueling bandits is especially useful when it is difficult to design a reward function or where the reward is highly subjective. The results provide useful evidence that this kind of adaptation process might be suitable for HRI. It might not only be used for task adaptation, but also for learning task skills or other preference-based learning situations. Besides, the application of dueling bandits for preference learning, the study on adaptation and adaptability implies that a certain degree of automation of the system increases the perceived competence of it and is beneficial for the relationship with its user. In a way, it supports research that works on increasingly adaptive systems from an interaction experience perspective and shows its benefits. It also shows that people seem to rely upon the competence of a system and do not fear to give it, in a soft sense, control. One can hypothesize that people prefer if the system has a higher control than controlling it by themselves because it reduces the responsibility of the user to control a robot in a correct way, which might feel relieving.

9.4 LIMITATIONS

Various factors and research choices limit the results of this thesis, which also influence the quality to answer the questions and hypotheses of it. Since each chapter already discussed the results and limitations, this section will only highlight the main limitations that account in all of the presented studies. The three most prominent limitations of this thesis are the choice of participants, the limited interaction time and the used robot platform.

POPULATION In the presented studies I mainly sampled participants from a healthy young student population. Not only healthy and young but my samples were also "drawn entirely from Western, Educated, Industrialized, Rich, and Democratic (WEIRD) societies" [Chi10]. Thus, I cannot generalize to other populations from eastern, non-industrialized, non-democratic or poor societies as well as for elderly, children, people suffering from Noncommunicable Diseases (NCDs) or obesity. Thus, it is somewhat difficult to answer the question of whether people that need more physical activity could benefit from a social robot. Healthy young students might be a population that is not affected by a lack of physical activity.

Thus, it is not easy to transfer the results to other groups by showing that SARs can enhance this population's motivation to exercise longer. People that are not facing any motivation or health issue might comply very well with study instructions and do not question them, especially when they are voluntarily taking part in an experiment. Motivating focus groups to exercise with a robotic system might be much harder compared with student groups. However, some reasons justify these study decisions.

First, finding a focus group (people suffering from obesity or NCDs, children or elderly people) and conducting studies with these groups would be much more challenging and time-consuming in a one-person project. It would have required to find clinical and health-care experts as project partners, negotiations about the project with all interested groups as well as getting an ethical consent. These requirements might be met in a large-scale project with more members but have been difficult to realize when working individually.

Second, even though students might be a particular section of the society, the results obtained with them are still vital to guide further research, get initial ideas on how motivational processes might work, and verify in-lab found results in the real world afterward. Thus, to confirm whether the discovered motivational gains can generalize to other populations will be an ongoing research concern.

INTERACTION TIME AND PLACE The second limitation of this thesis is the overall short interaction time between the robot and the user.

This limitation might restrict the results because the novelty effect could also attribute to the findings. Therefore, it is reasonable that the found effects wear off over repeated interaction with the system. Additionally, interaction times were also not long enough to study the effectiveness of the preference learning algorithm during an exploitation phase. The short interaction time also comes together with a study setting that only took place in the lab. The found effects might be very different in settings where people interact with the robot at home or in facilities like gyms or clinics. Both, the short interaction time and the location, influence the perception of the system and the exercising adherence. Following the same as the argument for the user population, these results might not be transferable to other scenarios, but they still give a good starting point that further investigations in this direction might be seminal. If this thesis proved the opposite, that a robot does not elicit any motivational effect for the user, then subsequent investigations would probably show no results also. Thus, the results are valuable to justify further research project and investigations. Further investigations would not be meaningful, if there were not any motivational effects, even though due to the novelty effect. Additionally, when users perceive the exploration phase of an adaptive agent as too cumbersome, then the usage of preference learning agents might be doubtful. Regarding these aspects, this thesis provides an exploratory investigation in this direction.

ROBOT PLATFORM The last general limitation of this thesis concerns the used robot platform, both in its dynamics and appearance. Nao is very restricted in its motion space, and it is almost not possible to animate fast and dynamic exercises. The missing difference in robot appearance limits the results because either static exercises (e.g., abdominal plank exercises) are possible to animate, or moderate dynamic activities, which are not as fast as humans could do them (e.g., jumping jacks). Thus, exercises offered by the system in Chapter 6 were limited to the constraints of the robots, and participants have exercised below what they are capable of. Perhaps, this might limit the user evaluation of the system, though, the participants did not complain about the exercising speed of Nao.

Moreover, the appearance of Nao is rather cute and toy-like than professional. This effect of the robot's appearance might also confound the results because the appeal of a robot triggers certain stereotypes and failures of the robot might not weight too much due to the cuteness of the robot. Thus, an evaluation of the confounding factors of the perceived cuteness of Nao could reveal whether these effects are genuine or due to the robot appearance.

9.5 FUTURE WORK

Based on the limitations mentioned above, I would make the following suggestions for future research in the field of assistive robots in exercising scenarios. To better anticipate the needs and requirements of affected people with a lack of PA, or suffering obesity or NCDs, studies with focus groups are needed. Working with focus groups would require a consortium of computer scientists, roboticists, psychologists, therapists and medical/healthcare experts to identify the needs of such groups and identify requirements for SAR to be deployed in such contexts. The goal should be to make clinical trials that evaluate the effectiveness of SAR not only in labs but also in-the-wild (e.g., people's home). Accordingly, it would also target the second limitation of this thesis as it would require to tackle the long-term interaction challenges and identify the contribution of the novelty effect. At last, such a project could also identify the requirements that need to be addressed by robotic industry partners.

Such work with focus groups requires researchers to draw more from a diverse set of research methodologies. The usage of qualitative methods has not widely spread in the domain of SAR yet. There are some approaches in this direction, as Winkle et al. [Win18] presented recently, but still they are rare. Winkle et al. [Win18] investigated therapist recommendations for the design of SARs for therapy and presented recommendations for SAR requirements. This piece of work shows the direction in which future research could go. Research will not only look at the empirical motivational effects of using SAR on the user's behavioral outcome in experiments, but will also address the needs and recommendation of therapists, physicians, nurses, family members, the person concerned and their outcome. Feil-Seifer et al. [Fei07] proposed different methods for benchmarking the usage and deployment of SAR in different scenarios. They proposed to evaluate how these robots might not only be beneficial for the users, but also the impact on user's care and life and the caregivers. But to the date of this thesis it is unknown to me wether there is a study that incorporated a multitude of this evaluation criteria yet.

Moreover, additional evaluation methods like social signal processing could be used in the future. None of my reviewed works incorporated social signal processing in the interaction and feedback loop for SAR in exercising tasks. However, with out-of-the box tools like OpenFace researchers could easily incorporate social signal processing in the interaction and the facial response from the user [Bal18]. Past research has used facial expression and the derived user valence as features to for empathic feedback in tasks like chess [Lei11]. Therefore, it is reasonable to use these features also in tasks like exercising or rehabilitation. In the course of my studies, I collected a great amount of video that could be explored, but due to the limitation of this thesis, I could not yet analyze the video material and look in to the effects that the robot had on the user's valence. This analysis will be part of future research.

Future work should also address the investigation of feasible measurements for SAR. In this thesis, I mainly used scales from the Godspeed questionnaire, the RoSAS, NARS, PAES and from the CASA studies. However, it is an open issue whether these scales are are appropriate for physical exercising interaction between partners. The used measurement mainly assessed the interaction and exercising enjoyment, but is this the same as the motivation to interact. To overcome this lack in questionnaire constructs, I used behavioral measurements to assess the users motivation. However, these measurements are also noisy and can be influenced by various factors (e.g., time of the day/year, health and mental condition). This variety requires to have repeated measurements and at best measurements that are standardized and used across different research groups to facilitate the comparison of research results.

Upcoming research could target the construct and validation of a scale specially suited for such kind of interactions. Important constructs would be the perceived extrinsic and intrinsic motivation of the user, as well as the self-efficacy beliefs. These could be used to measure the baseline levels of the user and conclude whether the interaction with a SAR somehow influenced these constructs.

Possible interesting mediating factors would be the embodiment of the robot, the perceived anthropomorphism, intelligence and likability. Hoffmann et al. [Hof18] recently proposed a scale to measure users' perceptions of an artificial entity's body-related capabilities. This could be helpful to determine what capabilities of the robot's embodiment might influence aspects of motivation. Anthropomorphism, intelligence and likability have been targeted by the Godspeed questionnaire, but the usefulness of these scales for specific capabilities of SARs are disputable [Car17].

As SAR are expected to become daily interaction partners one crucial aspect is dialogue capabilities of the agents. None of the reviewed literature from Chapter 2 looked into aspects of dialogue management or relation building. Most of the works are concerned with feasibility tests of their systems. Thus, these works mainly verified the perceptive or coaching abilities of the system. Still, after these problems have been solved, the open issue of building relationship with the users remains. Here, a look in the field of HCI could be valuable where dialogue or relationship management have already been targeted [Bic05]. However, a recent literature review found that research on ETs were poorly tested [Men17] and assume that their "results may indicate that a physical robot can communicate a sense of presence more efficiently than an ET" [Men17, p. 315].

In summary, I would suggest three research projects to tackle the challenges of deploying and using SAR in applications for exercising and increasing physical activity levels. A first project would be a longitudinal socio-linguistical investigation on how coaches and exercising partners interact with each other over an extended period of time. A qualitative analysis could be used to model the interaction between the partners and use them to build a computational model.

A second project should focus on the collection of data that can be used to investigate social signals in exercising. The data could be used to train a model that might be able to measure motivation using facial cues. Modern image processing is also capable of detecting heartrate from images, which could be an additional measurement. One could use wearable sensors, but I would favor approaches that do not need the user to wear any equipment. It would increase the deployability of the system and reduce faults and maintenance that is increased by using more technological devices. Thus, at best, everything would run on the robot.

As a third project, I would investigate more the motivation behind exercising with robots. I would use standardized exercises that can be used to measure behavioral motivation. Since, this thesis showed that working out together significantly increases exercising time, it is open to investigate whether this effect is replicable and how the embodiment and anthropomorphism of the robot might influence the interaction. Thus, I would first investigate the effects of having a virtual agent versus a robotic partner. Then, I would alter the appearance of the robot and would use smaller toy-like robots, human-size robot and at best human-like robots like the Geminoid [Nis07]. Additional investigations could investigate how the Köhler Effect effect might be applied to industrial or domestic settings.

Finally, the results could be used to develop a system that engages people in physical activity everyday at their home. Moreover, I would suggest to explore a hybrid approach. It is likely that there will be no sufficiently advanced humanoid robots that could accompany a person during the whole day. I assume that the robots might be stationary at homes. Those robots will be autonomous in the homes of their users, but will not be able to follow them to every location that the user is going (e.g., going downstairs on the street, to the forest or driving with them in a car). Thus, robots could be exercising tools for homes, but not usable outdoors. For such scenarios, a hybrid approach is useful where the robot is 'transffering' itself on the user's smartphone or smartwatch and can be used as a motivational companion during other activities like hiking, rowing, swimming. Current smart watches are already capable of tracking many activities, thus the robot could accompany the user as an avatar.

However, these wearable devices also introduce issues of privacy and further ethical challenges that researchers need to address.

9.6 ETHICAL CHALLENGES

The current trend in using applications to track every aspect of one's life is a major ethical issue. One can observe that people are readily available to share their personal information about everyday habits to optimize their current life style [Lup14]. People track their heart rate, their daily activities as well as their eating habits. They give those data out of their hands to fit into the demands of society.

Why should SAR researchers also be concerned with these trends? As the present research suggests, Embodied Robot have the potential to influence and persuade people more than other technologies [Li15]. Thus, researchers need to be aware of the effects that technological advancement might have on the human. I will briefly summarize this problem in the remaining of this chapter.

This thesis explored the usage of SAR to increase peoples' physical activity levels in light of a delicate motivation. Global overweight has made it to a severe problem for humankind. Worldwide, obesity killed more people than famine and malnutrition in 2010 [Abu15]. The reasons for overweight are diverse, and the usage of robots to tackle this problem fights only the symptoms. Overweight has come to a problem that is also firmly connected with our societies, the economy, and politics. In the face of neoliberalism, the problem of overweight has turned into a problem for the individual, where everyone is responsible for themselves while fading out the circumstances that also contribute [Mon16; Har07]. Education, environment (e.g., food supply, neighborhoods), psychological disorders, and social status have a powerful connection to overweight and the resulting health condition (e.g., [Veu05; Wil10; Zwa01; Fai98; Goo03]). However, changing education and food supply in schools or creating equal opportunities for people from different social status groups and supporting low-income neighborhoods is a political concern.

In this thesis, I investigated how social robots can be used as an exercising partner and how this might influence a person's motivation. The found results are relevant for the use of such a system as a tool for people with special needs, but as conventional technology, it has two sides: On the one hand, they can be helpful for people that need assistance. On the other hand, they can result in an undermining of the political discourse.

In the political research agendas robots and especially social robots are used as a solution to individualized or societal problems [Dip15]. In hospitals and elderly care homes there is a lack of healthcare specialists and caregivers [Org13]; a shortage of well-trained personnel like physicians or psychologist in rural areas [Org10]; trained teachers are missing in schools and schools handle this lack with hiring newcomers [Wel18]. Thus, one narrative is that a personal robot will be available for everybody to care about them, assist them, teach them, and give them psychological and emotional support (see [KPM18] as an example of how professional service companies market social robots). Those aspects were earlier responsibilities of strongly connected social groups like families and peers and part of political discourse about governmental responsibilities [Har14; Sha79; Bro15]. However, such tasks are not valued anymore, because they do not enhance the personal and economical growth. This moves the discussion from structural issues to individual problems. If one's children can not keep up with today's demands of schools, get a private tutor; if one has no time to support their parents when they are old, one can get a private caregiver. While private tutors and caregivers are currently still human, those people could be replaced by social robots in the future. This possibility opens a new market where start-ups could invest money, the economy could grow and investors could market social robots as a tool that might solve the problems mentioned above (e.g., [KPM18]). People could buy those technologies to free up time for other matters like investing in their career or education. Though, it might prevent a political discussion about the problems at hand and explorations of other ways to deal with them.

I would recommend as a requirement for research proposals to highlight the ethical and social issues. If one looks at the `Proposal Preparation Instructions`[1] supplemented by the German Research Foundation (DFG), one can see instructions to describe the ethical and legal aspects only when human subjects are involved. There is no advice to consider social aspects of the developed technology. Thus, my recommendation would be to justify research projects also by their social impacts, especially when the motivation of such projects is grounded on a societal problem (e.g., demographic change). Hence, the political discourse would also be reflected in the research proposals.

The development of robots for health service applications or social tasks could contribute to the privatization of such sectors. Thus, it is essential to regulate who owns the technological developments that were made by tax funded projects. If, in the end, industries own those products, because they are the ones that could invest sufficient money to build consumer products, and then sell them to the, e.g., health sector, who will cover the costs for this? The state, health insurances, or the end-user? This question is pivotal to answer, in order to guarantee that current research technologies will in the end also reach the target group for which the primary research funding was intended.

This problem is related to the issue that SARs are expensive tools and not everyone can afford them. Thus, it is likely that these technologies will be only available for those people that can afford them or have good health insurance thereby contributing to social schism. Looking back at the data from the World Health Organization (WHO)

1 `http://www.dfg.de/formulare/54_01/54_01_en.pdf`, visited on 2018/09/02

obesity is not only a problem for wealthy western societies with an insurance system, but it is also a dominant issue in developing countries. Thus, they should engage in multidisciplinary dialogues about future developments of the social sector. Questions of affordability and access for everyone must be considered. Additionally, societal consequences of increased automation should be expounded, not only regarding the impact it has on the user, but also with a focus on employees in the social sector whose fields of responsibility will change dramatically.

Summarizing, social robots have a high potential to affect our society in many different ways. They can be a valuable tool to support urgent concerns such as the shortage of therapists, caregivers, and teachers or to be deployed for special populations. The thesis contributes to the ongoing struggle with societal problems like overweight and obesity by showing how a robot could motivate people to maintain a behavior for an extended time. However, the development of these technologies can also shift the discourse from the causes of problems to fight only the symptoms. Hence, robotics researchers should bear in mind the ethical aspects mentioned, consider them in their everyday work, and take a broader perspective.

10 CONCLUSION

This thesis looked at the usage of Socially Assistive Robots for exercising applications from different angles. It investigated factors that are relevant for Socially Assistive Robots in future scenarios and questioned how social role, encouragement, embodiment and adaptation affect a user's exercising motivation and evaluation of a Socially Assistive Robot.

Therefore, the thesis presented four general hypotheses concerning these factors. It presented empirical investigations that tried to verify these hypotheses. The results show that

1. working out with robot companion enhance exercising times (proving Hypothesis 1.1)
2. encouraging feedback from robot instructors enhances exercising time (partially proving Hypothesis 1.2)
3. the embodiment might also influence exercising time (partially proving Hypothesis 1.3)
4. an adaptive robot is perceived as competent and users ally with it (proving Hypothesis 1.4).

The main contributions are a) an understanding of how a robot enhances a person's motivation by doing a task co-actively or giving a user encouraging feedback, b) additional perspectives for the debate about the importance of physically present systems, c) research on adaptive robots by presenting how a robot might learn a user's preference during interaction, d) insights about the influence of a proactively deciding robot on a user's evaluation and alliance with it.

The contributions could stimulated the ongoing efforts of increasing global physical activity levels by showing that robots could be used to enhance the maintaining aspect of exercising motivation.

Limitations of this thesis are the short interaction times, study samples only drawn from a student population, and a platform with limited exercising capabilities. Thus, future research should examine whether the found effects sustain in long-term interactions and try to verify the motivational gains with cohorts sampling from target groups.

.

BIBLIOGRAPHY

OWN PUBLICATIONS

[Sch12] S. Schneider et al. "Effects of Different Robot Interaction Strategies During Cognitive Tasks." In: *ICSR*. Lecture Notes in Computer Science. Springer, 2012, pp. 496–505. *used on: p. 4*

[Sch14] S. Schneider et al. "How Socially Assistive Robots Supporting on Cognitive Tasks Perform." In: *Proceedings of the 50th Anniversary Convention of the AISB*. 2014. *used on: pp. 35, 49, 55*

[Sch15] S. Schneider et al. "Long-Term Feedback Mechanisms for Robotic Assisted Indoor Cycling Training." In: *Proceedings of the 3rd International Conference on Human-Agent Interaction*. HAI '15. Daegu, Kyungpook, Republic of Korea: ACM, 2015, pp. 157–164. DOI: 10.1145/2814940.2814962. *used on: p. 7*

[Sch16a] S. Schneider et al. "Exercising with a humanoid companion is more effective than exercising alone." In: *Humanoid Robots (Humanoids), 2016 IEEE-RAS 16th International Conference on*. IEEE. 2016, pp. 495–501. *used on: pp. 35, 72*

[Sch16b] S. Schneider et al. "Motivational Effects of Acknowledging Feedback from a Socially Assistive Robot." In: *Social Robotics: 8th International Conference, ICSR 2016, Kansas City, MO, USA, November 1-3, 2016 Proceedings*. Springer International Publishing, 2016, pp. 870–879. *used on: p. 53*

[Sch17a] S. Schneider et al. "A framework for designing socially assistive robot interactions." In: *Cognitive Systems Research* 43 (2017), pp. 301–312. *used on: pp. 8, 39, 101, 102*

[Sch17b] S. Schneider et al. "Does the User's Evaluation of a Socially Assistive Robot Change Based on Presence and Companionship Type?" In: *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. ACM. 2017, pp. 277–278. *used on: p. 23*

[Sch17c] S. Schneider et al. "Exploring Embodiment and Dueling Bandits for Preference Adaptation in Human-Robot Interaction." In: *Proceedings of the 26th IEEE International Symposium on Robot and Human Interactive Communication*. 2017. *used on: p. 79*

[Süs14] L. Süssenbach et al. "A robot as fitness companion: towards an interactive action-based motivation model." In: *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*. IEEE. 2014, pp. 286–293. *used on: pp. 7, 53, 162*

GENERAL

[Abu15] I. Abubakar et al. “Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013.” In: *Lancet* 385.9963 (2015), pp. 117–171. *used on: p. 127*

[And15] S. Andrist et al. “Look like me: matching robot personality via gaze to increase motivation.” In: *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. ACM. 2015, pp. 3603–3612. *used on: p. 95*

[Aue02] P. Auer et al. “Finite-time analysis of the multiarmed bandit problem.” In: *Machine learning* 47.2-3 (2002), pp. 235–256. *used on: p. 80*

[Bai11] W. A. Bainbridge et al. “The benefits of interactions with physically present robots over video-displayed agents.” In: *International Journal of Social Robotics* 3.1 (2011), pp. 41–52. *used on: p. 68*

[Bal18] T. Baltrusaitis et al. “OpenFace 2.0: Facial Behavior Analysis Toolkit.” In: *Automatic Face & Gesture Recognition (FG 2018), 2018 13th IEEE International Conference on*. IEEE. 2018, pp. 59–66. *used on: p. 124*

[Ban77] A. Bandura. “Self-efficacy: toward a unifying theory of behavioral change.” In: *Psychological review* 84.2 (1977), p. 191. *used on: pp. 19, 156*

[Bar09] C. Bartneck et al. “Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots.” In: *International Journal of Social Robotics* 1.1 (2009), pp. 71–81. *used on: pp. 24, 28, 43*

[Bar15] K. Baraka et al. “Adaptive interaction of persistent robots to user temporal preferences.” In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 2015, pp. 61–71. DOI: 10.1007/978-3-319-25554-5_7. *used on: pp. 81, 82*

[Bau02] A. E. Bauman et al. “Toward a better understanding of the influences on physical activity: the role of determinants, correlates, causal variables, mediators, moderators, and confounders.” In: *American journal of preventive medicine* 23.2 (2002), pp. 5–14. *used on: p. 2*

[Bee14] J. M. Beer et al. “Toward a framework for levels of robot autonomy in human-robot interaction.” In: *Journal of Human-Robot Interaction* 3.2 (2014), pp. 74–99. *used on: pp. 98, 100*

[Bic01] T. Bickmore et al. “Relational agents: a model and implementation of building user trust.” In: *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM. 2001, pp. 396–403. *used on: p. 53*

[Bic05] T. W. Bickmore et al. “Establishing and maintaining long-term human-computer relationships.” In: *ACM Transactions on Computer-Human Interaction (TOCHI)* 12.2 (2005), pp. 293–327. *used on: pp. 42, 53, 104, 125*

[Biz07] R. Bize et al. “Physical activity level and health-related quality of life in the general adult population: a systematic review.” In: *Preventive medicine* 45.6 (2007), pp. 401–415. *used on: p. 1*

[Bor98] G. Borg. *Borg's perceived exertion and pain scales.* Human kinetics, 1998. *used on: p. 43*

[Bou13] M. N. Boulos et al. “Exergames for health and fitness: The roles of GPS and geosocial apps.” In: *International Journal of Health Geographics* 12 (2013), pp. 1–7. ISSN: 1476072X. DOI: 10.1186/1476-072X-12-18. *used on: p. 2*

[Bro15] W. Brown. *Undoing the demos: Neoliberalism's stealth revolution.* Mit Press, 2015. *used on: p. 128*

[Bro96] J. Brooke et al. “SUS-A quick and dirty usability scale.” In: *Usability evaluation in industry* 189.194 (1996), pp. 4–7. *used on: p. 87*

[Bru14] R. van den Brule et al. “Do robot performance and behavioral style affect human trust?” In: *International journal of social robotics* 6.4 (2014), pp. 519–531. *used on: pp. 26, 27, 33*

[Bus14] R. Busa-Fekete et al. “A survey of preference-based online learning with bandit algorithms.” In: *International Conference on Algorithmic Learning Theory*. Springer. 2014, pp. 18–39. *used on: pp. 80, 82*

[Cak10] M. Cakmak et al. “Designing interactions for robot active learners.” In: *IEEE Transactions on Autonomous Mental Development* 2.2 (2010), pp. 108–118. *used on: p. 96*

[Car17] C. M. Carpinella et al. “The robotic social attributes scale (RoSAS): development and validation.” In: *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. ACM. 2017, pp. 254–262. *used on: pp. 104, 119, 125*

[Cha12] J. Chan et al. “Social intelligence for a robot engaging people in cognitive training activities.” In: *International Journal of Advanced Robotic Systems* 9.4 (2012), p. 113. *used on: pp. 81, 82*

[Chi10] J. Y. Chiao et al. “The weirdest brains in the world.” In: *Behavioral and Brain Sciences* 33.2-3 (2010), pp. 88–90. *used on: p. 122*

[Coh88] J. Cohen. *Statistical power analysis for the behavioral sciences. 2nd.* 1988. *used on: p. 30*

[Cro49] L. J. Cronbach. “Essentials of psychological testing.” In: (1949). *used on: p. 117*

[Cro51] L. J. Cronbach. “Coefficient alpha and the internal structure of tests.” In: *psychometrika* 16.3 (1951), pp. 297–334. *used on: p. 29*

[Csi00] M. Csikszentmihalyi. *Beyond Boredom and Anxiety: Experiencing Flow in Work and Play*. 25th Anniversary. Jossey-Bass, 2000-04. ISBN: 0787951404. *used on: p. 21*

[Dau99a] K. Dautenhahn. “Embodiment and interaction in socially intelligent life-like agents.” In: *Computation for metaphors, analogy, and agents*. Springer, 1999, pp. 102–141. *used on: p. 25*

[Dau99b] K. Dautenhahn et al. “Bringing up robots or—the psychology of socially intelligent robots: From theory to implementation.” In: *Proceedings of the third annual conference on Autonomous Agents*. ACM. 1999, pp. 366–367. *used on: p. 156*

[De 10] M. De Gemmis et al. "Learning preference models in recommender systems." In: *Preference Learning*. Springer, 2010, pp. 387–407. *used on: p. 80*

[DeB87] V. DeBenedette. "With exercise videos, fitness can begin at home." In: *The Physician and Sportsmedicine* 15.10 (1987), pp. 164–170. *used on: p. 2*

[Dec06] E. L. Deci et al. *The Intrinsic Motivation Inventory*. http://www.selfdeterminationtheory.org/questionnaires/10-questionnaires/50. retrieved June 6, 2015. 2006. *used on: p. 87*

[Dip15] T. Diplomat. *Japans Robot Revolution*. 2015. URL: https://thediplomat.com/2015/07/japans-robot-revolution/ (visited on 2018-09-01). *used on: p. 127*

[Dis80] R. K. Dishman et al. "Self-Motivation and Adherence to Habitual Physical Activity 1." In: *Journal of Applied Social Psychology* 10.2 (1980), pp. 115–132. *used on: p. 19*

[End95] M. R. Endsley et al. "The out-of-the-loop performance problem and level of control in automation." In: *Human factors* 37.2 (1995), pp. 381–394. *used on: p. 98*

[Epl07] N. Epley et al. "On seeing human: a three-factor theory of anthropomorphism." In: *Psychological review* 114.4 (2007), p. 864. *used on: pp. 1, 6, 96, 119*

[Eri05] J. Eriksson et al. "Hands-off assistive robotics for post-stroke arm rehabilitation." In: *Rehabilitation Robotics, 2005. ICORR 2005. 9th International Conference on*. IEEE. 2005, pp. 21–24. *used on: pp. 13, 16, 17, 20, 165*

[Fai98] C. G. Fairburn et al. "Risk factors for binge eating disorder: a community-based, case-control study." In: *Archives of general psychiatry* 55.5 (1998), pp. 425–432. *used on: p. 127*

[Fan16] J. Fan et al. "A robotic coach architecture for multi-user human-robot interaction (RAMU) with the elderly and cognitively impaired." In: *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE. 2016, pp. 445–450. *used on: pp. 16, 17, 166*

[Fas10] J. Fasola et al. "A Socially Assistive Robot Exercise Coach for the Elderly." In: *Journal of Human-Robot Interaction* 1.1 (2010), pp. 1–16. *used on: pp. 14, 16, 35*

[Fas12] J. Fasola et al. "Using socially assistive human–robot interaction to motivate physical exercise for older adults." In: *Proceedings of the IEEE* 100.8 (2012), pp. 2512–2526. *used on: pp. 3, 4, 7, 13, 14, 16, 18, 53, 54, 165*

[Fas13] J. Fasola et al. "A socially assistive robot exercise coach for the elderly." In: *Journal of Human-Robot Interaction* 2.2 (2013), pp. 3–32. *used on: pp. 5, 16, 18, 21, 22, 25, 37, 67, 68, 165*

[Fei05] D. Feil-Seifer et al. "Defining socially assistive robotics." In: *in Proc. IEEE International Conference on Rehabilitation Robotics (ICORR 05*. 2005, pp. 465–468. *used on: pp. 1, 4, 11, 111, 156*

[Fei07] D. Feil-Seifer et al. "Benchmarks for evaluating socially assistive robotics." In: *Interaction Studies: Social Behaviour and Communication in Biological and Artificial Systems* 8.3 (2007), pp. 423–439. ISSN: 15720373. DOI: 10.1075/is.8.3.07fei. *used on: p. 124*

[Fel11] D. L. Feltz et al. "Buddy up: The Köhler effect applied to health games." In: *Journal of Sport and Exercise Psychology* 33.4 (2011), p. 506. *used on: p. 3*

[Fel14] D. L. Feltz et al. "Cyber buddy is better than no buddy: A test of the Köhler Motivation Effect in exergames." In: *GAMES FOR HEALTH: Research, Development, and Clinical Applications* 3.2 (2014), pp. 98–105. *used on: pp. 3, 26, 35–38, 40, 41, 44, 50, 64, 65, 69–75, 112, 113, 155*

[Fie12] A. Field et al. *Discovering Statistics Using R*. 2012. *used on: pp. 30, 189*

[Fis21] R. A. Fisher. "On the probable error of a coefficient of correlation deduced from a small sample." In: *Metron* 1 (1921), pp. 3–32. *used on: p. 29*

[Fon03] T. Fong et al. "A survey of socially interactive robots." In: *Robotics and autonomous systems* 42.3 (2003), pp. 143–166. *used on: p. 11*

[Fre07] A. Freedy et al. "Measurement of trust in human-robot collaboration." In: *Collaborative Technologies and Systems, 2007. CTS 2007. International Symposium on*. IEEE. 2007, pp. 106–114. *used on: p. 98*

[Fri14] M. Fridin et al. "Embodied robot versus virtual agent: Involvement of preschool children in motor task performance." In: *International Journal of Human-Computer Interaction* 30.6 (2014), pp. 459–469. *used on: pp. 15, 16, 18, 25, 165*

[Für11] J. Fürnkranz et al. "Preference Learning: An Introduction." In: *Preference Learning*. Ed. by J. Fürnkranz et al. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 1–17. ISBN: 978-3-642-14125-6. DOI: 10.1007/978-3-642-14125-6_1. *used on: p. 83*

[Gad11] P. Gadde et al. "Toward monitoring and increasing exercise adherence in older adults by robotic intervention: a proof of concept study." In: *Journal of Robotics* 2011 (2011). *used on: pp. 14, 17, 18, 20, 165*

[Gao15] Z. Gao et al. "A meta-analysis of active video games on health outcomes among children and adolescents." In: *obesity reviews* 16.9 (2015), pp. 783–794. *used on: pp. 7, 23*

[Goc05] R. Gockley et al. "Designing robots for long-term social interaction." In: *Intelligent Robots and Systems, 2005.(IROS 2005). 2005 IEEE/RSJ International Conference on*. IEEE. 2005, pp. 1338–1343. *used on: p. 53*

[Goc06] R. Gockley et al. "Encouraging physical therapy compliance with a hands-off mobile robot." In: *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*. ACM. 2006, pp. 150–155. *used on: pp. 14, 16, 37, 165*

[Goo03] E. Goodman et al. “Impact of objective and subjective social status on obesity in a biracial cohort of adolescents.” In: *Obesity Research* 11.8 (2003), pp. 1018–1026. *used on: p. 127*

[Gör13] B. Görer et al. “A robotic fitness coach for the elderly.” In: *International Joint Conference on Ambient Intelligence*. Springer. 2013, pp. 124–139. *used on: pp. 13, 14, 16, 17, 20, 37, 165*

[Gor16] G. Gordon et al. “Affective Personalization of a Social Robot Tutor for Children’s Second Language Skills.” In: *AAAI*. 2016, pp. 3951–3957. *used on: p. 82*

[Gou08] D. Gouaillier et al. “The nao humanoid: a combination of performance and affordability.” In: *CoRR abs/0807.3223* (2008). *used on: p. 12*

[Gun17] A. Guneysu et al. “Socially assistive child-robot interaction in physical exercise coaching.” In: *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. 2017-08, pp. 670–675. DOI: 10.1109/ROMAN.2017.8172375. *used on: pp. 16–18, 20, 37, 166*

[Hal12] P. C. Hallal et al. “Global physical activity levels: surveillance progress, pitfalls, and prospects.” In: *The lancet* 380.9838 (2012), pp. 247–257. *used on: p. 1*

[Ham14] J. Ham et al. “A persuasive robot to stimulate energy conservation: the influence of positive and negative social feedback and task similarity on energy-consumption behavior.” In: *International Journal of Social Robotics* 6.2 (2014), pp. 163–171. *used on: pp. 7, 53*

[Han11] P. A. Hancock et al. “A meta-analysis of factors affecting trust in human-robot interaction.” In: *Human Factors* 53.5 (2011), pp. 517–527. *used on: pp. 99, 100*

[Har07] D. Harvey. *A brief history of neoliberalism*. Oxford University Press, USA, 2007. *used on: p. 127*

[Har14] Y. N. Harari. *Sapiens: A brief history of humankind*. Random House, 2014, pp. 433–444. *used on: p. 128*

[Hau04] H. A. Hausenblas et al. “Self-presentation and exercise.” In: *Journal of Applied Sport Psychology* 16.1 (2004), pp. 3–18. *used on: p. 3*

[Hem17] J. Hemminghaus et al. “Towards Adaptive Social Behavior Generation for Assistive Robots Using Reinforcement Learning.” In: *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. ACM. 2017, pp. 332–340. *used on: pp. 81, 82*

[Hil24] A. V. Hill et al. “Muscular exercise, lactic acid, and the supply and utilisation of oxygen.—Parts -.” In: *Proc. R. Soc. Lond. B* 96.679 (1924), pp. 438–475. *used on: p. 116*

[Ho10] C.-C. Ho et al. “Revisiting the uncanny valley theory: Developing and validating an alternative to the Godspeed indices.” In: *Computers in Human Behavior* 26.6 (2010), pp. 1508–1518. *used on: p. 27*

[Hof18] L. Hoffmann et al. "The Peculiarities of Robot Embodiment (EmCorp-Scale): Development, Validation and Initial Test of the Embodiment and Corporeality of Artificial Agents Scale." In: *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. ACM. 2018, pp. 370–378. *used on: p. 125*

[Hol79] S. Holm. "A simple sequentially rejective multiple test procedure." In: *Scandinavian journal of statistics* (1979), pp. 65–70. *used on: p. 30*

[Hoo15] D. Hood et al. "When Children Teach a Robot to Write: An Autonomous Teachable Humanoid Which Uses Simulated Handwriting." In: *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*. HRI '15. Portland, Oregon, USA: ACM, 2015, pp. 83–90. DOI: 10.1145/2696454.2696479. *used on: p. 51*

[Hor89] A. O. Horvath et al. "Development and validation of the Working Alliance Inventory." In: *Journal of counseling psychology* 36.2 (1989), p. 223. *used on: pp. 101, 104*

[Hua02] W. Huang et al. "Camera Angle Affects Dominance in Video-mediated Communication." In: *CHI '02 Extended Abstracts on Human Factors in Computing Systems*. CHI EA '02. Minneapolis, Minnesota, USA: ACM, 2002, pp. 716–717. DOI: 10.1145/506443.506562. *used on: p. 33*

[Iol13] C. M. Iolanda Leite et al. "Social Robots for Long-Term Interaction: A Survey." In: *I. J. Social Robotics 5(2)* (2013), pp. 291–308. *used on: pp. 95, 111*

[Irw12] B. C. Irwin et al. "Aerobic exercise is promoted when individual performance affects the group: a test of the Kohler motivation gain effect." In: *Annals of Behavioral Medicine* 44.2 (2012), pp. 151–159. *used on: pp. 3, 21, 44*

[Irw13] B. C. Irwin et al. "Silence is golden: Effect of encouragement in motivating the weak link in an online exercise video game." In: *Journal of medical Internet research* 15.6 (2013). *used on: pp. 55, 64, 112*

[Jun09] Y. Jung et al. "Games for a Better Life: Effects of Playing Wii Games on the Well-being of Seniors in a Long-term Care Facility." In: *Proceedings of the Sixth Australasian Conference on Interactive Entertainment*. IE '09. Sydney, Australia: ACM, 2009, 5:1–5:6. DOI: 10.1145/1746050.1746055. *used on: p. 2*

[Kah01] L. E. Kahn et al. "Comparison of robot-assisted reaching to free reaching in promoting recovery from chronic stroke." In: *Proceedings of the International Conference on Rehabilitation Robotics*. IOS Press. 2001, pp. 39–44. *used on: p. 11*

[Kas17] S. Kashi et al. "Smooth leader or sharp follower? Playing the mirror game with a robot." In: *Restorative neurology and neuroscience* Preprint (2017), pp. 1–13. *used on: pp. 16, 18, 166*

[Ken15] J. Kennedy et al. "Comparing robot embodiments in a guided discovery learning interaction with children." In: *International Journal of Social Robotics* 7.2 (2015), pp. 293–308. *used on: pp. 67, 68*

[Ken91] D. Kendzierski et al. “Physical Activity Enjoyment Scale: Two validation studies.” In: *Journal of Sport & Exercise Psychology* (1991). *used on: p. 43*

[Ker07] N. L. Kerr et al. “Psychological mechanisms underlying the Köhler motivation gain.” In: *Personality and Social Psychology Bulletin* 33.6 (2007), pp. 828–841. *used on: p. 3*

[Ker11] N. L. Kerr et al. “The Köhler group motivation gain: How to motivate the ‘weak links’ in a group.” In: *Social and Personality Psychology Compass* 5.1 (2011), pp. 43–55. *used on: pp. 21, 44*

[Kid08a] C. D. Kidd et al. “Robots at home: Understanding long-term human-robot interaction.” In: *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*. 2008, pp. 3230–3235. DOI: 10.1109/IROS.2008.4651113. *used on: p. 79*

[Kid08b] C. D. Kidd et al. “Robots at home: Understanding long-term human-robot interaction.” In: *Intelligent Robots and Systems, 2008. IROS 2008. IEEE/RSJ International Conference on*. IEEE. 2008, pp. 3230–3235. *used on: pp. 4, 53, 104*

[King92] A. C. King et al. “Determinants of physical activity and interventions in adults.” In: *Medicine & Science in Sports & Exercise* (1992). *used on: p. 19*

[Koh26] O. Kohler. “Kraftleistungen bei Einzel-und Gruppenabeit [Physicalperformance in individual and group situations].” In: *Industrielle Psychotechnik* 3 (1926), pp. 274–282. *used on: p. 21*

[Kom15] J. Komiyama et al. “Regret lower bound and optimal algorithm in dueling bandit problem.” In: *Conference on Learning Theory*. 2015, pp. 1141–1154. *used on: p. 84*

[Kom16] J. Komiyama et al. “Copeland dueling bandit problem: Regret lower bound, optimal algorithm, and computationally efficient algorithm.” In: *arXiv preprint arXiv:1605.01677* (2016). *used on: p. 84*

[Kör08] A. Körner et al. “Persönlichkeitsdiagnostik mit dem NEO-Fünf-Faktoren-Inventar: Die 30-Item-Kurzversion (NEO-FFI-30).” In: *PPmP-Psychotherapie· Psychosomatik· Medizinische Psychologie* 58.06 (2008), pp. 238–245. *used on: p. 87*

[KPM18] KPMG. *Social Robots*. 2018-09-01. URL: https://assets.kpmg.com/content/dam/kpmg/pdf/2016/06/social-robots.pdf. *used on: p. 128*

[Kru52] W. H. Kruskal et al. “Use of ranks in one-criterion variance analysis.” In: *Journal of the American statistical Association* 47.260 (1952), pp. 583–621. *used on: p. 29*

[Kuc12] D. Kuchenbrandt et al. “Keep an eye on the task! How gender typicality of tasks influence human–robot interactions.” In: *Social Robotics*. Springer, 2012, pp. 448–457. *used on: pp. 26, 33*

[Lam14] L. Lammer et al. “Mutual care : How older adults react when they should help their care robot.” In: *AISB2014: Proceedings of the 3rd International Symposium on New Frontiers in Human-Robot Interaction* June 2016 (2014). *used on: p. 120*

[Lee06] K. M. Lee et al. "Can robots manifest personality?: An empirical test of personality recognition, social responses, and social presence in human–robot interaction." In: *Journal of communication* 56.4 (2006), pp. 754–772. *used on: pp. 26, 33*

[Lee12] M. K. Lee et al. "Personalization in HRI: A longitudinal field experiment." In: *7th ACM/IEEE International Conference on Human-Robot Interaction*. 2012-03. *used on: pp. 81, 82*

[Lee14] J. Y. Lee et al. "The Effect of Human Intervention Levels of a Robot on People's Motivation." In: *International Journal of Robots, Education and Art* 4.1 (2014), p. 9. *used on: pp. 13, 15, 16*

[Lei11] I. Leite et al. "Modelling empathy in social robotic companions." In: *International Conference on User Modeling, Adaptation, and Personalization*. Springer. 2011, pp. 135–147. *used on: pp. 97, 124*

[Lei12] I. Leite et al. "Modelling empathy in social robotic companions." In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 2012, pp. 135–147. DOI: 10.1007/978-3-642-28509-7_14. *used on: pp. 80–82*

[Lei13] I. Leite et al. "Social Robots for Long-Term Interaction: A Survey." In: *International Journal of Social Robotics* 5.2 (2013), pp. 291–308. ISSN: 18754791. DOI: 10.1007/s12369-013-0178-y. *used on: p. 79*

[Lei14] I. Leite et al. "Empathic robots for long-term interaction." In: *International Journal of Social Robotics* 6.3 (2014), pp. 329–341. *used on: pp. 7, 53, 55*

[Lev61] H. Levene. "Robust tests for equality of variances." In: *Contributions to probability and statistics. Essays in honor of Harold Hotelling* (1961), pp. 279–292. *used on: p. 29*

[Lew16] L. Lewis et al. "Evaluating Human-Robot Interaction Using a Robot Exercise Instructor at a Senior Living Community." In: *International Conference on Intelligent Robotics and Applications*. Springer. 2016, pp. 15–25. *used on: pp. 16, 17, 166*

[Ley12] D. Leyzberg et al. "The physical presence of a robot tutor increases cognitive learning gains." In: *Proceedings of the 34th Annual Conference of the Cognitive Science Society. Austin, TX: Cognitive Science Society*. 2012. *used on: pp. 25, 32, 35, 68, 93*

[Ley14a] D. Leyzberg et al. "Personalizing Robot Tutors to Individuals Learning Differences." In: *Proceedings of the 2014 ACM/IEEE International Conference on Human-robot Interaction*. HRI '14. Bielefeld, Germany: ACM, 2014, pp. 423–430. DOI: 10.1145/2559636.2559671. *used on: pp. 53, 55, 79, 81, 82*

[Ley14b] D. Leyzberg et al. "Personalizing robot tutors to individuals' learning differences." In: *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*. ACM. 2014, pp. 423–430. *used on: pp. 4, 53, 97*

[Li15] J. Li. "The benefit of being physically present: A survey of experimental works comparing copresent robots, telepresent robots and virtual agents." In: *International Journal of Human-Computer Studies* 77 (2015), pp. 23–37. *used on: pp. 23, 25, 26, 34, 38, 67, 68, 81, 91, 92, 127*

[Lim13] G. H. Lim et al. "Robot recommender system using affection-based episode ontology for personalization." In: *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*. 2013, pp. 155–160. DOI: 10.1109/ROMAN.2013.6628437. *used on: pp. 81, 82*

[Lot17] A. Lotfi et al. "Active and healthy ageing: development of a robotic platform as an exercise trainer." In: *Proceedings of the 10th International Conference on PErvasive Technologies Related to Assistive Environments*. ACM. 2017, pp. 275–279. *used on: pp. 15–17, 20, 166*

[Lup14] D. Lupton. "Self-tracking cultures: towards a sociology of personal informatics." In: *Proceedings of the 26th Australian Computer-Human Interaction Conference on Designing Futures: the Future of Design*. ACM. 2014, pp. 77–86. *used on: p. 127*

[Mad00] M. Madsen et al. "Measuring human-computer trust." In: *11th australasian conference on information systems*. Citeseer. 2000, pp. 6–8. *used on: p. 99*

[Mar18] D. Marriam-Webster. *The Merriam-Webster Dictionary*. Merriam-Webster, Incorporated, 2018. *used on: p. 99*

[McA95] D. J. McAllister. "Affect-and cognition-based trust as foundations for interpersonal cooperation in organizations." In: *Academy of management journal* 38.1 (1995), pp. 24–59. *used on: p. 99*

[Men17] L. Menvielle et al. "The digitization of healthcare: New challenges and opportunities." In: *The Digitization of Healthcare: New Challenges and Opportunities* (2017), pp. 1–454. DOI: 10.1057/978-1-349-95173-4. *used on: pp. 16, 125*

[Mid09] C. Midden et al. "Using negative and positive social feedback from a robotic agent to save energy." In: *Proceedings of the 4th international conference on persuasive technology*. ACM. 2009, p. 12. *used on: p. 5*

[Mit05] N. Mitsunaga et al. "Robot behavior adaptation for human-robot interaction based on policy gradient reinforcement learning." In: *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*. 2005-08, pp. 218–225. DOI: 10.1109/IROS.2005.1545206. *used on: pp. 80–82*

[Mit08] N. Mitsunaga et al. "Adapting robot behavior for human-robot interaction." In: *IEEE Transactions on Robotics* 24.4 (2008), pp. 911–916. *used on: p. 97*

[Mon16] G. Monbiot. "Neoliberalism–the ideology at the root of all our problems." In: *The Guardian* 15.04 (2016). *used on: p. 127*

[Moo00] R. J. Mooney et al. "Content-based book recommending using learning for text categorization." In: *Proceedings of the fifth ACM conference on Digital libraries*. ACM. 2000, pp. 195–204. *used on: p. 95*

[Nas96] C. Nass et al. "Can computers be teammates?" In: *International Journal of Human-Computer Studies* 45.6 (1996), pp. 669–678. *used on: pp. 49, 58*

[Nis07] S. Nishio et al. "Geminoid: Teleoperated android of an existing person." In: *Humanoid robots: New developments*. InTech, 2007. *used on: p. 126*

[Noa05] T. D. Noakes et al. "From catastrophe to complexity: a novel model of integrative central neural regulation of effort and fatigue during exercise in humans: summary and conclusions." In: *British journal of sports medicine* 39.2 (2005), pp. 120–124. *used on: p. 116*

[Nom06] T. Nomura et al. "Experimental investigation into influence of negative attitudes toward robots on human–robot interaction." In: *Ai & Society* 20.2 (2006), pp. 138–150. *used on: p. 58*

[Org10] W. H. Organisation. *Increasing access to health workers in remote and rural areas through improved retention*. 2010. URL: http://www.who.int/hrh/retention/guidelines/en/ (visited on 2018-09-01). *used on: p. 127*

[Org13] W. H. Organisation. *Health Workforce Shortage*. 2013. URL: http://www.who.int/mediacentre/news/releases/2013/health-workforce-shortage/en/ (visited on 2018-09-01). *used on: p. 127*

[Par11] E. Park et al. "The effects of a robot instructor's positive vs. negative feedbacks on attraction and acceptance towards the robot in classroom." In: *Social robotics*. Springer, 2011, pp. 135–141. *used on: p. 55*

[Par16] C. Park et al. "Robot social skills for enhancing social interaction in physical training." In: *The Eleventh ACM/IEEE International Conference on Human Robot Interaction*. IEEE Press. 2016, pp. 493–494. *used on: pp. 15, 16, 20, 166*

[Plo00] D. J. Plonczynski. "Measurement of motivation for exercise." In: *Health Education Research* 15.6 (2000), pp. 695–705. *used on: p. 19*

[Pot09] E. Pot et al. "Choregraphe: a graphical tool for humanoid robot programming." In: *Robot and Human Interactive Communication, 2009. RO-MAN 2009. The 18th IEEE International Symposium on*. IEEE. 2009, pp. 46–51. *used on: p. 12*

[Pow03] A. Powers et al. "Matching robot appearance and behavior to tasks to improve human-robot cooperation." In: *Human-Computer Interaction Institute* (2003), p. 105. *used on: pp. 13, 16, 26, 33, 49, 165*

[Pre08] K. J. Preacher et al. "Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models." In: *Behavior research methods* 40.3 (2008), pp. 879–891. *used on: pp. 108, 170*

[Pre10] M. Prem et al. "Recommender system." In: *Encyclopedia of Machine Learning* (2010), pp. 1–9. *used on: p. 95*

[Qui09] M. Quigley et al. “ROS: an open-source Robot Operating System.” In: *ICRA workshop on open source software*. 3.2. Kobe, Japan. 2009, p. 5. *used on: p. 12*

[R C13] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2013. URL: http://www.R-project.org/. *used on: pp. 30, 189*

[Rau13] P.-L. P. Rau et al. “Effects of a social robot’s autonomy and group orientation on human decision-making.” In: *Advances in Human-Computer Interaction* 2013 (2013), p. 11. *used on: pp. 99, 100, 109*

[Ree97] B. Reeves et al. “The media equation: How people treat computers, television and new media like real people.” In: *Computers and Mathematics with Applications* 5.33 (1997), p. 128. *used on: pp. 1, 4, 33, 119, 155*

[Rho06] R. Rhodes et al. “Personality correlates of physical activity: a review and meta-analysis.” In: *British journal of sports medicine* 40.12 (2006), pp. 958–965. *used on: pp. 5, 93*

[Rie12] N. Riether et al. “Social facilitation with social robots?” In: *Human-Robot Interaction (HRI), 2012 7th ACM/IEEE International Conference on*. IEEE. 2012, pp. 41–47. *used on: p. 50*

[Rit17] H. Ritschel et al. “Adapting a Robot’s Linguistic Style Based on Socially-Aware Reinforcement Learning.” In: *Robot and Human Interactive Communication (RO-MAN), 2017 26th IEEE International Symposium on*. IEEE. 2017, pp. 378–384. *used on: pp. 81, 82*

[Ros16] A. M. Rosenthal-von der Pütten et al. “Robots or Agents–Neither Helps You More or Less During Second Language Acquisition.” In: *International Conference on Intelligent Virtual Agents*. Springer. 2016, pp. 256–268. *used on: pp. 67, 68, 91*

[Ros94] R. Rosenthal. “Parametric measures of effect size.” In: *The handbook of research synthesis* 621 (1994), pp. 231–244. *used on: p. 30*

[Rus16] S. J. Russell et al. *Artificial intelligence: a modern approach*. Malaysia; Pearson Education Limited, 2016. *used on: p. 98*

[Rya00] R. M. Ryan et al. “Intrinsic and extrinsic motivations: Classic definitions and new directions.” In: *Contemporary educational psychology* 25.1 (2000), pp. 54–67. *used on: p. 20*

[Sek13] A. Sekmen et al. “Assessment of adaptive human-robot interactions.” In: *Knowledge-Based Systems* 42 (2013), pp. 49–59. ISSN: 09507051. DOI: 10.1016/j.knosys.2013.01.003. *used on: p. 81*

[Sha65] S. S. Shapiro et al. “An analysis of variance test for normality (complete samples).” In: *Biometrika* 52.3/4 (1965), pp. 591–611. *used on: p. 29*

[Sha79] E. Shanas. “The family as a social support system in old age.” In: *The gerontologist* 19.2 (1979), pp. 169–174. *used on: p. 128*

[She78] T. B. Sheridan et al. *Human and computer control of undersea teleoperators*. Tech. rep. Massachusets Insti of Tech Cambrdge Man-Machine Systems Lab, 1978. *used on: p. 98*

[Sie88] S. Siegel et al. "Nonparametric statistics for the behavioural sciences 2. ed." In: *Statistics. New York: McGraw-Hill* (1988). *used on: p. 30*

[Str02] B. Strauss. "Social facilitation in motor tasks: a review of research and theory." In: *Psychology of sport and exercise* 3.3 (2002), pp. 237–256. *used on: pp. 50, 112, 156*

[Stu12] G. Stulp et al. "High and mighty: Height increases authority in professional refereeing." In: *Evolutionary Psychology* 10.3 (2012), p. 147470491201000314. *used on: p. 33*

[Stu13] G. Stulp et al. "Tall claims? Sense and nonsense about the importance of height of US presidents." In: *The Leadership Quarterly* 24.1 (2013), pp. 159–171. *used on: p. 33*

[Swi13] K. Swift-Spong. "Studying the Köhler Effect with a Robot Exercise Teammate." In: *Proceedings of the Human Robot Interaction Pioneers Workshop at HRI*. 2013. *used on: pp. 35, 37*

[Swi15] K. Swift-Spong et al. "Effects of comparative feedback from a Socially Assistive Robot on self-efficacy in post-stroke rehabilitation." In: *Rehabilitation Robotics (ICORR), 2015 IEEE International Conference on*. IEEE. 2015, pp. 764–769. *used on: pp. 5, 7, 15, 16, 18, 53, 55*

[Swi16] K. Swift-Spong et al. "Comparing backstories of a socially assistive robot exercise buddy for adolescent youth." In: *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE. 2016, pp. 1013–1018. *used on: pp. 4, 15, 16, 18, 20, 166*

[Tap07a] A. Tapus et al. "Hands-off therapist robot behavior adaptation to user personality for post-stroke rehabilitation therapy." In: *Proceedings - IEEE International Conference on Robotics and Automation*. 2007, pp. 1547–1553. DOI: 10.1109/ROBOT.2007.363544. *used on: pp. 14, 79–82*

[Tap07b] A. Tapus et al. "Socially assistive robotics [grand challenges of robotics]." In: *Robotics & Automation Magazine, IEEE* 14.1 (2007), pp. 35–42. *used on: p. 16*

[Tap08a] A. Tapus et al. "Socially Assistive Robots: The Link between Personality, Empathy, Physiological Signals, and Task Performance." In: *AAAI Spring Symposium* (2008), pp. 133–141. *used on: p. 165*

[Tap08b] A. Tapus et al. "User personality matching with a hands-off robot for post-stroke rehabilitation therapy." In: *Experimental robotics*. Springer. 2008, pp. 165–175. *used on: pp. 16, 49, 97*

[Tat15] D. F. Tate et al. "High-tech tools for exercise motivation: use and role of technologies such as the internet, mobile applications, social media, and video games." In: *Diabetes Spectrum* 28.1 (2015), pp. 45–54. *used on: p. 2*

[The16] S. Thellman et al. "Physical vs. Virtual agent embodiment and effects on social interaction." In: *International Conference on Intelligent Virtual Agents*. Springer. 2016, pp. 412–415. *used on: p. 68*

[Tro02] S. G. Trost et al. "Correlates of adults' participation in physical activity: review and update." In: *Medicine & Science in Sports & Exercise* 34.12 (2002), pp. 1996–2001. *used on: p. 1*

[Tsi16] K. Tsiakas et al. "An Interactive Learning and Adaptation Framework for Adaptive Robot Assisted Therapy." In: *PETRA*. 2016. DOI: 10.1145/2910674.2935857. *used on: pp. 81, 82, 97*

[Tur18] K. L. Turabian. *A manual for writers of research papers, theses, and dissertations: Chicago style for students and researchers*. University of Chicago Press, 2018. *used on: p. 189*

[Veu05] P. J. Veugelers et al. "Prevalence of and risk factors for childhood overweight and obesity." In: *Canadian Medical Association Journal* 173.6 (2005), pp. 607–613. *used on: p. 127*

[Vir13] M. Vircikova et al. "Experience with the children-humanoid interaction in rehabilitation therapy for spinal disorders." In: *Robot Intelligence Technology and Applications 2012*. Springer, 2013, pp. 347–357. *used on: pp. 13, 14, 16, 17, 165*

[Wai07] J. Wainer et al. "Embodiment and human-robot interaction: A task-based perspective." In: *The 2007 16th IEEE International Symposium on Robot and Human interactive Communication (RO-MAN)*. IEEE. 2007, pp. 872–877. *used on: pp. 25, 32, 33*

[Wal11] M. L. Walters et al. "Evaluating the Robot Personality and Verbal Behavior of Domestic Robots Using Video-Based Studies." In: *Advanced Robotics* 25.18 (2011). DOI: 10.1163/016918611X603800. *used on: pp. 26, 27, 33*

[Way14] A. Waytz et al. "The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle." In: *Journal of Experimental Social Psychology* 52 (2014), pp. 113–117. *used on: pp. 99, 109*

[Web07] B. Weber et al. *Motivation gains of inferior group members: a meta-analytical review.* 2007. *used on: pp. 3, 21*

[Wel18] D. Welle. *Germany is desperate for teachers*. 2018. URL: https://www.dw.com/en/germany-is-desperate-for-teachers/a-45246978 (visited on 2018-09-01). *used on: p. 127*

[Wel47] B. L. Welch. "The generalization ofstudent's' problem when several different population variances are involved." In: *Biometrika* 34.1/2 (1947), pp. 28–35. *used on: p. 29*

[Wer13] F. Werner et al. "Evaluation of the Acceptance of a Social Assistive Robot for Physical Training Support Together with Older Users and Domain Experts." In: *Assistive Technology: From Research to Practice: AAATE 2013* 33 (2013), p. 137. *used on: pp. 14, 17*

[WHO18] WHO. *Physical Activity*. http://www.afro.who.int/health-topics/physical-activity. 2018. (Visited on 2018-02-12). *used on: p. 2*

[Wie11] J. Wienke et al. "A middleware for collaborative research in experimental robotics." In: *2011 IEEE/SICE International Symposium on System Integration (SII)*. IEEE. 2011, pp. 1183–1190. *used on: p. 8*

[Wil10] R. Wilkinson et al. *The spirit level: Why equality is better for everyone*. Penguin UK, 2010. *used on: p. 127*

[Wil45] F. Wilcoxon. "Individual comparisons by ranking methods." In: *Biometrics bulletin* 1.6 (1945), pp. 80–83. *used on: p. 29*

[Win03] C. J. Winstein et al. "Methods for a multisite randomized trial to investigate the effect of constraint-induced movement therapy in improving upper extremity function among adults recovering from a cerebrovascular stroke." In: *Neurorehabilitation and neural repair* 17.3 (2003), pp. 137–152. *used on: p. 11*

[Win18] K. Winkle et al. "Social Robots for Engagement in Rehabilitative Therapies." In: *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction - HRI '18* (2018), pp. 289–297. ISSN: 21672148. DOI: 10.1145/3171221.3171273. *used on: p. 124*

[Wu16] H. Wu et al. "Double Thompson Sampling for Dueling Bandits." In: *Advances in Neural Information Processing Systems*. 2016, pp. 649–657. *used on: pp. 83–85*

[Zen17] N. Zeng et al. "A systematic review of active video games on rehabilitative outcomes among older patients." In: *Journal of Sport and Health Science* 6.1 (2017), pp. 33–43. ISSN: 22132961. DOI: 10.1016/j.jshs.2016.12.002. *used on: p. 23*

[Zwa01] M. de Zwaan. "Binge eating disorder and obesity." In: *International Journal of Obesity* 25.S1 (2001), S51. *used on: p. 127*

ACRONYMS

A

ACME
: Average Causal Mediation Effects. *used on: p. 108*

ADE
: Average Direct Effect. *used on: p. 108*

AI
: Artificiall Intelligence. *used on: p. 12*

ANOVA
: Analysis of Variance. *used on: pp. 29–31, 44, 46, 48, 59, 61, 88–90*

API
: Application Programming Interface. *used on: p. 12*

AR
: Assistive Robots. *used on: p. xiii, 11*

AVG
: Active Video Game. *used on: pp. 2, 23, 121*

B

BN
: Bayesian Network. *used on: pp. 81, 82*

BWE
: Bodyweight Exercise. *used on: pp. 7, 8*

C

CBF
: Content-based Filtering. *used on: p. 95*

CF
: Collaborative Filtering. *used on: pp. 81, 82, 95, 96*

CRC
: co-located Robot Companion. *used on: pp. 24, 28, 31*

CRI
: co-located Robot Instructor. *used on: pp. 24, 28, 31*

D

DBP
: Dueling Bandit Problem. *used on: p. 112*

DoF
: Degree of Freedom. *used on: pp. 38, 40, 69*

DSL
: Domain Specific Language. *used on: p. 8*

DTS
: Double Thompson Sampling. *used on: pp. 83–85, 91*

E

ECW-RMED
Efficient Copeland Winners Relative Minimum Empirical Divergence. *used on: p. 84*

ER
Embodied Robot. *used on: pp. 7, 67, 127*

ET
Embodied Trainer. *used on: pp. 16, 125*

ETSAR
Exercise Trainer Socially Assistive Robot. *used on: p. 15*

F

FET
Fisher's Exact Test. *used on: pp. 47, 91*

G

GUI
Graphical User Interace. *used on: pp. 40, 85, 87, 90, 92, 93*

H

HCI
Human-Computer Interaction. *used on: pp. 16, 99, 104, 125*

HHI
Human-Human Interaction. *used on: pp. v, 1, 3*

HHP
Hardly Human Partner. *used on: pp. 37, 69–75*

HP
Human Partner. *used on: pp. 37, 38, 69, 71–76*

HRI
Human-Robot Interaction. *used on: pp. 1, 3, 8–10, 18, 23, 26–28, 33–35, 53, 64, 77, 79–82, 84, 91, 93, 95–100, 104, 110, 112, 114, 121*

I

IC
Individually Exercising. *used on: pp. 37–39, 41, 44, 45, 49, 51, 56, 58–60, 63, 69–72*

K

KW-Test
Kruskal-Wallis Rank Sum Test. *used on: pp. 29–31, 44, 46, 47, 59, 71–74, 88, 90, 91*

L

LAW
Lethal Autonomous Weapon. *used on: p. 98*

LoA
Level of Automation. *used on: pp. 6, 96, 98–100, 109, 110*

M

MAB

Mulit-armed Bandit. *used on: pp. 80–83*

N

NARS

Negative Attributes Towards Robots Scale. *used on: pp. xiii, 57, 60, 62, 104, 125*

NCDs

Noncommunicable Diseases. *used on: pp. v, 2, 122, 124*

NHP

Nearly Human Partner. *used on: pp. 37, 69, 71, 72, 74–76*

NHS

National Health Services. *used on: p. 15*

P

PA

physical activity. *used on: pp. v, 1–3, 5, 7, 11, 36, 79, 110, 111, 119, 120, 122, 124, 126, 127*

PAES

Physical Activity Enjoyment Scale. *used on: pp. xiv, 43, 45, 48, 73, 104, 105, 125*

PL

Preference Learning. *used on: pp. 9, 80–85, 88, 91, 93, 112, 114*

Q

QOL

Quality of Life. *used on: p. v, 1*, Glossary: *Quality of Life*

R

RC

Robot Companion. *used on: pp. 4–6, 24, 28, 36–39, 41, 44–47, 49–51, 53, 54, 56–60, 63, 69, 70, 72–75, 111, 113, 116, 118*

RCF

Robot Companion with Feedback. *used on: pp. 5, 54, 56, 57, 59–63, 69–75*

RI

Robot Instructor. *used on: pp. 4–6, 24, 28, 36–39, 41, 42, 45–47, 49–51, 54, 56–61, 63, 69–71, 74, 75, 111, 113, 115, 118*

RIF

Robot Instructor with Feedback. *used on: pp. 5, 54, 56, 57, 59–63, 69, 70, 72, 73, 118*

RL

Reinforcement Learning. *used on: pp. 81–83*

RMED

Relative Minimum Empirical Divergence. *used on: p. 84*

ROS

Robot Operating System. *used on: p. 12*

RoSAS
Robotic Social Attribute Scale. *used on: pp. xiv, 104, 106, 109, 117, 119, 125*

RRC
remote-located Robot Companion. *used on: pp. 24, 28, 31, 32*

RRI
remote-located Robot Instructor. *used on: pp. 24, 28, 31*

RSB
Robot Service Bus. *used on: p. 8*

S

SAR
Socially Assistive Robot. *used on: pp. v, xiii, xvi, 1, 3–17, 19–25, 35, 37, 49, 51, 53–55, 63–65, 67, 68, 76, 79, 80, 98, 99, 101, 104, 109, 111–113, 119, 122, 124–128, 131*, Glossary: *Socially Assistive Robot*

SCT
Social Cognitive Theory. *used on: pp. 19, 120*

SDK
Software Development Kit. *used on: p. 12*

SDT
Self-Determination Theory. *used on: p. 20*

SIR
Socially Interative Robots. *used on: p. xiii, 11*

SUS
System Usability Scale. *used on: pp. xiv, 87, 89, 90, 92, 105*

SW-Test
Shapiro-Wilk test. *used on: p. 29*

U

UCB
Upper Confidence Bounds. *used on: p. 80*

USC
University of Souther California. *used on: pp. 13, 16*

V

VA
Virtual Agent. *used on: pp. 5, 7, 10, 23, 25, 35–38, 50, 53, 67, 69, 81, 112, 121*

vHRI
video Human-Robot Interaction. *used on: pp. xiii, 10, 23, 24, 26, 27, 31–35, 112*

W

WAI
Working Alliance Inventory. *used on: pp. xiv, 104, 105, 108, 109*

WC-Test
Wilcoxon Rank Sum Test. *used on: pp. 29, 30, 45, 71, 105*

WHO
World Health Organization. *used on: pp. v, 2, 3, 128*

WoZ
Wizard of Oz. *used on: pp. 82, 102*

GLOSSARY

A

adaptability

Adaptability is the system's ability to be adjustable by the user. *used on: pp.* *96*, *97*, *105*, *106*, *112*, *117*

adaptivity

Adaptivity is the system's ability to adapt automatically to changing users or conditions. *used on: pp.* *v*, *1*, *9*, *96*, *97*, *105*, *106*, *110*–*112*, *114*, *117*

C

cold start

Cold start is the problem when an information system cannot draw any inference for users about which it has not yet gathered information. *used on: pp.* *83*, *93*, *95*, *96*

E

embodiment

Embodiment refers to the agents representation. It can either be physically or virtually embodied. *used on: pp.* *v*, *xvi*, *1*, *4*, *7*, *9*, *23*–*28*, *30*–*34*, *38*, *65*, *67*, *68*, *74*, *76*, *80*, *84*, *91*–*93*, *95*, *111*–*114*, *117*, *118*, *131*

encouragement

Encouragement is type of feedback used in this thesis. This type of feedback only acknowledges the current state and does not include comparative, positive or negative feedback. *used on: pp.* *v*, *xiii*, *1*, *5*, *9*, *53*, *54*, *56*, *60*–*62*, *111*–*114*, *116*, *117*, *131*

K

Köhler Effect

In this effect "the least capable group member exhibits a motivation gain (relative to individual performance) when performing as part of a group on effort-based tasks" [Fel14, p. 99]. *used on: pp.* *xiii*, *3*, *4*, *6*, *9*, *21*, *22*, *35*–*38*, *41*, *44*–*47*, *49*–*51*, *53*, *55*, *64*, *65*, *69*, *112*, *113*, *115*, *116*, *119*, *120*, *126*

M

Media Equation

The media equation is a theory from Reeves et al. [Ree97] that states that people treat media as people. *used on: p.* *4*

Q

Quality of Life

Quality of Life (QOL) is the general well-being of individuals and societies, outlining negative and positive features of life. *used on: pp. v, 1, 151*

S

self-efficacy belief

Self-efficacy belief is a concept from the Social Cognitive Theory (SCT). It states that motivation is a cognitive function and that self-efficacy is the primary mediator for behavior change [Ban77]. *used on: pp. xiv, xv, 19, 21, 22, 61, 113, 120*

social facilitation effect

The social facilitation or audition effect states that people behave differently in the presence of others [Str02]. *used on: pp. 112, 113, 115*

social robot

Social robots are "embodied agents that are part of a heterogeneous group: a society of robots or humans. They are able to recognize each other and engage in social interactions, they possess histories (perceive and interpret the world in terms of their own experience), and they explicitly communicate with and learn from each other." [Dau99b] *used on: pp. 96, 111, 129*

social role

The social role in this thesis refers to the role of an exercising companion that is coactively working out with the human or an exercising instructor, that is not exercising with the human but instructing them. *used on: pp. v, 1, 4, 5, 9, 23, 24, 26–28, 30–35, 45, 56, 65, 111, 112, 114, 117, 118, 131*

Socially Assistive Robot

Socially Assistive Robot (SAR) are a type of social robot. They are defined as"the intersection of AR and SIR" [Fei05, p. 465]. Thus, they assist people by there social presence. *used on: pp. v, 1, 9, 11, 22, 23, 67, 131, 152*

T

trust

An assured reliance on the character, ability, strenght or truth of someone or something. In the course of this thesis trust is measured as the alliance with the robot using the Working Alliance Scale scale. *used on: pp. 98–101, 104, 110, 114*

α Cronbach's alpha is a realiability measure of a psychometric test. *used on: pp. 29, 30, 45, 46, 48, 59–61, 104, 105*

d Cohen's d (an effect size measure). *used on: pp. 45, 60, 61, 105–108*

χ^2 Chi-square test statistics. *used on: pp. 32, 47, 107*

η_p^2 Partial eta-squared (an effect size measure). *used on: pp. 30–32, 44, 48, 59, 88–91*

F F-ratio (test statistics used in ANOVA). *used on: pp. 30–32, 44, 48, 59, 88–91*

H Kruskal-Wallis test statistics. *used on: pp. 31, 44, 46, 47, 59, 72, 73, 88, 90, 91*

ω^2 Omega squared (an effect size measure). *used on: pp. 30, 44, 59, 88–91*

r Pearson r correlation coefficient (an effect size measure). *used on: pp. 46, 71, 105–108*

t Test statistic for Student's t-test. *used on: pp. 60, 61, 104–108*

W_s Test statistic for the Shapiro-Wilk test and the Wilcoxon's rank-sum test. *used on: pp. 46, 71, 104–107*

APPENDICES

A SYSTEM OVERVIEW

A.1 SYSTEM COMPOSITION

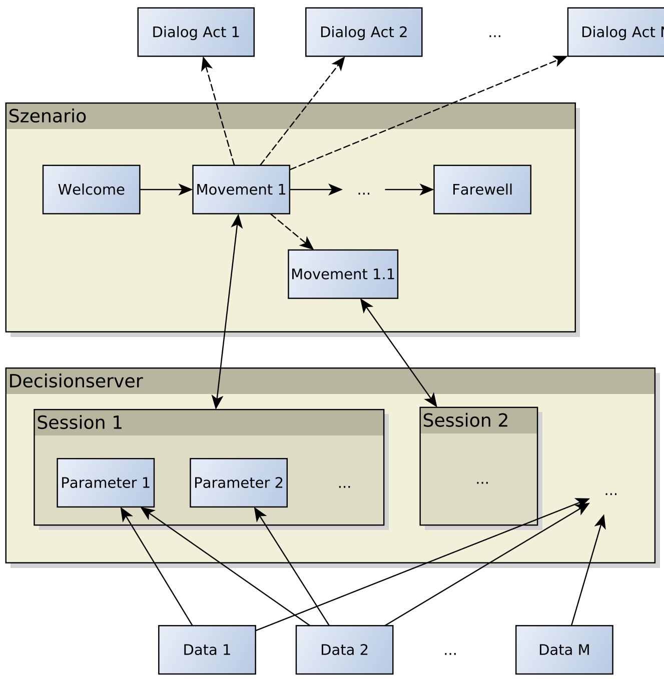


Figure A.1.1: Used framework to design the socially assistive scenario in this thesis. It is composed of a scenario manager that manages the interaction. It configures state transitions based on events happening in a distributed system. The decision server is configured according to the configuration of the scenario and sends new decision information based on the received data. The scenario manager can trigger state changes based on this decisions which leads to new dialog acts. These acts trigger movements and speech utterances of the robot. Responses from the user, either motion or verbal responses, are in turn received by the decision server as data inputs and thus closes the (inter-)action-perception loop.

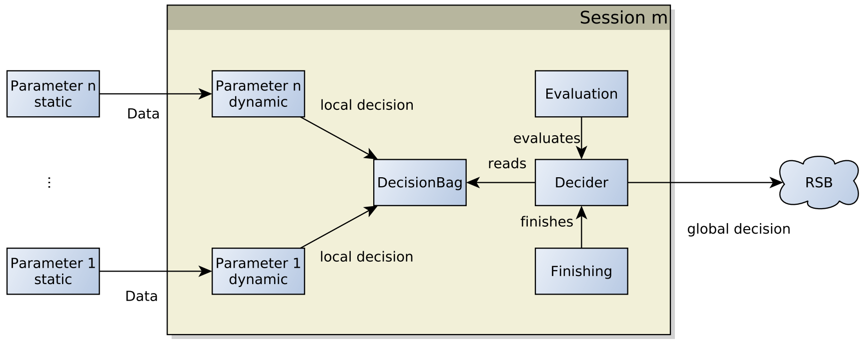


Figure A.1.2: Overview of the decision system that triggers state transitions between the states. The decision server collects the data in a decision bag. This bag is read by an decider that sends out decisions based on the chosen evaluation or finishing methods.

A.2 INTERACTION MODELS

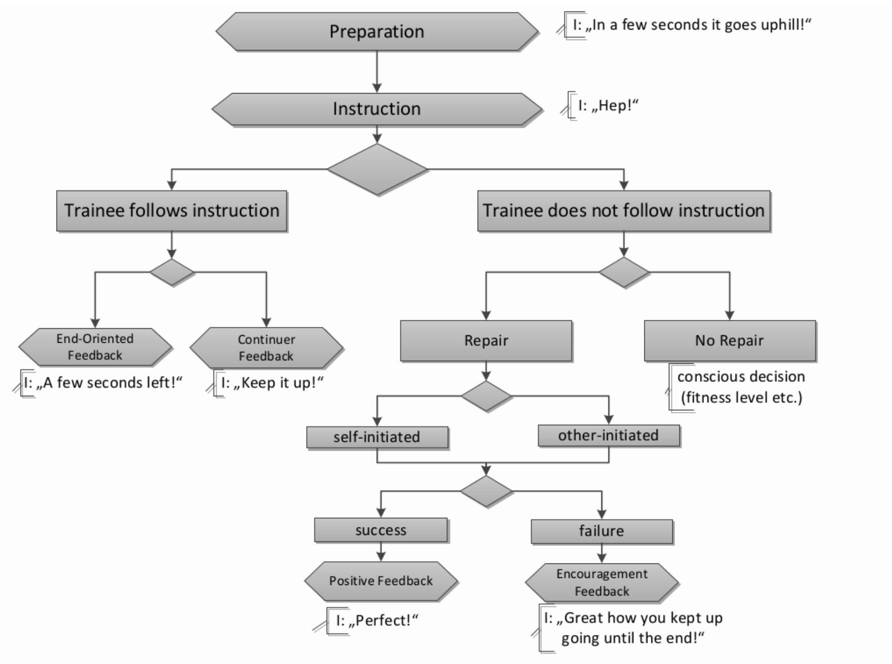


Figure A.2.3: Interactive action-based motivation model that captures the basic motivational interaction strategy of a coach in a training session. It is the basis for the developed interaction model. See [Süs14] for an in-depth explanation.

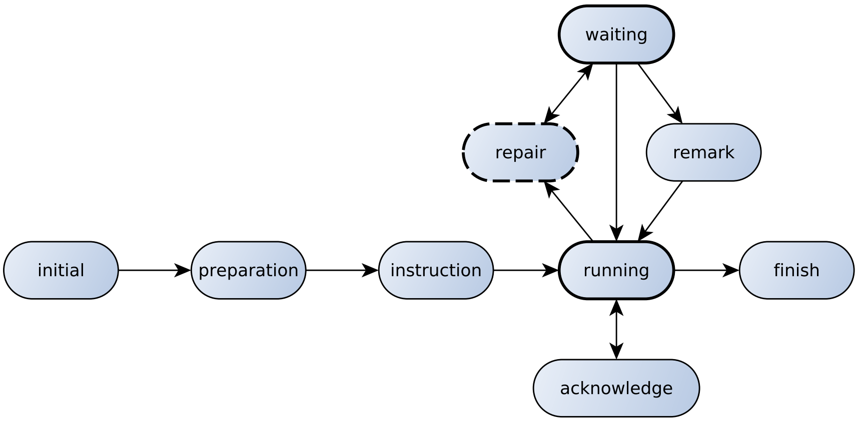


Figure A.2.4: Implementation of the interactive action-based motivational model as a state machine for static exercising movements (e.g., holding a plank position).

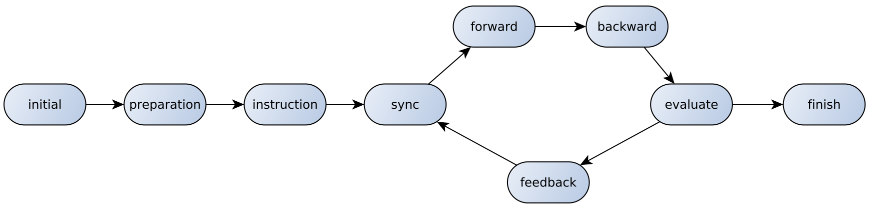
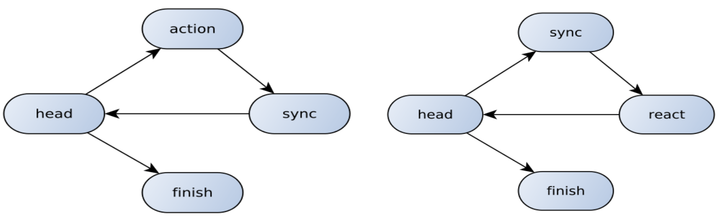


Figure A.2.5: Implementation of the interactive action-based motivational model as a state machine for cycling/repeating exercising movements (e.g., movements like pushups or squats, going up and down).



(a) The act-action. (b) The react-action.

Figure A.2.6: Two possible actions for the *forward-backward* states of the cyclic movement. In **act-actions** the robot starts the exercises and waits for the user to follow. During **react-actions**, the robot follows the user's lead.

Figure A.2.7: Implementation of the interactive action-based motivational model as a state machine for static exercising movements without any acknowledgement and reparation as used in Chapter 4 (e.g., holding a plank position).

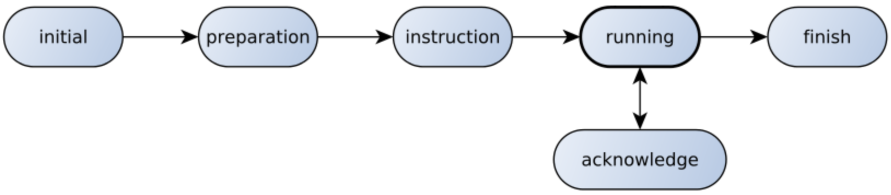


Figure A.2.8: Implementation of the interactive action-based motivational model as a state machine for static exercising movements with an acknowledgement node to trigger encouraging feedback as used in the Chapter 5 (e.g., holding a plank position).

B
RELATED WORK

Table B.1: Research on SAR for exercising and physical rehabilitation.

Reference	Task	Measures	Supportive Behavior	Robot	Sample Size (N) / Subject's Age (year)	Conditions, Interaction Time (min)	Results
[Pow03]	breathing, stretching, balancing	duration	encouraging remarks every five seconds	Nurse Bot Pearl	n=21, average 25y	playful vs. serious robot	participants exercised longer with serious robot
[Eri05]	post-stroke rehabilitation, putting books in a shelf	arm angles, no questionnaire information	feedback (no specific information)	Pioneer 2-DX	n=6, middle-aged	sound feedback vs. synthesized voice vs. recorded voice	compliance higher with robot
[Goc06]	moving pens, lifting magazines, flipping through pages	arm angles, no questionnaire information	feedback (no specific information)	Pioneer 2-DX	n=11, university students	2 (proxemics) x 2 (engagement) 6 min	people exercises longer with an engaged robot, negative correlation between extroversion and proxemics.
[Tap08a]	moving pencils	Eysenck Personality Inventory	adapt to user personality	Pioneer 2-DX	n=11, 19-35y	maximum 15 min	user prefer personality matching, robot behavior adaptation to user personality and performance is effective
[Gad11]	exercise demonstration	usability measures ,no performance measures	positive feedback	RoboPhilo	n=10, no demographics	no conditions 10 min	supports feasibility of approach
[Fas12]	exercise games using arm raising	arm position	corrections, praise, guidance, encouragement	Bandit	n=13 77 to 92y / n=24 68 to 89y	relational vs. non-relational/ activity choice vs. no choice	praise and relational discourse preferred, no difference for choice condition
[Fas13]	exercise games using arm raising	arm position	relational vs. non-relational robot, embodied vs. virtual robot	Bandit	n=14/77-92y/ n=37 68-88y	relational vs. non-relational robot/VA vs. ER 10 min/5x20 min	relational and embodied robot preferred
[Gör13]	stretching and strength exercises	arm position game experience questionnaire	verbal feedback on the success of exercises	Nao	n=8 25 - 35y	none	survey showed participants enjoy interacting with the robot
[Vir13]	exercises for scoliosis reduction	post-video analysis: interest, docility, joy of exercise, correctness, repeatability	no information	Nao	n=50 5-7y	no conditions, 20min	joy of children rose from the interaction, the accuracy of the motions and their repeatability
[Fri14]	repeating movements	eye contact, emotional reaction	verbal reinforcing feedback	Nao	children (w/o experience)	VA vs. ER	experienced group involvement in motor task was induced by both conditions, but children interacted less with VA

Table B.1: Research on SAR for exercising and physical rehabilitation.

Reference	Task	Measures	Supportive Behavior	Robot	Sample Size (N) / Subject's Age (year)	Conditions, Interaction Time (min)	Results
[Lew16]	warm-up/exercise routine	motivating, highly interactive, intelligent, task driven	compliments	Nao	n=6 elderly residents	-	seniors modereatley accept robots, nurses are enthusiastic
[Fan16]	Simon says game	arm joints subjective measurements	no supportive behavior	Nao	n=8 elderly	-	survey showed participants enjoyed interacting with the robot
[Par16]	yoga poses	modified Godspeed	motivational feedback on user performance	Nao	n=28 age 20-60y	social vs. no-social skills	robot social skills should be considered as effective social cues in physical training.
[Gun17]	arm rehabilitation	arm position Interpersonal Attraction Scale	motivational feedback and real time guidance	Nao	n=19 children	-	children engaged in physical exercise throughout the interaction sessions and rated the interaction highly in terms of enjoyableness, social attraction, social presence, and companionship
[Lot17]	exercises recommended by NHS	skeleton	showing videos, exercise recognition and giving video and audio feedback and smiley	Double robot	no information , 10 min	-	system engaged the participants, feedback was appropriate and timely
[Swi16]	different exercises	pre- and post-enjoyment of physical activity, perceived stress, and intrinsic motivation	congratulatory, encouraging, feedback, and testimonial	Nao	adolescent 11-14y	realistic vs. fictional backstory	no differences between the ratings for the robot characters, participants reacted positively to the robot exercise buddy
[Kas17]	mirror game	Godspeed questionnaire and own questions	no supportive behavior	robot arm	n=23 age 18-29	VA vs. ER	no real difference between ER and VA

C

SUPPLEMENTARY RESULTS

C.1 SELF EFFICACY BELIEFS

Studies in Chapter 4 and 5 assessed the self-efficacy belief before and after the study. Participants were asked to rate how long they belief to persist the exercises in seconds. Thus, difference in these ratings might indicate whether participants that exercised together with the robot had higher self-efficacy beliefs. The difference between prior and post average self-efficacy beliefs were calculated to measure for any difference between the conditions. A KW-Test showed a significant main effect, $H(4) = 10.41$, $p = .03$. Though, post hoc tests revealed no significant differences between the five conditions. Figure C.1.9 shows the difference between post and prior self-efficacy beliefs.

C.1.1 *Factors Influencing Exercising Time on Block 1*

The data from Chapter 4 and 5 also allow to investigate what might influence a participants exercising time in general, this means without having a robot partner or instructor. In the first block every participant is individually exercising and I collected their initial self-efficacy beliefs about their exercising abilites and information about their

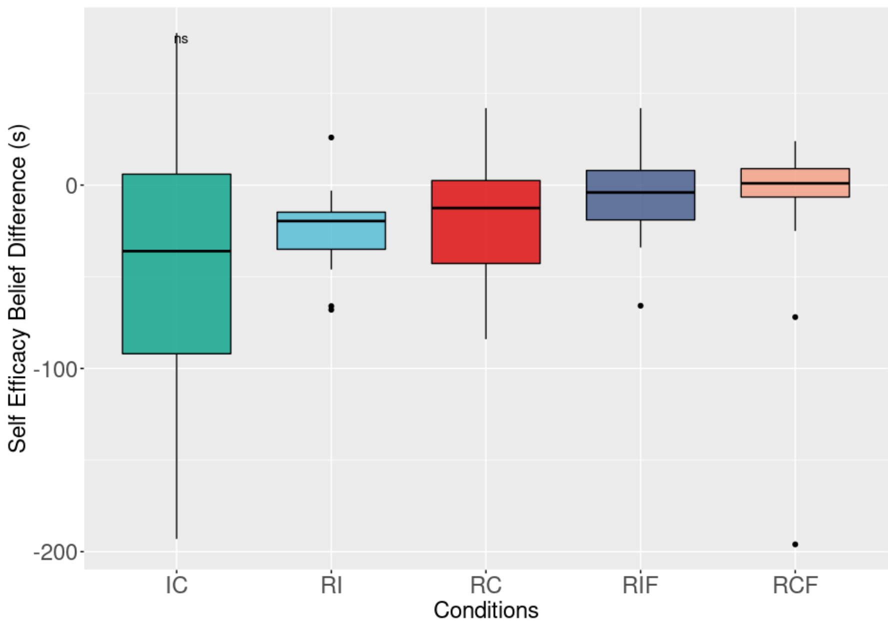


Figure C.1.9: Box plot showing the difference betweeen before and after average self-efficacy beliefs from the studies of Chapter 4 and 5.

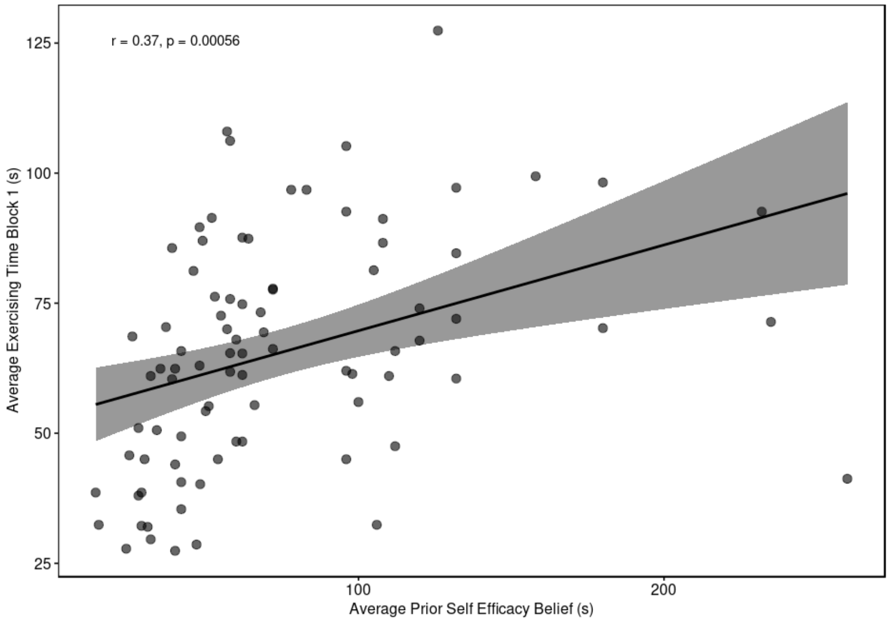


Figure C.1.10: Scatter plot of self-efficacy belief and average exercising time on Block 1.

weekly exercising time. This allows to analyse how the different exercising times on Block1 for the each each person might be explained.

Figure C.1.10 shows a scatter plot between the self-efficacy belief, which is before the exercises, and the average persistence time in Block 1. As the graph shows, there is a correlation between the initial effects and the persistence (even though the residual errors are quite high). To further analyze the data, I evaluated whether the persistence on Block 1 depends on the participant's average weekly exercising time. For this analysis, the average exercising times have been cluster into three groups. Participants stating that they exercise between 0h to 2h have been assigned the group 'low', participants stating they exercise between 3h to 5h have been assigned to the group "medium" and everything above has been assigned to the group "high". The results of this grouping is depicted in Figure C.1.11.

Considering these groups for the regression analysis, one can see that only the participants in the *medium* group correctly estimate their exercising capabilities. Individuals from the *low* group overestimate their capabilities and individuals from the *high* group underestimate their capabilities. However, these results should be interpreted with cautious due to the limited amount of individuals in the *high* group. Kruskal-Wallis test on exercising time on Block 1 showed no significant main effects, $H(2) = 3.57$, $p = .17$. Also, the self-efficacy belief before the experiment were not affected by the grouped exercising time per week, $F_{2,79} = 2.34$, $p = .1$.

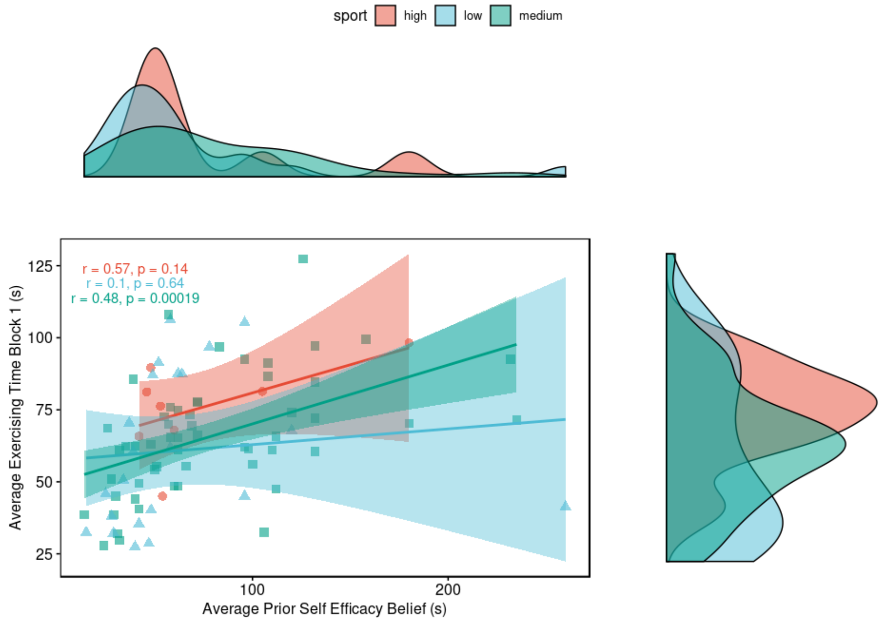


Figure C.1.11: self-efficacy belief and persistence on Block 1 with highlighting of the different exercising times per week. Exercising time per week was grouped as follows, less than two hours is labeled as `low`, between two hour and six hours per week as `medium` and more than six hour per week as `high`.

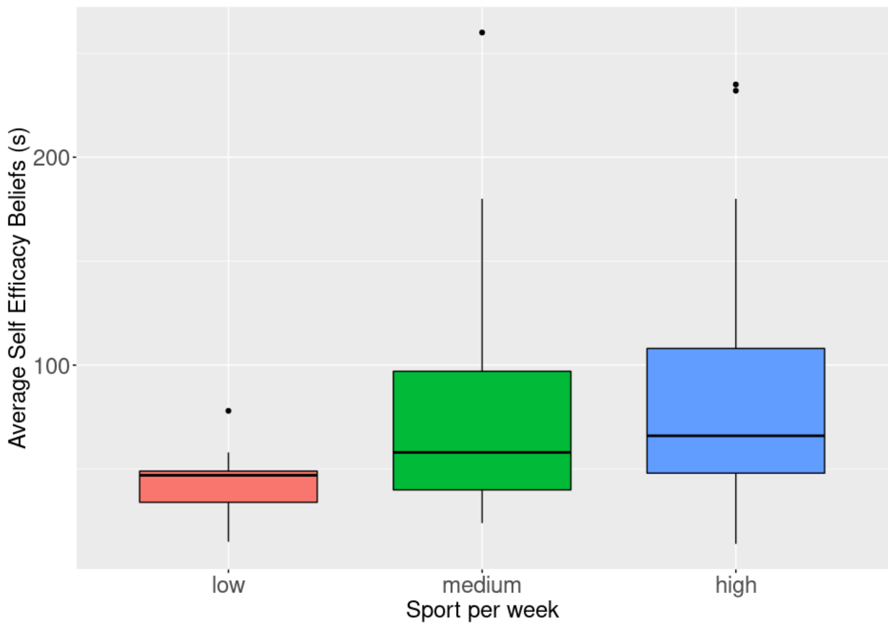


Figure C.1.12: Box plot showing the different self-efficacy belief for the different categories sport per week. Exercising time per week was grouped as follows, less than two hours is labeled as low, between two hour and six hours per week as medium and more than six hour per week as high.

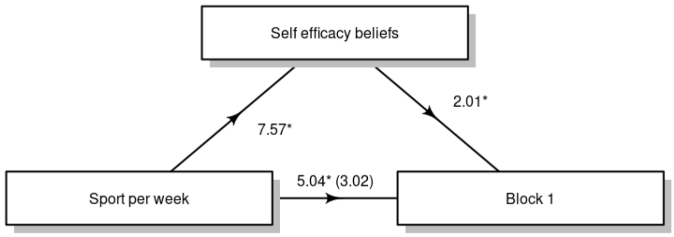


Figure C.1.13: Standardized regression coefficients for the relationship between the weekly exercising time and user's exercising time in Block 1 as mediated by the user's self-efficacy belief. The standardized regression coefficient between sport per week and the exercising time in Block 1, controlling for self-efficacy belief, is in parentheses.

To assess whether the exercising per week's effect on overall exercising time in Block 1 was statistically mediated by the self-efficacy, I used non-parametric bootstrapping method based on the method from Preacher et al. [Pre08]. Using this approach I could not find a statistically significant ACME. However, when I cleaned the data set from outliers (Figure C.1.12 shows a boxplot with the outliers. I used 200 as a cut-off for the outliers), the analysis confirmed that self-efficacy belief statistically mediated the relationship between exercising time per week and exercising time in Block 1 (ACME = 2.01, $p < .05$, 95% CI = .34 to 4.31; see Figure 30; 10,000 resamples) with no direct effect of autonomy of the system (ADE = 3.02, $p = .12$) and significant total effect ($p < .05$).

However, if I use self-efficacy belief as a moderator for exercising time in Block 2 and the conditions as treatment, self-efficacy belief have no significant influence anymore.

Additionally, when using the self-efficacy beliefs after the experiment, the exercising times on Block 2 did not significantly differ due to the self-efficacy beliefs ($F_{2,83} = .45$, $p = .5$), and the conditions did not significantly change the self-efficacy beliefs, $F_{4,80} = 1.45$, $p = .22$.

C.2 EMBODIMENT

Table C.2: Multiple comparison test after Kruskal-Wallis for persistence. This table only shows significant differences.

comparison	observed difference	critical difference
HHP-HP	51.6	48.96
HHP-RC	1.56	57.57
HHP-RC	61.5	57.57
HHP-RCF	67.05	56.54
HP-IC	78.35	48.96
HP-IC_2	74.91	57.57
IC-RC	88.31	57.57
IC-RCF	93.80	56.54
IC-RIF	84.79	58.71
IC_2-RC	84.88	65.04
IC_2-RCF	90.37	64.14
IC_2-RIF	81.36	66.05
NHP-RCF	59.93	56.54

C.3 EMBODIMENT AND PERSONALIZATION

Table C.3: Mean and standard deviation for the Godspeed questionnaire scales, System Usability Scale, intrinsic motivation and interaction satisfaction and learned preference quality of the embodiment and preference learning study.

item	condition	M	SD
Animacy	*computer*	1.6	.46
	virtual	2.41	.75
	robot	2.34	.63
Anthropomorphism	*computer*	2.1	.6
	virtual	3.02	.74
	robot	2.98	.67
Likeability	*computer*	3.1	.4
	virtual	3.94	.67
	robot	4.24	.53
Intelligence	*computer*	3.2	.56
	virtual	3.54	.81
	robot	3.41	.68
SUS	*computer*	3.2	.66
	virtual	3.9	.56
	robot	3.5	.88
Intrinsic Motivation	*computer*	3.0	.58
	virtual	3.61	1.0
	robot	3.67	.87
Interaction	*computer*	3.8	.7
	virtual	3.78	5.6
	robot	3.98	.55
learned preference quality	*computer*	3.3	1.0
	virtual	3.7	.87
	robot	3.9	.75

Table C.4: Results from post-hoc analysis using pairwise t-test with pooled SD and Bonferroni correction for the Godspeed subscales.

Godspeed item	conditions	p - value	effect size d
Animacy	*real* vs. *computer*	$< .01$	1.34
	real vs. *virtual*	n.s.	-.1
	virtual vs. *computer*	$< .01$	-1.33
Anthropomorphism	*robot* vs. *computer*	$< .01$	0.2
	real vs. *virtual*	n.s.	-.06
	virtual vs. *computer*	$< .001$	1.37
Intelligence	*real* vs. *computer*	n.s.	.34
	real vs. *virtual*	n.s.	-0.17
	virtual vs. *computer*	n.s.	-.49
Likeability	*real* vs. *computer*	$< .0001$	2.35
	real vs. *virtual*	n.s.	.49
	virtual vs. *computer*	$< .001$	-1.52

Table C.5: Multiple comparison test after Kruskal-Wallis for preference ranking error. This table only shows only significant observed and critical differences.

metric	comparison	observed difference	critical difference
position error	random-computer	31.77	18.96
	random-robot	32.71	18.96
	random-virtual	38.40	19.30
discounted error	RAND-C	32.76	18.96
	random-robot	40.04	18.96
	random-virtual	40.16	19.30

D QUESTIONNAIRES

The following sections present the used scales in this thesis in their German version.

D.1 GODSPEED QUESTIONNAIRE

Bitte bewerten Sie das System auf der folgenden Skala:

*

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

	1	2	3	4	5	
maschinenhaft	○	○	○	○	○	menschenähnlich
hat kein Bewusstsein	○	○	○	○	○	hat ein Bewusstsein
künstlich	○	○	○	○	○	realistisch
bewegt sich steif	○	○	○	○	○	bewegt sich flüssig
unecht	○	○	○	○	○	natürlich
tot	○	○	○	○	○	lebendig
unbewegt	○	○	○	○	○	lebhaft
mechanisch	○	○	○	○	○	organisch
träge	○	○	○	○	○	interaktiv
teilnahmslos	○	○	○	○	○	ansprechbar
unsympathisch	○	○	○	○	○	sympathisch
unfreundlich	○	○	○	○	○	freundlich
unhöflich	○	○	○	○	○	höflich
unangenehm	○	○	○	○	○	angenehm
furchtbar	○	○	○	○	○	nett
inkompetent	○	○	○	○	○	kompetent
ungebildet	○	○	○	○	○	gebildet
verantwortungslos	○	○	○	○	○	verantwortungsbewusst
unintelligent	○	○	○	○	○	intelligent
unvernünftig	○	○	○	○	○	vernünftig
ängstlich	○	○	○	○	○	entspannt
aufgewühlt	○	○	○	○	○	ruhig
still	○	○	○	○	○	überrascht
demotivierend	○	○	○	○	○	motivierend

D.2 ROBOT SOCIAL ATTRIBUTE SCALE

[]Wie nah sind die untenstehenden Wörter mit dem Roboter assoziiert? *

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

	(definitiv nicht mit assoziiert) 1	2	3	4	5	6	7	8	(definitiv mit assoziiert) 9
fröhlich	○	○	○	○	○	○	○	○	○
fühlend	○	○	○	○	○	○	○	○	○
sozial	○	○	○	○	○	○	○	○	○
biologisch	○	○	○	○	○	○	○	○	○
mitfühlend	○	○	○	○	○	○	○	○	○
emotional	○	○	○	○	○	○	○	○	○
kompetent	○	○	○	○	○	○	○	○	○
responsiv	○	○	○	○	○	○	○	○	○
interaktiv	○	○	○	○	○	○	○	○	○
verlässlich	○	○	○	○	○	○	○	○	○
klug	○	○	○	○	○	○	○	○	○
gruselig	○	○	○	○	○	○	○	○	○
befremdlich	○	○	○	○	○	○	○	○	○
unbehaglich	○	○	○	○	○	○	○	○	○
gefährlich	○	○	○	○	○	○	○	○	○
scheußlich	○	○	○	○	○	○	○	○	○
aggressiv	○	○	○	○	○	○	○	○	○

D.3 PHYSICAL ACTIVITY ENJOYMENT SCALE

[]Bitte geben Sie an wie Sie sich bei den Übungen, die Sie gerade durch geführt haben, gefühlt haben. *

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

	1	2	3	4	5	
Ich genieße es	○	○	○	○	○	Ich hasses es
Ich fühle micht gelangweilt	○	○	○	○	○	Ich fühle mich interessiert
Ich mag es nicht	○	○	○	○	○	Ich mag es
Ich finde es angenehm	○	○	○	○	○	ich finde es unangenehm
Ich bin total von der Aufgabe absorbiert	○	○	○	○	○	Ich bin überhaupt nicht von der Aufgabe absorbiert
Es ist überhaupt nicht spaßig	○	○	○	○	○	Es ist total spaßig
Ich finde es antreibend	○	○	○	○	○	Ich finde es ermüdend
Es bedrückt mich	○	○	○	○	○	Es macht mich freudig
Es ist sehr angenehm	○	○	○	○	○	Es ist sehr unangenehm
Ich habe mich währendedessen physisch gut gefühlt	○	○	○	○	○	Ich habe mich währenddessen physisch schlecht gefühlt
Es ist sehr stärkend	○	○	○	○	○	Es ist überhaupt nicht stärkend
Ich bin davon frustriert	○	○	○	○	○	Ich bin davon überhaupt nicht frustriert
Es ist sehr belohnend	○	○	○	○	○	Es ist überhaupt nicht belohnend
Es ist sehr erheiternd	○	○	○	○	○	Es ist überhaupt nicht erheiternd
Es ist überhaupt nicht stimulierend	○	○	○	○	○	Es ist sehr stimulierend
Es ist sehr wohltuend	○	○	○	○	○	Es ist überhaupt nicht wohltuend
Es fühlte sich so an als würde ich etwas ganz anderes machen	○	○	○	○	○	Es fühlte sich überhaupt nicht so an als würde ich etwas ganz anderes machen

D.4 SYSTEM USABILITY SCALE

Stellen Sie sich vor Sie würden in Zukunft so ein System verwenden, welches Sie dabei unterstützen soll regelmäßig sportlich aktiv zu sein in dem es ihre sportlichen Vorlieben lernt.

Wie würden Sie die folgenden Fragen hinsichtlich der Systemfunktionalität, ihre Vorlieben zu lernen, beantworten?
*****

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

	(Ich stimme voll zu) 1	2	3	4	(Ich stimme gar nicht zu) 5
Ich kann mir sehr gut vorstellen, das System regelmäßig zu nutzen.	○	○	○	○	○
Ich empfinde das System als unnötig komplex.	○	○	○	○	○
Ich empfinde das System als einfach zu nutzen.	○	○	○	○	○
Ich finde, dass die Funktionen des Systems gut integriert sind.	○	○	○	○	○
Ich finde, dass es im System zu viele Inkonsistenzen gibt.	○	○	○	○	○
Ich empfinde die Bedienung als sehr umständlich.	○	○	○	○	○
Ich habe mich bei der Nutzung des Systems sehr sicher gefühlt.	○	○	○	○	○
Ich kann mir vorstellen, dass die meisten Leute das System schnell mögen werden.	○	○	○	○	○
Ich habe zuerst nicht verstanden was das System macht.	○	○	○	○	○
Ich denke, dass das System ohne weiteren Support nutzlos ist.	○	○	○	○	○

D.5 NEGATIVE ATTITUDES TOWARDS ROBOTS SCALE

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

	(Ich stimme gar nicht zu)				(Ich stimme voll zu)
Ich würde mich unwohl fühlen, wenn ich auf der Arbeit mit Robotern zu tun hätte.	○	○	○	○	○
Es könnte etwas Schlimmes passieren, wenn sich Roboter zu Lebewesen entwickeln würden.	○	○	○	○	○
Ich wäre entspannt, wenn ich mit einem Roboter spräche.	○	○	○	○	○
Wenn Roboter Emotionen besäßen, könnte ich mich mit ihnen anfreunden.	○	○	○	○	○
Ich würde mich wohl fühlen, wenn Roboter tatsächlich Emotionen besäßen.	○	○	○	○	○
Das Wort "Roboter" hat keine Bedeutung für mich.	○	○	○	○	○
Ich wäre nervös, wenn ich vor anderen Leuten einen Roboter bedienen müsste.	○	○	○	○	○
Ich verabscheue die Vorstellung, dass Roboter oder künstliche Intelligenzen sich Urteile über Dinge bilden können.	○	○	○	○	○
Vor einem Roboter zu stehen, würde mich nervös machen.	○	○	○	○	○
Ich glaube, dass etwas Schlimmes passiert, wenn ich zu sehr von Robotern abhängig wäre.	○	○	○	○	○
Ich würde mich paranoid fühlen, wenn ich mit einem Roboter spräche.	○	○	○	○	○
Ich mache mir Sorgen, dass Roboter einen schlechten Einfluss auf Kinder haben könnten.	○	○	○	○	○
Ich glaube, dass die Gesellschaft in Zukunft von Robotern dominiert wird.	○	○	○	○	○
Ich fühle mich beruhigt mit Robotern die Emotionen haben.	○	○	○	○	○

D.6 TEAM PERCEPTION SCALE

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

	(Ich stimme gar nicht zu) 1	2	3	4	(Ich stimme voll zu) 5
Ich war Teil einer Gruppe	○	○	○	○	○
Der Roboter war mein Partner	○	○	○	○	○
Mir kam es so vor als würden wir gemeinsam trainieren	○	○	○	○	○
Mir kam es so vor als würden wir miteinander trainieren	○	○	○	○	○
Mir kam es nicht so vor als würden wir getrennt trainieren	○	○	○	○	○

D.7 COOPERATION SCALE

[]Bitte beantworten Sie die folgenden Items *

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

	(Ich stimme gar nicht zu) 1	2	3	4	(Ich stimme voll zu) 5
Ich kooperiere mit dem Roboter	○	○	○	○	○
Ich wollte die Zustimmung des Roboters	○	○	○	○	○
Die Hinweise des Roboters haben mir geholfen	○	○	○	○	○

D.8 WORKING ALLIANCE SCALE

[]Denken Sie an ihre vorherige Erfahrung während der Interaktion und entscheiden Sie welche der untenstehenden Kategorien ihre Erfahrung am besten beschreibt. *

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

	Weniger	Ein wenig	Ziemlich	Stark	Sehr stark
Als Ergebniss dieser Sitzung bin ich mir klarer wie ich mich verändern könnte und welche Sportübungen ich mag.					
Was ich in der Sitzung getan habe, gibt mir eine neue Betrachtung auf meine sportlichen Vorlieben.					
Ich glaube Nao mag mich.					
Nao und ich haben während der Sitzung gemeinsam auf unsere gemeinsamen Ziele hingearbeitet.					
Nao und ich respektieren uns.					
Nao und ich sind uns einig darüber was meine bevorzugte Sportübungen sind.					
Ich fühle, dass Nao mich würdigt.					
Nao und ich sind uns darüber einig was für mich wichtig ist um daran weiter zu arbeiten.					
Ich glaube das sich Nao um mich sorgt, auch wenn ich Dinge machen die es nicht gut heißt.					
Ich habe das Gefühl, dass die Sportübungen die ich in der Sitzung getan habe mir dabei helfen werden meine Bewegungssiele zuerreichen					
Nao und ich haben ein gutes Verständnis aufgebaut über die Sportübungen die gut für mich wären.					
Ich glaube, dass die Art wie wir zusammengearbeitet haben gut für mich ist.					

D.9 PERCEIVED INFORMATION QUALITY SCALE

D.9.1 *Feedback Study*

[]Bitte beantworten Sie die folgenden Items *

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

	(Ich stimme gar nicht zu) 1	2	3	4	(Ich stimme voll zu) 5
Die Informationen des Roboters waren relevant.	○	○	○	○	○
Die Informationen des Roboters waren nützlich.	○	○	○	○	○
Die Informationen des Roboters waren erkenntnisreich.	○	○	○	○	○

D.9.2 *Preference Learning Studies*

[]Bitte beantworten Sie die folgenden Items *

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

	(Ich stimme gar nicht zu) 1	2	3	4	(Ich stimme voll zu) 5
Die Information zu den Übungen waren sinnvoll.	○	○	○	○	○
Die Informationen zu den Übungen waren nützlich.	○	○	○	○	○
Die Informationen zu den Übungen waren erkenntnisreich.	○	○	○	○	○

[]Bitte beantworten Sie die folgenden Items *

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

	(Ich stimme gar nicht zu) 1	2	3	4	(Ich stimme voll zu) 5
Die verglichenen Übungen waren sinnvoll.	○	○	○	○	○
Die Übungsvergleiche des Systems waren nützlich.	○	○	○	○	○
Die Übungsvergleiche des Systems waren erkenntnisreich.	○	○	○	○	○
Die Übungsvergleiche des Systems haben genervt.	○	○	○	○	○
Die Auswahl der Übungen hatten eine Systematik.	○	○	○	○	○

D.10 NEO-FFI

[]Inwieweit treffen die folgenden Aussagen auf Sie zu? *

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

	trifft überhaupt nicht zu	trifft eher nicht zu	weder noch	trifft eher zu	trifft voll und ganz zu
Ich fühle mich anderen oft unterlegen.	○	○	○	○	○
Wenn ich unter starkem Stress stehe, fühle ich mich manchmal, als ob ich zusammenbräche.	○	○	○	○	○
Ich fühle mich oft angespannt und nervös.	○	○	○	○	○
Manchmal fühle ich mich völlig wertlos.	○	○	○	○	○
Zu häufig bin ich entmutigt und will aufgeben, wenn etwas schief geht.	○	○	○	○	○
Ich fühle mich oft hilflos und wünsche mir eine Person, die meineProbleme löst.	○	○	○	○	○
Ich habe gern viele Leute um mich herum.	○	○	○	○	○
Ich bin leicht zum Lachen zu bringen.	○	○	○	○	○
Ich bin gerne im Zentrum des Geschehens.	○	○	○	○	○
Ich habe oft das Gefühl, vor Energie überzuschäumen.	○	○	○	○	○
Ich bin ein fröhlicher, gutgelaunter Mensch.	○	○	○	○	○
Ich bin ein sehr aktiver Mensch.	○	○	○	○	○
Ich finde philosophische Diskussionen langweilig.	○	○	○	○	○
Mich begeistern die Motive, die ich in der Kunst und in der Natur finde.	○	○	○	○	○
Poesie beeindruckt mich wenig oder gar nicht.	○	○	○	○	○
Wenn ich Literatur lese oder ein Kunstwerk betrachte, empfinde ich manchmal ein Frösteln oder eine Welle der Begeisterung.	○	○	○	○	○

	trifft überhaupt nicht zu	trifft eher nicht zu	weder noch	trifft eher zu	trifft voll und ganz zu
Ich habe wenig Interesse, über die Natur des Universums oder die Lage der Menschheit zu spekulieren.	◯	◯	◯	◯	◯
Ich habe oft Spaß daran, mit Theorien oder abstrakten Ideen zu spielen.	◯	◯	◯	◯	◯
Ich bekomme häufiger Streit mit meiner Familie und meinen Kollegen.	◯	◯	◯	◯	◯
Manche Leute halten mich für selbstsüchtig und selbstgefällig.	◯	◯	◯	◯	◯
Im Hinblick auf die Absichten anderer bin ich eher zynisch und skeptisch.	◯	◯	◯	◯	◯
Manche Leute halten mich für kalt und berechnend.	◯	◯	◯	◯	◯
Ich versuche stets rücksichtsvoll und sensibel zu handeln.	◯	◯	◯	◯	◯
Um zu bekommen, was ich will, bin ich notfalls bereit, Menschen zu manipulieren.	◯	◯	◯	◯	◯
Ich halte meine Sachen ordentlich und sauber.	◯	◯	◯	◯	◯
Ich kann mir meine Zeit recht gut einteilen, sodass ich meine An-gelegenheiten rechtzeitig beende.	◯	◯	◯	◯	◯
Ich versuche, alle mir übertragenen Aufgaben sehr gewissenhaft zu erledigen.	◯	◯	◯	◯	◯
Wenn ich eine Verpflichtung eingehe, so kann man sich auf mich bestimmt verlassen.	◯	◯	◯	◯	◯
Ich bin eine tüchtige Person, die ihre Arbeit immer erledigt.	◯	◯	◯	◯	◯
Ich werde wohl niemals fähig sein, Ordnung in mein Leben zu bringen.	◯	◯	◯	◯	◯

D.11 OPENNESS TO INFLUENCE SCALE

D.11.1 *Feedback Study*

[]Bitte beantworten Sie die folgenden Items *

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

	(Ich stimme gar nicht zu) 1	2	3	4	(Ich stimme voll zu) 5
Ich bin offen für die Vorschläge des Roboters	○	○	○	○	○
Ich nehme die Hinweise des Roboters ernst	○	○	○	○	○
Meine Leistung hängt von den Hinweise des Roboters ab	○	○	○	○	○
Ich akzeptiere die Ratschläge des Roboters	○	○	○	○	○
Ich stimme dem Roboter zu	○	○	○	○	○
Ich bin mag die Anregungen des Roboters	○	○	○	○	○
Ich vertraue den Informationen des Roboters	○	○	○	○	○

D.11.2 *Preference Learning Studies*

[]Stellen Sie sich vor Sie würden in Zukunft dieses System nutzen und es würde Ihnen Übungen vorschlagen die sie dann ausprobieren sollen damit das System ihre Vorlieben lernt, würden Sie *

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

	(Ich stimme gar nicht zu) 1	2	3	4	(Ich stimme voll zu) 5
offen sein für die Übungsvorschläge des Systems	○	○	○	○	○
die Übungsvorschläge ernst nehmen	○	○	○	○	○
die vogeschlagenen Übungen akzeptieren	○	○	○	○	○
den vogeschlagenen Übungen zu stimmen	○	○	○	○	○
die Anregungen des Systems mögen	○	○	○	○	○
den Informationen des Systems vertrauen	○	○	○	○	○

D.12 INTRINSIC MOTIVATION

Bitte bewerten Sie die Interaktion mit dem System. Geben Sie für jede Aussage an, inwiefern diese für Sie zutrifft.
*

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

	(Ich stimme gar nicht zu) 1	2	3	4	(Ich stimme voll zu) 5
Die Interaktion mit dem System hat mir Spaß gemacht.	○	○	○	○	○
Bei der Interaktion mit dem System habe ich mich angespannt gefühlt.	○	○	○	○	○
Ich fand die Interaktion mit dem System sehr interessant.	○	○	○	○	○
Die Interaktion mit dem System war unterhaltsam.	○	○	○	○	○
Ich fühlte ich mich gestresst während der Interaktion	○	○	○	○	○
Ich hatte Bedenken, ob ich die Interaktion mit dem System gut hinbekomme.	○	○	○	○	○
Die Interaktion mit dem System war unterhaltsam.	○	○	○	○	○

D.13 INTERACTION SCALE

[]Die Interaktion mit dem System *

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

	1	2	3	4	5	
fiel mir schwer	○	○	○	○	○	fiel mir leicht
war ineffizient	○	○	○	○	○	war effizient
war verwirrend	○	○	○	○	○	war selbsterklärend
war frustrierend	○	○	○	○	○	war angenehm
war kompliziert	○	○	○	○	○	war leicht zu verstehen
hat Spaß gemacht	○	○	○	○	○	hat keinen Spaß gemacht

DECLARATION OF AUTHORSHIP

According to the Bielefeld University's doctoral degree regulations §8(1)g: I hereby declare to acknowledge the current doctoral degree regulations of the Faculty of Technology at Bielefeld University. Furthermore, I certify that this thesis has been composed by me and is based on my own work, unless stated otherwise. Third parties have neither directly nor indirectly received any monetary advantages in relation to mediation advises or activities regarding the content of this thesis. Also, no other person's work has been used without due acknowledgment. All references and verbatim extracts have been quoted, and all sources of information, including graphs and data sets, have been specifically acknowledged. This thesis or parts of it have neither been submitted for any other degree at this university nor elsewhere.

Sebastian Schneider	Place, Date

COLOPHON

This document was typeset using the typographical look-and-feel `classicthesis` developed by André Miede and Ivo Pletikosić. The style was inspired by Robert Bringhurst's seminal book on typography "*The Elements of Typographic Style*". The analysis style has been influenced by Field et al. [Fie12]. Plots and figures have been generated using [R C13] The writing style has been influenced by Turabian [Tur18].