Integration and Evaluation of a Gaming Situation for Long-Term Human-Robot Interaction

Playing a Game of Pairs with Flobi using Contextual Knowledge

Andreas Kipp

Declaration of Authorship

According to 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.

Andreas Kipp	Place, Date

Integration and Evaluation of a Gaming Situation for Long-Term Human-Robot Interaction

Playing a Game of Pairs with Flobi using Contextual Knowledge

Andreas Kipp

March 2015

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

Faculty of Technology
Bielefeld University
Applied Informatics
Inspiration 1
33619 Bielefeld
Germany

Reviewers

Prof. Dr.-Ing. Franz Kummert
Dr.-Ing. Marc Hanheide

Examination Board

Prof. Dr. Philipp Cimiano
Dr. rer. nat. Thies Pfeiffer

Defended and approved at Tuesday 11^{th} August, 2015

Printed on permanent paper as per ISO 9706.

Acknowledgments

There are many people I would like to thank, making it possible to complete another step in my life.

First of all, I would like to thank my father, accepting and supporting all the decisions I made in my life. Thank your for creating the foundation of all this.

Thank you dad!

Then I would like to thank my girlfriend, my relatives and closest friends, that gave me advices and feedback throughout all the time it took me to complete this task. Thank you for your understanding in times I was stressed and edgy.

I am grateful for the support of Franz and Britta, for the possibility to work in the applied informatics group and on the topic of social robotics. Thanks to Ingo, Simon and Frank for their support with our robot Flobi and the difficulties that had to be solved now and then. I also like to thank all my colleagues for their feedback and the many discussions, mostly near our coffee machine. Especially I would like to thank Patrick and Torben, for the many detailed discussions on my topic, their feedback and helpful ideas.

Thank you Franz, for being my supervisor and for being there, whenever I had questions (and I asked a lot), for all the helpful input upon all stages I went through. Furthermore, I would like to thank Marc, for accepting to review my thesis.

Thank you all!

Abstract

This doctoral thesis proposes a system for long-term interaction between a robot and a human using a gaming context. A robot plays a game of pairs autonomously with a human. The system was developed to evaluate how to implement social assistive robots during space missions that occur under isolation conditions.

The first part of the thesis presents the system as designed and implemented. Described are the different components used to realize autonomous interaction. The study itself was conducted in cooperation with the German Aerospace Center. Throughout the study, the proposed system performed robustly, and without major system failures. The participants interacted with the system continuously, and gave it average ratings for acceptability. No significant extraneous effects, such as those related to novelty were found. Nevertheless, problems with perception and classification led to negative ratings for the system's competence.

The second part of the thesis was motivated by findings from the isolation study. Integrated is a context knowledge system to increase interest in the interactions during long-term use. This made it possible to collect data on past interactions for use with further interactions.

The results of a study showed that greater commitment to gaming interaction could be promoted by using context knowledge. In addition, implemented is a remotely controlled version of the system to evaluate, whether a robot without visual or speech recognition problems who played perfectly could influence the way the system was perceived. Results showed that ratings on participant enjoyment while playing with the system decreased when the playing system was too perfect. Additionally, a robot that played too perfectly appeared to promote unfair game-play behaviors in the human, likely as an attempt to cope with disadvantages such as lower memory capacity.

Contents

List of Tables

List of Figures

1. Introduction

Maintaining social contacts has become an increasing challenge for many people in modern society. People with disabilities or elderly persons often stay in one place for long periods of time. Relatives or caregivers may not always be available to occupy or entertain people. The resultant social isolation has been shown to lead to negative health effects [3].

Other groups, for example astronauts, are unable to simply “go outside” to meet with people. Although there are ways to communicate with support staff, in most cases communication is impersonal. In addition there is often a delay between communicative exchanges. Staying in the same place without personal contact with others can lead to boredom or depression. These psychological effects need to be identified and researched [18, chap.5], so that effective counter measures can be developed to support these groups in their daily lives [7].

Social robotics has developed several approaches to help bring robots into peoples homes as entertainment and social partners. Robots can help satisfy communication needs, as well as accompany people during their daily routines. Their presence can help to alleviate tedium and loneliness. The use of robots in these context raises questions about just how robots can support humans. How do the people they are intended to support, respond to them? Could the role of a robot become that of a social companion? What might the effects be in the long term for the people interacting with these social robots as a normal part of their daily lives?

1.1. The SoziRob Project - Robots as Social Companions

In 2010, Bielefeld University started the SoziRob project[1] in cooperation with the German Aerospace Center[2]. The goal of the project was to investigate how supportive social robotic systems could assist and accompany humans that live and work under isolation conditions [2]. The idea was to use robots as supporters and social partners in daily routines that simulate tasks performed by astronauts in space missions. One type of interaction to be implemented and evaluated was social human-robot interaction. Due

[1]https://aiweb.techfak.uni-bielefeld.de/projekt-sozirob
[2]http://www.dlr.de/

to the number of tasks, the goal for the interaction was to be entertaining so as to reduce stress and loneliness.

This one prerequisite produced requirements defined in the following terms:

Requirement 1:
Create a social and entertaining human-robot interaction to reduce isolation effects.

The interactions should also be cognitively challenging. They should help participants improve memory recall, which leads to the second requirement.

Requirement 2:
The entertaining interaction should provide a cognitive benefit.

Selected was a gaming situation in which a robot plays against a human counterpart using the game “pairs”. It offers the possibility of training the participants memory and improving concentration skills. The goal of the game is to find and remember pairs of matching cards in a set of hidden cards. The simple structure of the game can be communicated to the player using a dialog system. The robot can specify cards and ask the human player to turn them. Cards placed in front of the robot can be detected and classified using the robot’s vision elements.

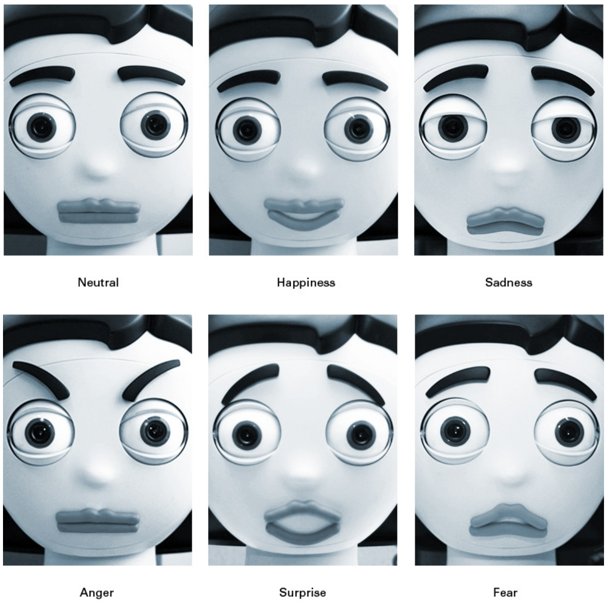
The robotic platform used for the interaction was the anthropomorphic robot head Flobi (see fig. 1.1) developed at Bielefeld University [32]. The robot has a human-like facial appearance, and is mounted on a small tub torso. Flobi uses 16 degrees of freedom and can rotate its head (pan, tilt and roll) to focus on points of interest. Additionally, it can move its eyes (up / down and left / right) allowing it to directly focus on points in its field of view. It can move its eyelids and eyebrows as well as its mouth (left + right up/down and center upper and lower up/down each). By combining these elements, Flobi can display basic emotions like anger, disgust, fear, happiness, sadness and surprise. The lip movement is used whenever Flobi uses speech production. The robot has two cameras placed inside the eyeballs. Both cameras can be used for image processing (tracking faces, classification of items, etc.). Because the robot possesses no manipulators to move objects, the interaction needed to be managed using only verbal communication.

A third requirement for the interaction and the robot was to create an interaction that could be used in repeated interactions over a period of 18 days. The isolation conditions required that needed maintenance be reduced. Problems should be managed remotely as much as possible.

Requirement 3:
Create a robust interaction that could be used over longer periods.

The implementation of the system, and the evaluation during the isolation study will be described later in this thesis (see chapter 3).

(a) The robot head Flobi



(b) Basic emotions shown by Flobi

Figure 1.1.: The robot head Flobi (a). Flobi is capable of expressing different emotions (b) using 16 manipulators. The robot is equipped with two cameras placed inside the eyeballs.

1.2. Long-Term Motivation: Keeping the Game Interesting

For most people interacting with a robot is exciting and interesting. This can be observed from the first interactions. A robot's ability to act and to react in a human-like manner can make a person curious about what the robot may be capable of. But if people repeat the interaction and get used to the robots behaviors, they tend to establish static behavior patterns. If the interaction remains static and nothing new occurs, then they may get bored, and the novelty effect wears off (see [12, 19, 6, 45]). Consequently, most people tend to minimize or even avoid further interactions.

This thesis describes a human-robot interaction about playing a game. The goal of the system was to allow autonomous game-play. To investigate long-term effects, the system was designed for repeated interactions.

There are several approaches to applying a memory component to enrich interactions over time (see [14, 21]). Most constructs require complex implementations. Data must be acquired over longer periods to train the system and generate appropriate behaviors.

A simpler memory approach is used for the system described in this thesis. By counting different events occurring in every game, a context for each interaction partner can be built. This context can then be used in later interaction by integrating results into the robots dialog.

During gaming situations, the robot can store and analyze events while the game-play remains ongoing. For instance, the robot can count how often interactions occur, or how many turns are made in a round. Based on these information, the dialog system can select a corresponding output. Also, the data itself can be used to output dialog. Listing 1.1 shows what a sample dialog with contextual information may look like.

Listing 1.1: Example of how the robot can record and use contextual information for dialog with a human.

```
<Query context: Player has started more often>
Robot: "You have started more often,
        may I start this time?"
Human: "Ok, you start this time."
```

This raised the question of whether a robot that knows how a person performs while playing, could use the data to motivate the player to play again. To evaluate the effects of such a contextual system the following hypothesis were formulated:

H1 Collecting simple data based on counted events and utilize it in the dialog system will result in positive ratings of the system after repeated interactions.

H2 By altering the dialog based on the collected contextual information, the user will play more games even in a later interaction.

The system described in this thesis works autonomously. Flaws are expected due to the complex vision pipeline and unforeseen participant behaviors. These errors must be handled through dialog to keep the interaction going.

In an interaction that is not too perfect and that allows errors to occur, humans tend to anthropomorphize the robot instead of rating it as a machine [37].

In this thesis the effect of a system with and without interaction problems will be evaluated. To this end, a remotely controlled version will be tested against the autonomous system. To evaluate effects how the autonomous system is rated against a system that interacts nearly perfectly, the following hypothesis has been formulated:

H3 A system with perfect game-play will result in less commitment for later interactions and will be perceived more like a tool.

To test these hypotheses, the interaction system will be extended to include a context system for recurring games (see chapter 4). The effects of a system that plays perfectly using a remote control component will be investigated. The effects of the context system will be evaluated in a long-term study.

The context system will be evaluated during the first part of the study. It will be compared to a system without knowledge about past interactions. This second system will skip context in the dialog part. In both cases, the system will play autonomously against human opponents.

In the second part of the study, the autonomous system will be evaluated against a remote controlled system. The goal of the remote control system is to play perfectly, and to make no mistakes. To this end, the visual component and voice recognition will be controlled by a human examiner. The player will not know about the control mechanism until the end of the study.

1.3. Thesis Prospect

The next chapter (2) will provide an introduction to the field of human-robot interaction in terms of long-term interactions, interactions within the scope of gaming situations, and approaches to using memory from past events. Chapter 3 describes in detail the system implemented for the SoziRob project. It covers all components implemented for human-robot interaction. The results of the isolation study will be presented and discussed here. Chapter 4 provides information about the context system integrated into the system described in chapter 3, as well as provide description of the component used to remotely control the system. This chapter will describe the evaluation of these components and discuss their results. Chapter 5 will provide a conclusion to the results of using the systems implemented and the studies conducted. The final chapter (see chapter 6) will indicate directions for further work on the system and its usage in the field of long-term interaction.

2. Related Work

A great deal of work has been done on single interaction in the field of human-robot interaction (HRI). Long-term interaction, on the other hand has only recently begun to be heavily investigated motivated by the difficulties that arise in its implementation. In contrast to a single interaction study, a study with reoccurring interactions is time consuming and needs more structured planning. This often leads to much more overhead [11]. Measuring data over longer periods requires careful consideration of the methods used, as well as how to store and analyze the data. In addition, interpretation can be more difficult. The nature of changing perceptions in the interaction gives rise to multiple extraneous effects. One such effect is the novelty effect. It describes situations where humans tend to rate new experiences more positively early on, and less positively once the novelty has worn off. In this case, human participants react very positively when first interacting with the robot, only to later become bored if its behavior becomes predictable. Another effect is familiarization or habituation [26]. Once humans know how the interaction needs to be handled, they tend to create static behaviors for the interaction. Humans tend to utilize these behaviors for all upcoming interaction when the system proceeds in a static manner.

These considerations lead to the following questions: How do humans perceive a robot over repeated interactions? What role does the robot play? Can the robot become a companion?

Currently, robots are finding their way more and more into normal every day life. Robots can be found helping to clean a house [10], or as toys reacting to a child's behavior [17]. As such, robots needs to be able to in interact repeatedly with humans. For industry, it is of interest how robots should behave during long-term interactions.

For the scope of this thesis three topics are relevant to long-term interaction: long-term interaction in social robotics, representation of memory for social robotics and interactions for social robotics in gaming situations.

The field of long-term interaction deals with human-robot interaction over longer periods of time. A single human or a group of humans interact with one or more robots. Studies can help to investigate how humans behave when exposed to robots repeatedly. Section 2.1 focuses on publications and studies conducted in the context of long-term interactions.

Using memory for human-robot interaction allows the system to remember events and to alter upcoming sessions based on them. Section 2.2 describes different approaches to how events and data from interactions can be handled and stored, how to analyze them and how to make them usable in later interactions.

In section 2.3, the use of robots as companions or opponents in interactions where a human plays with or against a robot is considered. What effects may emerge when a human plays with a robot? Do humans like to play with robots? Studies investigating these situations will be discussed.

2.1. Long-Term Human-Robot Interaction

A good starting point for work on long-term interaction is the survey by Leite et al. [28]. The authors looked at different research fields where long-term interacting systems can be applied. Examples include home environments and health care facilities. Several studies described situations where a robot and a human interact in repeated sessions. One limitation of the survey is that it only covers studies in real-world environments. Studies conducted in laboratories are left out. That said, the authors created categories for health care and therapy, education, work environments and public spaces, and home environments. In most cases, the studies tested a small part of a long-term interaction. There were mostly situations in which a human needed to fulfill a simple task in cooperation with a robot. They pointed out the many details that need to be considered when running a long-term study: how long should it be? What is the tested target group and will it represent the whole population? Even differences in age can lead to broadly divergent results.

One prominent example mentioned in the survey is the robotic seal Paro (see fig. 2.1). Paro is a robot designed to look like a toy seal. The robot uses tactile sensors and can move its eyelids, neck and paddles. Using acoustic and pressure sensors, the robot reacts to being cared for. In 2006 Wada and Shibata [42] presented the first results of a two month study. In the study, 12 elderly people interacted with Paro over a five week period for about nine hours a day. The analysis was based on questionnaires, interviews and urinary tests. The results showed that most participants observed positive changes in themselves. The robot encouraged them to interact more often with the robot itself and with the other participants. The urinary tests showed that the participant's vital organs' reaction to stress improved. This suggests that using robots in health care scenarios can lead to psychological and physiological benefits when used over long periods.

Kanda et al. [19] conducted a long-term study in a Japanese elementary school. Two robots of the type Robovie (see fig. 2.2a) performed over 18 days with first and six graders. The task of both robots was to engage the children to speak English and to support them in their pronunciation. The children could play and communicate with the robot. The robot was capable of identifying its interacting partner using a wireless ID tag system. The main research questions concerned whether a robot could establish a relationship with a child. The authors wanted to investigate whether the robot was perceived as a partner or more like a machine. The results showed that many details must be considered before conducting a field study over several sessions. The children started to loose interest in the robot after one week, and some started to mob the robot due to its lack of abilities. The authors mentioned that the robot's abilities needed to be more robust to allow to respond to difficult situations. Interactions in public places cannot be controlled as in a laboratory environment. There was also an effect related to

Figure 2.1.: The robotic seal paro designed by the Intelligent System Research Institute of Japan's AIST. The robot is used in long-term interaction with elderly people in heath care facilities.

learning English, but the results were limited in there explanatory power given that no comparison was made to other teaching methodologies such as computer-aided learning. Nevertheless, the results indicate that although long-term interaction is challenging, it is a well worth the effort.

Kidd and Breazeal [22] performed a long-term interaction study where a robot acts as a weight loss coach. The robot Autom (see fig. 2.2b) interacted two times a day for five minutes. The robot logged information about how humans performed on their diet plans. The author's goal was to investigate how a robot should perform in a long-term study, and evaluate its as a weight loss coach. The study was conducted with 45 participants. Fifteen participants worked with the robot, the other 30 where distributed among a software program and a paper log. The results showed that participants enjoyed the interaction and developed a closer alliance with the robot compared to the computer or the paper system. This suggests that embodiment and interaction can lead to better performance.

One publication mentioned in the survey was of particular interest to the system described in this thesis. In 2014 Leite et al. [27] published results of a long-term interaction

(a) Robovie-2

(b) Autom ©2014 Intuitive Automata

Figure 2.2.: The robot Robovie (a) was used as teacher for the English language in a Japanese elementary school. The robot Autom (b) acts as a weight loss coach.

between a robot and children. The robot iCat (see fig. 2.3) was used in a school as a companion playing games over several weeks. The game of choice was chess and the robot played the counterpart to one child at a time. The study was conducted at a school with 16 third grade children. Over five consecutive weeks, each child played a game of chess with the iCat. Each game started with a predefined chess situation chosen by the teacher. To challenge the child, the difficulty of each game varied between the sessions. One session took from 10 to 25 minutes. After the session in week one, and the last session in week five, a questionnaire was completed by the child, after which the child was interviewed.

Several factors were analyzed: social presence, engagement, help and self-validation, and perceived social support. Social presence was equivalent between weeks one and five, indicating no effect of decrease over the long term. The same effect was found for engagement and help and self-validation. The authors concluded that in the given setting, the children saw the robot as a supportive companion. The authors noted that a limitation of the study was that the children did not understand the questionnaires as well as the adults. Also for the interviews, children sought to please the interviewer in following with the effect of suggestibility[8].

Conclusion to Long-Term Human-Robot Interaction

Long-term interactions and studies are essential to understanding how humans and robots can possibly work together or next to each other. To cope with effects like novelty, the system must be capable of change and to adapt its behaviors. For studies on long-term interaction, many considerations need to be taken into account. Topics such as how to store information and how to analyze and utilize the data need to be specified. It is crucial to consider how to collect and analyze data of impressions and experiences

Figure 2.3.: The robot iCat. In the scenario the robot plays the game chess with participants. The robot advises the player on the turns to be played.

from participants. Direct techniques like questionnaires can lead to subjective results that could possibly impact analyses. Implicit estimations could lead to more objective results, although they might lack information due to the complex structure of the interaction. Nevertheless, once these issues are addressed, the results can be very valuable. Long-term interaction can lead to much important information on how to design social robots intended for long-term use.

2.2. Memory for HRI: Remembering Events and Interactions

Work by Leite et al. [28] raised the question of the importance of memory for HRI. One important element for long-term interaction is to recognize the information provided and events as they occur, and to store this information for later use. Doing so can lead to a more positive perception of the robot and the interaction.

When thinking about how to remember past events in repeated interactions, the model proposed by Endel Tulving [40, 41] for *episodic memory* comes to mind. He stated that *"Episodic memory receives and stores information about temporally dated episodes or events, and temporal-spatial relations among these events."* [41].

Episodic memory has been studied extensively in intelligent virtual agents (IVA). Given that they are less complex than robotic systems, memory systems can be implemented and tested more quickly.

Lim et al. [29] worked on socially-aware memory for companion agents. The authors extended a computational model to emotional agents, based on a system for short- and long-term memory, that also allowed them to forget past events. The prototype was implemented and tested with the conversational embodied agent Greta. Although no results were presented, the system showed how a simple memory model could be implemented in social interactive companions.

Brom et al. carried out a similar approach on an IVA [4] . The authors' goal was to create episodic memory for generic and believable agents found in video games scenarios. Episodic memory was implemented as life-long autobiographic memory that allowed the agent to plan its reactions based on actions that occurred in the past. Similarly to Lim's work, the system was only tested in scenarios running for several days, and no study was presented. Nevertheless, the authors pointed out that episodic memory could be represented in IVAs. They also proposed that IVA's are better suited to study these effects as compared to robotic platforms due to their greater accessibility.

Memory representation is also of interest for robotic systems. Providing memory capabilities can help to support natural and social interactions.

In 2005, Dodd and Gutierrez [9] introduced an episodic memory system for a cognitive robot. The system was implemented on the humanoid robot ISAC, and stored temporally-based experiences from past interactions. A retrieval component allowed evaluation of the correct episode for a given situation. The authors argued, based on experimental evidence, that *"the system is able to draw attention to information that is both novel and salient"*. Different types of emotional feedback can be used to rate the importance of the information.

Ho et al. [14] investigated how episodic memory could be presented in a robot for an HRI environment, namely for a showcase in the Robot House at the University of Hertfordshire, UK. The implementation consisted of semantic memory described as *"simple"* and *"passive"*. It was capable of collecting data about the user in an HRI. The authors said that many lessons were based on to the technical difficulties emerging from integrating memory components in cooperation with information providing sensors and actuators. Ho et al. [13] showed how the system could be used in a robot to visualize episodic memory, and to support elderly people. For instance, when elderly people need to remember past events like when they took medicine. The user could browse the robot's data to check if the interaction (e.g. taking medicine) took place or not.

Kasap and Magnenat-Thalmann [20, 21] investigated the influence of knowledge about past events in a tutoring scenario. In the first study, the robot Eva acted as a teacher for a course on computer networks. Throughout the interaction, the robot asked participants predefined questions and reacted to their responses. Preliminary results showed positive reactions to the perceptions of the system. In the second study (from 2012) the system was extended and used as a tutor for digital photography. The model for knowledge about past interactions was based on the similarities between episodic memory, a Belief-Desire-Intention architecture and Hierarchical Task Network Planning. The authors used a 2x2 design to test support/non-support in combination with and absence of memory about past interactions. For the study, 52 participants interacted with the system four times over two weeks. It should be mentioned here that the participants where members of the computer science department (32) and of the media design department (20). As such, they may have possessed some knowledge about how to interact with a robot. During the interaction, the robot greeted the participant and asked predefined questions in accordance with a course on color digital photography. The system used participants responses to decide whether memories from past interactions should be used or not. Each session lasted approximately 15 minutes. Results showed that memory of past events significantly affected the users engagement. An effect was found for the supportive factor items friendliness and dominance, although not for intrinsic or extrinsic motivation. Comparing effects between the first and the last session showed that participants in the condition with memory became more interested in the interaction, while those without memory tend to lose interest quickly.

The results indicate that a memory component for past events can establish a greater engagement between human and robot, and therefore should be integrated in gaming scenarios.

Conclusion on Memory for HRI

The ability to create memories of past interactions is essential for human beings to understand and learn. New conclusions can be drawn from past experiences and used in future interactions. For a social robot, this capability is also of importance. Being capable to remember past events helps to evaluate situations, which is of interest especially for social interactive robotics. To understand upcoming situations, it is important that a robot be able to learn from past interactions. This can help to solve problems based on actions learned. How a robot is perceived is influenced by whether it remembers and utilizes past information. In social interactions, humans tend to rate robots with memory capabilities more positively. When a robot facilitates past information in repeated interactions, the system tended to be perceived less like a tool.

2.3. Interaction Scenario: Playing Games with a Robot

Most interactions between humans and robots are designed in a simple manner to limit the number of variables to be analyzed. Whenever an autonomous robot interacts in a complex scenario, the possibility for errors and misbehaviors increases. Often scenarios are chosen that do not represent interactions in the real world or are not familiar to a naive user.

Interactions based on playing games can help humans to interact with robots more naturally. In research scenarios, the gaming context can help to detract participants from the artificiality of their surroundings and situations.

Also playing games with robots creates a one-to-one situation in which the robot can be a companion or an opponent. This then creates scenarios in which tasks need to be solved together or in which competition arises. Several approaches have been developed to play games with robots.

Janssen et al. [15] investigated how a robot could support children in learning arithmetic principles. They proposed that learning by playing could result in better performance. A study conducted with the robot NAO in the ALIZ-E project showed that even after three sessions the children continued to like playing with the robot. The children continued to interact with the robot even in their free time, as was also highlighted by the authors. Allowing participants to interact freely with a robot can lead to more natural interaction, and can offer an opportunity to study how motivated humans interact with robotic systems.

Jost et al. [16] used the robot NAO to study how a robot was perceived as compared to a tablet device. The authors implemented a version of "Simon's game" where a sequence of colors is shown and must be memorized by each player. The authors research question was how a stressful interaction with a robot compared to the same interaction with a tablet device. The also evaluated whether there was an effect if the tablet was present when the participants interacted with the robot. The hypothesis was that participants would be less concentrated. 67 participants (20 male / 47 female) played in three condition (only robot, only tablet, robot and tablet). The results supported the hypotheses that a robot was more enjoyable. When compared to a tablet device, the authors found that people tended to be less positive toward facial expressions in contrast to an interaction with only a robot.

Using childhood games renders explanation about the structure of the interaction unnecessary. Another advantage is that playing games can result in more common ground about what to do and how to interact. For instance playing a game of tic-tac-toe is simple and very fast to learn even if never played before. The vocabulary for elements of the game is simple, as well.

Short et al. [38] implemented the game "Stone-Paper-Scissors" (RPS) for the humanoid robot Nico. The authors investigated the effect of a cheating robot in the context of this game. They were interested in research on how people characterize cheating, and how the mental state and engagement of the robot is rated. To this end, they used a robotic hand that was capable of displaying the three figures of the game. A dialog system announced the outcome of a round. For the sake of speed and responsiveness, the experiment used a Wizard of Oz (WoZ) setup to remotely control the robot, and provide the answers to the results of each round. Three conditions were tested. In condition one, the robot cheated by announcing an incorrect result. In condition two, the robot changed the outcome after the result of the participant is known. Condition three was a control-condition where no cheating was used. 55 participants (23 male / 32 female) each played 20 rounds, and the robot cheated in four of those rounds. The authors found that a cheating robot was more engaging compared to a fair robot. The participants rated an incorrect announcement in the dialog as a malfunction. These behaviors are not rated as bad, although the action-cheating was rated as cheating, and is not rated very positively.

Walters et al. [43] implemented the RPS game on the robot platform CHARLY. The robot CHARLY has a "simplified human-like appearance (humanoid)" and can move autonomously in people oriented environments. The Goal of the study was to investigate the effect of different faces showing emotions while playing the game with a human partner. For the study the robot interacted with 82 participants (56 male / 26 female) that could play voluntarily whenever they passed the robot. The authors mentioned that the study showed differences on how participants evaluate the use of the robot. From the results the authors stated that male participants preferred a simple and unchanging face display while female participants preferred a richer and more dynamic display.

A more complex game situation was used by Kim and Suzuki [23], where they implemented the game of poker. The authors compared human-human interaction (HHI) to HRI. In the HRI setting, the robot was controlled by a WoZ system. Ten participants were recruited, and each played five rounds. While playing, the robot used facial expression (smiling), hand movement, eye blink and eye movements. They investigated bluffing between the HHI and HRI conditions. The authors argued that their result showed that there was a difference in the humans behaviors, and that the bluffing decisions made differed between the conditions. The authors suggested that these findings *"provide the possibility to build a humanoid robot to perform complex behaviors by modeling human opponents based on a given situation and human responses in a poker game"*.

Becker-Asano and Meneses [1] created a gaming interaction with the hybrid agent MARCO. The authors combined the virtual agent with a robotic arm and implemented the game of chess. While playing, the system showed emotions in response to game events. The system was designed to evaluate how artificial agents could influence humans emotions. The authors speculated that *"a human player's enjoyment will increase together with higher levels of emotional contagion"*.

For the system described later in this thesis (see chapter 3.2), the classical game of pairs was implemented. The game offers a one to one interaction with cognitive challenges. Louie et al. [30] conducted a study with a robot playing the game of pairs. The robot Brian interacted as a socially assistive companion in the game. The study was run in a long-term care facility with 22 participants (8 male / 14 female). The robot acted as an assistance partner that was capable of recognizing turned cards, announcing the result, and trying to cheer up the participants whenever no pair was found. The results showed that the interaction promotes positive attitudes towards the robot. Also, the selected scenario was rated to be funny and easy, even for participants without prior knowledge of intelligent systems.

Conclusion for Interaction Context

These studies as a group make several points about human-robot gaming interactions. Whenever a game is played with a robot, people tend to be more positive towards the interaction. When errors occur during the interaction, the robot tends to be perceived as less machine-like, which is also positive. Playing games with robots allow for a more natural way to analyze how motivated persons interact. This also helps to evaluate what elements are needed to make the robot behave in a more social manner.

Also, the interactions can be less sterile, such as with artificial situations that are often used in laboratories. Choosing a game context that is already known can enhance the perception of the interaction. Games that are known can establish a common ground for cooperative or competitive tasks.

2.4. Summary

In this chapter, three fields were examined: long-term interaction with robots, approaches to memory representation, and interaction scenarios using gaming situations. For the first, much information can be derived about creating new interactions with robots. Choosing a gaming interaction can lead to a more natural interaction, and may result in a more companion like perception of the robot. Using memory approaches can support natural and social interactions. The following chapter (see chapter 3) will focus on long-term interaction in a gaming situation. In chapter 4 the gaming system will be enhanced to use a simple memory approach.

3. Long-Term Interaction in an Isolation Study

Astronauts working and living in space must deal with many physiological and psychological challenges. The solitary environment requires that support usually be provided remotely by ground control. Most vital parameters for each astronaut can be determined externally. Psychological parameters are more difficult to observe, and must typically be assessed verbally. Normal communication is often indirect. Moreover, communication during deep space missions can be very delayed. The lack of communicative exchanges can result in changes and possible decreases in task performance.

In cooperation with the German Aerospace Center[1], Bielefeld University initiated the SoziRob[2] project. The goal of the project was to investigate the benefits of robotic systems to space missions as assistive partners. A long-term study was carried out to simulate isolation similar to that experienced by astronauts on their missions. The research questions concerned social interaction and their benefits, as well as long-term technical challenges to robotic systems. Two interaction scenarios were investigated throughout the study.

The first focused on physical exercise. Astronauts need to exercise to cope with the effects of zero gravity on the human body. To support physical maintenance, a training scenario guided by a robot was designed. Selected was indoor cycling as the sports activity. Indoor cycling provides interval training with the guidance by an instructor. The human-robot training was implemented using human-human training as a model. The humanoid robot Nao[3] served as instructor and guided the participants through each session, giving advice at each training step.

The second interaction scenario focuses on the social interaction between a human and a robot. Astronauts face several different psychological effects under isolation, including depression (see chapter 1). A conversation partner is not always available, communication with ground control is not instantaneous. This can lead to boredom and loneliness, in particular during recreational periods. Social assistive robots can be provided to act

[1]DLR - http://www.dlr.de

[2]SoziRob Project - https://aiweb.techfak.uni-bielefeld.de/projekt-sozirob
The project was funded by the Federal Ministry of Economics and Technology due to resolution of the German Bundestag (support code 50RA1023) and by the the DFG (EXC 277 CITEC)

[3]NAO - http://www.aldebaran.com/en/humanoid-robot/nao-robot

as communication partners in these situations. Social interaction with a robotic partner should help to investigate effects known to emerge under isolation. To ensure that the interaction itself has no effects on mood, a more entertaining context was chosen for the study. The interaction was based on a game, namely the game of pairs, that allows a human to play against a robot. The interaction was designed to cheer the participants and to decrease stress. As a benefit, the game of pairs provides a cognitive challenge that can help to improve participants' memory. The competitive factor of the game may help to motivate participants to apply themselves throughout repeated interactions. A technical goal of the system was to be functional and operational during the entire study. The implemented interaction was intended to provide a fully autonomous robotic system.

To simulate the work of real astronauts, the isolation study was conducted in the laboratories of the German Aerospace Center. Throughout the study, the participants completed many tasks during their daily routines. A timetable was designed with slots for every task.

For the study, two different conditions were investigated. The experimental group was accompanied by robots during their daily routines. One robot functioned as a cycling instructor, and the second as a game partner for social interaction. The control group worked on the same tasks without support from robotic systems. During the sports part, the robot was substituted by a display providing feedback similar to the robots instructions. During social interaction, the game of pairs was not supported by an interactive system. For the gaming task, the participants played alone by just looking up pairs.

The topic described throughout this thesis is based on the social interaction scenario developed in the project. The next part describes the robotic system used for in interaction. Due to the comparison in the sports scenario, comparing a robotic system with a system using a display, no similar interactive system was provided for the gaming situation in the control group. Because the control group is not truly comparable, this thesis only focuses on the robotic part of the social interaction used in the study.

The following chapter describes the implementation of the different components used for the scenario. This covers the definition of the game-play, the system components to model the game-play, the dialog system for communication, the different vision components as well as the control elements for the robot. The system description will be followed by the evaluation of the study conducted in 2013. The chapter will close with a discussion of the results and an analysis of ways to extend the system for further long-term interaction.

3.1. Game-Play Definition for the Interaction

The game of pairs is a structured game that can be played by two players. The interaction between a human player and a robot was based on this structure. A state graph was defined that contains the main actions of the game (see fig. 3.1), and formed the basis of the different components implemented. The following six items represent the structure of the implemented interaction for the game of pairs.

Interaction Initiation

To initiate the interaction, the human needs to address the robot with a greeting call. The robot reacts by greeting the player and initiates a small talk dialog. This dialog contains questions about how the player performed in the previous task or how he feels at the moment. The game introduction is initiated thereafter.

Game Introduction

During the game's introduction, the player is asked if the game rules need to be explained. This dialog is followed by asking whether the field setup requires explanation. The player is then asked to distribute the cards on the playing field. While the cards are being placed, the robot observes the field. The number of cards detected is communicated to the player. This intended to help the player to understand how the robot perceives the cards. After all cards have been placed, the player who will begin the round is negotiated. A dialog allows the human player to choose whether the robot or the player begins. Based on this selection, the robot's turn or the human's turn is initiated.

Robot Turn

If the robot starts the round, or a turn for the robot begins, the first card is selected using vision and control components. The selected card is then announced to the player via dialog by providing coordinates given by column and row. After the player turns the card, the system evaluates the turned card. If the card matches the card requested, the robot goes on to the next card. If the wrong card was turned, the robot announces the mistake and requests that the correct card be turned. For the second card, the same procedure takes place. After the correct card is turned, the robot evaluates the result. If no pair is found, the robot requests turning the cards. After all cards are turned,the robot cedes the game to the other player. If a pair is found, the robot requests removing the cards. When the cards are removed, the robot starts its next turn. If no cards are left after a pair has been found, the system moves to the round closing part.

Human Turn

If the player starts the round or begins a turn, the robot asks the player to turn his cards. When two cards are turned, the robot evaluates and announces the result. If no pair is found, the robot asks to turn the cards and begins its turn. If a pair is found, the robot asks to remove the pair. After removing the cards, the robot asks the player to start his next round. If no cards are left after a pair is found, the system moves to the round closing part.

Round Closing

If no more cards are left the robot announces the winner based on the amount of pairs found by both parties. If the player and the robot parties found an equal pairs, a draw is announced. After the announcement, the robot offers the player the opportunity to begin another round. If the player wishes to play again, the robot returns to the point of setting up the field. If another round is not requested, the robot initiates the closing part of the interaction.

Interaction Closing

The closing part of the interaction begins with a small talk dialog. This dialog contains questions about whether the player enjoyed the game or what the next task to be performed will be. Once answered, the robot wishes the player farewell, and switches its state to idle mode, awaiting for the next player.

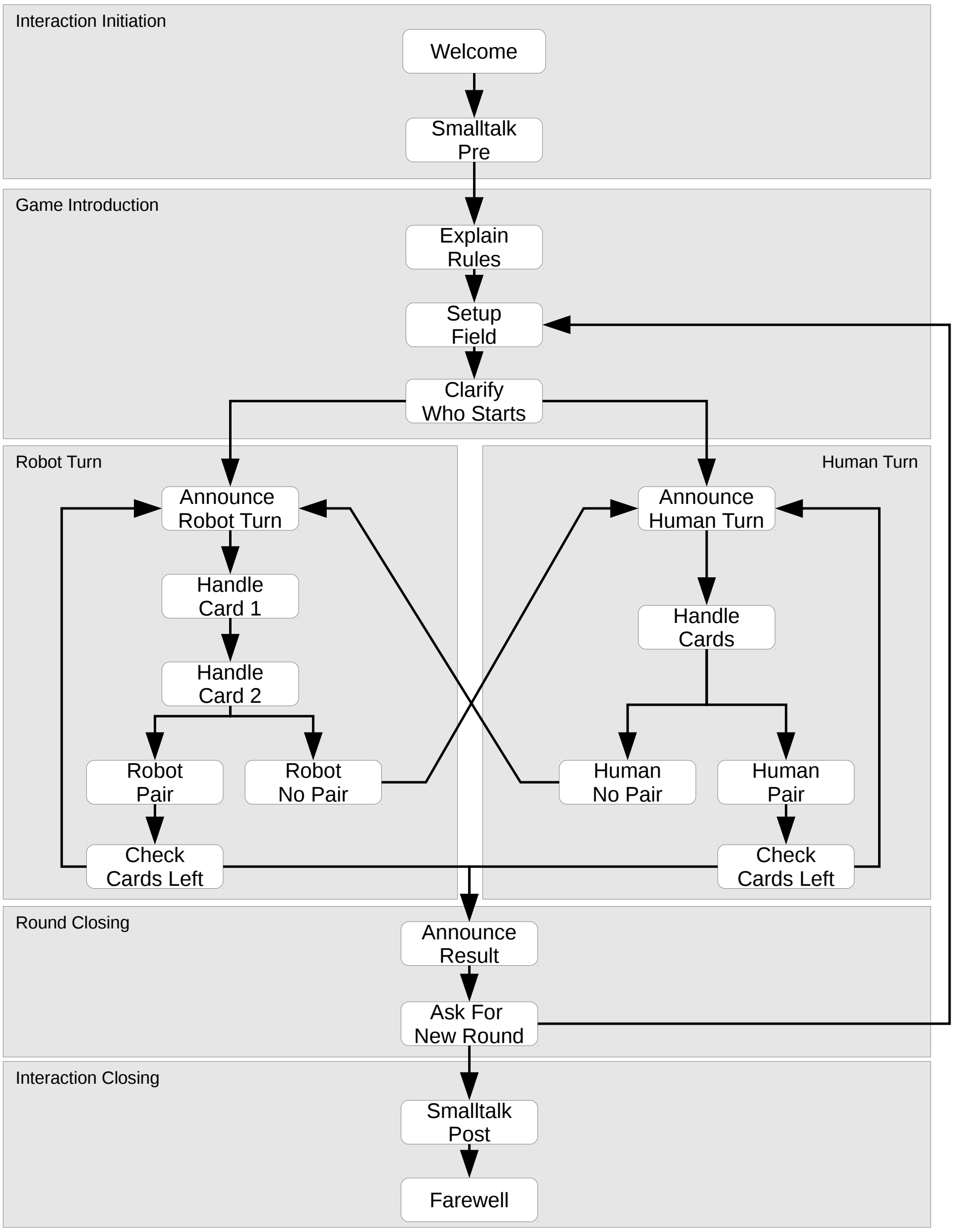


Figure 3.1.: The game of pairs structured for human-robot interaction. The left side represents the robot's turn. The right side models the human's turn.

3.2. Implementation of the Scenario

Based on the definition of the interaction, several components are implemented and combined (see fig. 3.2). Throughout the interaction, several vision components detect the human player and the cards placed on the playing field. Based on game events, the dialog component triggers the production of speech or reacts to speech recognition. Events from the control engine trigger movements by the robot head. This allows focus on important elements (e.g. looking at the player's face or at the playing field) or to show facial expressions.

To establish communication between the components, interprocess communication was used. The selected framework was the XML enabled Communication Framework [44].

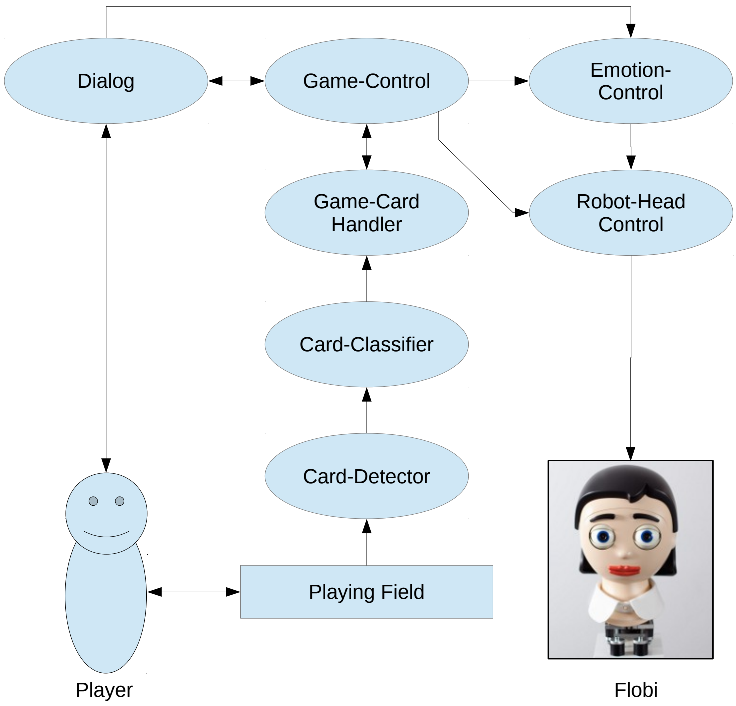


Figure 3.2.: The components implemented for the game of pairs. The system's core is the state-engine. Based on external events, the state-engine controls the dialog system and the robot's motion.

The following sections describe all components implemented to create the interaction for the game of pairs. First, the control engine used to model the structure of the game is described. Following, all necessary components to perceive and manipulate the interaction are described.

3.2.1. Control Engine: Modeling the Interaction

The structure of the game (see section 3.1) was implemented using a state machine. This game-control component represents the process of starting the interaction, the main game-play and the closing of the interaction. The game-control engine was realized by using a state-driven-engine based on the state machine notation for control abstraction (SCXML[4]). For the execution of the state machine, the StateChartExecutive of the eXtensible Task Toolkit Framework (XTT[5]) was integrated. [31]. The StateChartExecutive handles the state chart, triggers events and executes transitions for state changes. Additionally, a data model can be provided to store simple information inside the state chart. The structure of the game is represented by three main states (see fig. 3.3).

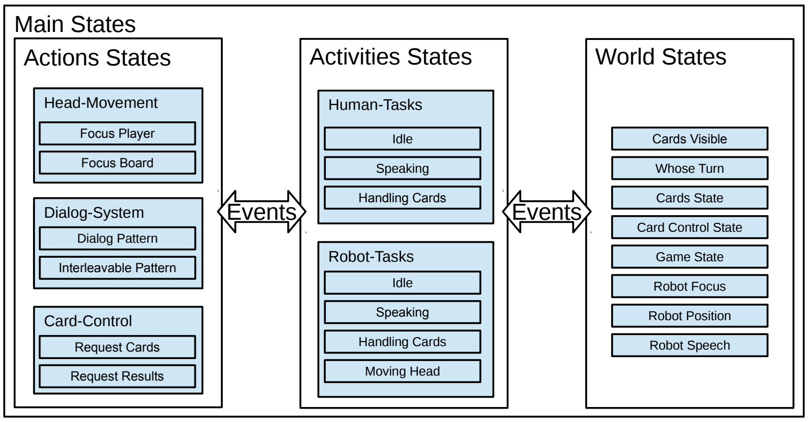


Figure 3.3.: The main states implemented in the state chart component. Actions trigger events for external components. Activities represent tasks to be performed by the robot or the human. World state represents the current situation for the whole game.

Action States

The action state defines how external components are controlled to realize the game-play. Inner states defined in the action states provide control for communication with the dialog component and the card control. Additionally, one inner state controls the movement of the robot head. Two positions are used throughout a game (see fig. 3.4). The board position is used to focus on the playing field. A position towards the player is used to track faces whenever the robot addresses the player. Each inner state listens on changes in the activity states and the world states. When changes occur, the inner

[4] SCXML - http://www.w3.org/TR/scxml/
[5] XTT - http://openresearch.cit-ec.de/projects/xtt

states change based on predefined conditions (e.g. execute dialog for card request if the robot's turn begins).

Activity States

The activity states represents the current task performed by the player or by the robot. The states are used by the conditions in the action states. For the robot the inner states contain speaking, handling of cards and head movements. For the human part, the inner states contain speaking and handling of cards. Inner states are switched based on the actions currently executed by action states.

(a) Robot position for the view of the playing field

(b) Robot position for the view of the player

Figure 3.4.: Camera views for the positions of the robot head. Image (a) shows the view used for card detection and classification. Image (b) shows the position used for verbal communication with the player.

World States

The world states represent the current situation for the whole game. Each inner state represents configurations of external and internal situations (e.g. whose turn it is, how many cards are turned, where the robot is looking). Inner states activate based on results of actions executed by the action states. Additionally, external events can activate inner states (e.g. a card result received from a card detection). Based on elements of world states, a data model stores values about received events (e.g. number of cards visible).

3.2.2. Dialog Component

For the dialog component, the Pamini framework [36] is integrated and enhanced. This framework makes it possible to model communication between a human and a robot by using different predefined *patterns*. A pattern defines how dialog tasks are processed. One pattern consists of definitions for inputs (e.g. speech recognition) and outputs (e.g. speech production). For a pattern, different phrases can be defined that are selected by the component throughout the interaction.

Listing 3.1: An example dialog pattern used if the robot wins the round. The element <*phrase*> configures the text output to be executed by the speech output component. The *impulse* value defines the emotional effect to be forwarded to the emotion-control component.

```
 <?xml version="1.0" encoding="UTF-8"?>
<patternConfiguration>
 <robotDialogAct state="initial" type="R.notify">
  <output>
   <phrases choice="random" impulse="0.3">
    <phrase> Diese Runde habe ich gewonnen. </phrase>
    <phrase> Ich gewinne dieses mal. </phrase>
    <phrase> Damit habe ich gewonnen. </phrase>
   </phrases>
  </output>
 </robotDialogAct>
</patternConfiguration>
```

For the game of pairs, each verbal interaction between the robot and the human is defined as a pattern (e.g. explain rules, ask to turn a card). For the game scenario, a total of 61 patterns are defined (for an example pattern see listing 3.1). For the robot, 41 patterns are defined to communicate actions and intension. These patterns provide information about the game or announce tasks the robot requires the player to perform. For the player, 20 task patterns are defined. A player task is used to request manipulation of cards or to give explanations about the current game state. These patterns also define possible requests to be triggered by the player (e.g. requests about who currently leads, correction of a wrong result announcement).

Additionally, the dialog component is used to control emotional feedback. This is done by extending a value for the emotional impulse. Here, positive impulses are defined in case of an advantage by the robot (e.g. robot found a pair, robot wins the round). Negative impulses are used in case of an advantage by the player (e.g. player found a pair). When a pattern is triggered, the defined value is forwarded to the emotion-control component (see section 3.2.5).

3.2.3. Vision-Components for Card Handling

For the game of pairs, cards placed on the game field need to be detected and classified. To communicate cards of interest, the detections need to be transformed into a coordinate system understandable by the player.

A vision pipeline is implemented to handle the detection and classification of cards (see fig. 3.5). The pipeline is based on the iceWing framework[6]. The framework allows combining vision plug-ins and takes care of acquiring images. To simplify the detection of square shaped cards an existing plug-in was used to create a top down view of the playing field.

The following paragraphs describe the detection of cards on the game field and the classification of detected cards. The last paragraph describes the component implemented to transform the image coordinates of the cards into a coordinate system used for verbal exchanges with the player.

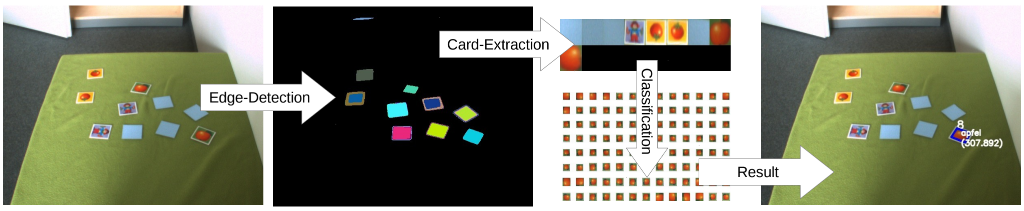


Figure 3.5.: The vision pipeline. First the cards are detected using an edge detection approach. Detected cards are classified by linear regression based on the training material.

Game-Card Detection

To differentiate the game cards from the background and the game field, a game card detection component was implemented based on an edge detection approach (see fig. 3.6). It uses the sobel operator[39] and the canny edge algorithm [5]. Contours are detected in a binary image containing only edge information. For each detected contour, a rectangle that fits into the contour is selected. All found rectangles get checked for overlap and whether they fit a predefined size. Matching rectangles are extracted as thumbnails. Each thumbnail is exported to external components for further processing.

Additionally to the detection of cards, the component allows manually creating training data for the classifier component. The detected image thumbnails can be stored to create training material. Creating a wide range of training data requires that each card is moved to different positions in the training field throughout the recording process.

[6] iceWing - http://icewing.sourceforge.net/

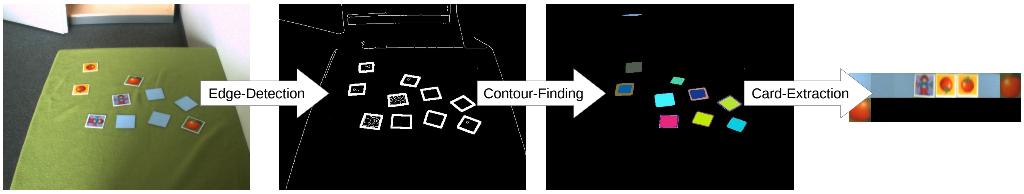


Figure 3.6.: The detection pipeline. First, a binary image is created by applying edge-detection. Possible detections are selected for extraction using contour finding and rectangle fitting.

Classification of Cards

The implementation for the classification of cards is based on a linear regression approach. It was reproduced from the face identification approach proposed by Naseem et al. [34], and is applied to the detected cards. Training results are then transformed to class specific models. Each class model is represented by a matrix that contains all training images as column vectors. Each input card from the detector is transformed to the same type of vector. The distance for each model is computed for the given input vector. The shortest distance between input vector and all models is regarded as the classification result. The classname is also stored. For the exact mathematical definition see [34, chap., 2.1]. A parameter allows defining a threshold for the maximum distance to the class model. Cards with a distance above the threshold are regarded as not flipped (called background). For the study described later in this thesis (see section 3.3), a total of 400 training examples was recorded per class using cards from a child's game (see fig. 3.7). The creation process involves the corresponding functions in the detection component.

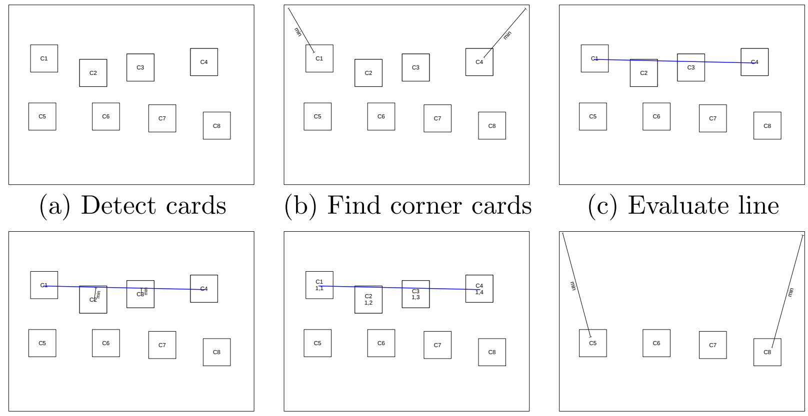
The classifier evaluates all detected cards and creates an output containing the results. For external components, the number of flipped cards is determined and added to the result (no cards turned, one card turned, two cards turned, more than two cards turned). If the amount matches two cards, the result is also determined (pair or no pair).

Figure 3.7.: The set of cards used for the interaction (©Ravensburger). In total, eleven pairs were trained. The unturned side of all cards is shown in the lower left corner of the image.

3.2.4. Game Card Handler Component

The game card handler component was implemented to store information about cards throughout the interaction. To keep track of the cards, the component stores references for detected cards in a card model. This model contains information on positions and classification results. The model is updated based on new detections and classifications.

To communicate card positions to the player, the component evaluates coordinates for the column and the row based on pixels (see fig. 3.8). From the list of all detected cards, the nearest cards to the upper corners are selected. Between both corner cards a line is evaluated. For each card within a predefined distance to the line the corresponding value for row and column is assigned. For the maximum distance to the line the width of a card is used. Once marked, the card gets removed from the list of detected cards. When no more cards are found near the line the corner cards get their numbers assigned. These cards are also removed from the list. Now the algorithm begins from the beginning until all cards are successfully labeled. The coordinate system was evaluated [25], and results showed that the system was well understood by the participants.



(a) Detect cards (b) Find corner cards (c) Evaluate line

(d) Find nearest cards (e) Assign numbers (f) Remove cards

Figure 3.8.: Creating card coordinates from detected cards. All detected cards are labeled according to their positions.

After creating the coordinates, a lookup for pairs is performed. The lookup compares the classification results between all detected cards. Cards with matching class names are marked as pairs. The results of the coordinate system and the pair matching process is published for use by external components.

Additionally to handling cards, the component acts as an information provider. As such, the component forwards information and cards based on requests by the external components. When a card is requested, the component evaluates the card model by searching for a known pair. If a pair is known, one card of the pair is provided to the request. If no pair is known, a random card is selected.

3.2.5. Emotion-Control for the Facial Expression of the Robot

The robot uses facial expressions to provide emotional feedback for the player. The emotion-control component was implemented as a layer between the current emotional state of the game and the robot. The emotional state of the robot is represented by a value for pleasure. The range for the pleasure value runs from -1 (very sad) to 1 (very happy). These values were distributed into five emotional ranges (see table 3.1). The value for pleasure is changed by the dialog component (see section 3.2.2).

(a) Happy (high) (b) Happy (low) (c) Neutral (d) Sad (low) (e) Sad (high)

Figure 3.9.: Configurations for the emotional faces. Flobi uses five configurations in the studies to show emotions. Facial configurations go from very happy (a) to very sad (e).

Like the game-control component (see section 3.2.1), the emotion-control component was designed as a state machine. Three main states were defined to control the robot's emotional feedback: the emotional states, the emotional handler and the emotional face-control.

Name	**Range**	
Happy High	>0.5	<= 1.0
Happy Low	>0.1	<= 0.5
Neutral	>= -0.1	<= 0.1
Sad Low	<-0.1	>= -0.5
Sad High	<-0.5	>= -1.0

Table 3.1.: Names and ranges of the emotional states. States change on receiving updates for the value pleasure. If a state is changed, the corresponding emotional face configuration is sent to the head-control component.

The emotional states represent the five ranges for the emotion. The inner states are changed based on the pleasure value. Every change of the value results in changes to the current emotional state.

The emotional handler is responsible for listening to external events that change the pleasure value. Events are based on emotional impulses sent by the dialog component. Positive impulses result in an increase in the pleasure value. Negative impulses decrease the value. Additionally, the emotional handler continuously increases or decreases the pleasure value whenever no new input occurs until zero is reached.

The emotional face-control represents the current face configuration executed in the robot (see fig. 3.9). Given the case of a state change in emotion, the corresponding face configuration is triggered.

Figure 3.10.: The robot Flobi used for the entertaining interaction in the study of the SoziRob project.

3.3. Evaluation for the SoziRob Study

The goal of the SoziRob project was to evaluate how social assistive robots could support humans that live and work in an isolated environment. As such, two interactive systems were implemented, one using a robot as a sports trainer and one with a robot playing a game. This thesis refers to the latter scenario.

The system described in section 3.2 was evaluated in a long-term study under isolation conditions. The study was conducted in 2013 in cooperation with the German Aerospace Center. Sixteen participants were isolated for a total of 18 days in the ANSAM research laboratory at the German Aerospace Center in Cologne.

The study consisted of two conditions, a group supported by the robotic systems was compared to a group without robots. Contrary to the sport interaction, no comparable system was provided for the social interaction in the control group. As such, the following evaluation focuses on social interactions with the robot Flobi. Throughout the study, the robot was placed inside the living room of the research laboratory, and was active during the entire study (see fig. 3.10). According to a daily time table, each participant was instructed to play alone with the robot for one hour per day.

The following section begins with the design of the study. Described is the setup inside the laboratory, and how the system was integrated into the daily routines of the participants. The follow up section describes the design of the questionnaires. Each participant was asked to complete a questionnaire after each interaction with the system.

Figure 3.11.: Floor plan of the AMSAN research laboratory at the German Aerospace Center in Cologne. The robot setup for the interaction scenario with the robot Flobi was located in the living room.

The last section covers the analysis of the results.

3.3.1. Study Design of the Interaction

The isolation study was conducted inside the AMSAN research laboratory (see fig. 3.11). The laboratory offers a closed habitat for up to 8 participants. Inside the laboratory, there are several rooms for work related tasks, a living room, a kitchen and bathrooms. For the study, a total of seven rooms were used for daily work tasks by the participants. The setup for the robot Flobi was located in the living room. The robot was placed at the end of the dining table.

The study was designed to run for a total of 21 days. Two days before the main isolation section began, the participants were introduced to the different tasks and systems. For the social interaction task, two participants at a time were introduced to the robot Flobi. While the introduction took place, the robot focused on faces and looked around. An examiner explained the game to be played and showed the cards to be used. At this point, no actual interaction was shown or exercised by the participants. The examiner strictly explained how the interaction would begin and end once the main part of the isolation began. The participants were advised that the robot itself would instructs the participants while interacting, and explains the interaction in case of further inquiries.

The main section of the study was designed to take place over 18 days. During this period, predefined timetables were handed out to each participant. Social interaction with the robot represented one task in the timetable. Two days (day 7 and day 14) were defined as weekends. On these days, only sport tasks were obligatory. The HRI with the robot Flobi was optional. Nevertheless, the robot was active and ready to play on both days.

Each session with the robot was designed to last 45 minutes. Throughout the interaction, the participant played alone with the robot, although other participants could stay in the room in their leisure time. Each participant was advised to play a minimum of two rounds per session. Depending on the time left, participants were allowed to play as many rounds as they liked. Participants completed a questionnaire when they were finished.

On the last day of the study, after the participants were released from the isolation part, each participant was interviewed. The interviews were conducted by an examiner using a predefined questionnaire concerning the isolation part.

3.3.2. Questionnaire Design for the Interaction

This section describes the questionnaires completed by participants throughout the study. They include questionnaires intended for before and after the study, and those conducted as interviews after the isolation part of the study ended.

Questionnaires for Social Interaction

After each session, participants were asked to rate different items concerning their interaction with the robot. Each questionnaire consisted of three sets of questions with a total of fourteen items. To ease discomfort from completing so many questionnaires throughout the whole study, all questionnaires were presented on a tablet PC.

In the first group, the participants rated how they perceived the *physical and mental requirements* of the interaction. This group consists of seven items using a scale from one (rated low) to twenty-one (rated high). The items rated were:

- Mental Demand
- Physical Demand
- Time Demand
- Performance
- Exertion
- Frustration

The second group evaluated their *positive attitudes towards the interaction* using a 7-point Likert scale. The questions were:

- How excited were you about the interaction? (Excitement)
- How motivated were you while interacting with the robot? (Motivation)
- Did you enjoy the interaction with the robot? (Joy)

The third group was used for the participant's self-assessment on *physical and mental condition* using a 7-point Likert scale:

- How exhausted do you feel? (Exhaustion)
- How agitated do you feel? (Agitation)
- How positive is your current mood? (Mood)

In additionally, all participants indicated the number of rounds played. Also provided was one free form field for remarks about the interaction. This field was not obligatory.

Additional Ratings on the Final Day

In additional to the questionnaires for each session, a paper and pencil questionnaire was provided on the final day of the isolation part of the study. This questionnaire contained three free form fields for remarks on the following items:

- Please describe your impressions of the interaction throughout the study.
- Have you experienced technical difficulties at any point during the interaction? If yes, please describe in brief.
- Please provide brief concluding remarks of you impressions of the entire interaction.

Also rated was the participants' perception of the presence of the robot using a 7-point Likert scale for the items:

- Disturbing
- Distracting
- Suitable
- Annoying
- Intimidating
- Irritating
- Motivating
- Confusing
- Pleasant
- Helpful
- Observing

The questionnaire ended with general questions using a 7-point Likert scale for the following items:

- I would interact with the robot again.
- The system needs improvement.
- I enjoyed interacting with the robot.
- The interaction improved over time.
- I felt well prepared for the interaction.
- I knew how to react to the system.
- The system reacted as expected.
- The system worked reliable.
- I understood the systems intention.
- The system motivated me.
- I would prefer playing with the robot instead of a human.

Questionnaire on Negative Attitudes Towards Robots

All participants were asked to complete a questionnaire on negative attitudes towards robots (NARS, [35]). It was distributed to the participants before and after the study. The questionnaire consists of three groups with a total of fourteen items. The first group evaluated *Negative Attitudes towards Situations of Interaction with Robots* (NARS1, 6 items). The second group rated the *Negative Attitude toward Social Influence of Robots* (NARS2, 5 items) and the third group rated the *Negative Attitude towards Emotions in Interactions with Robots* (NARS3, 3 items). For each item, a scale from 1 (*I strongly disagree*) to 5 (*I strongly agree*) was defined.

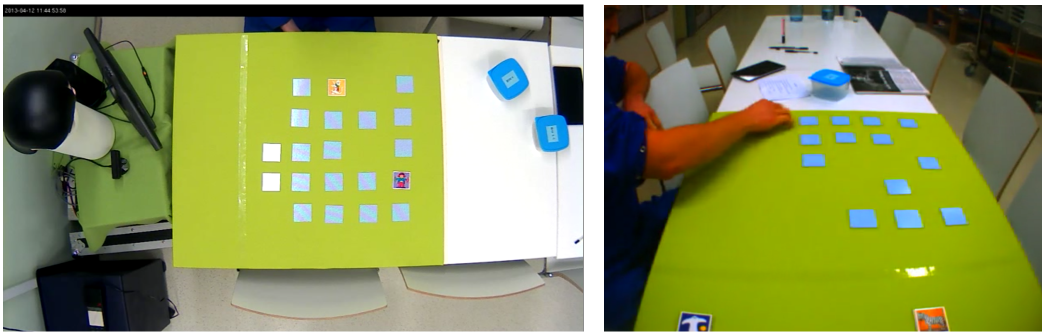
Interviews after the Isolation Study

After the isolation study, an interview was conducted with each participant. The examiner asked various questions concerning the perception of the isolation study. Answers were collected as short notes. A total of 37 predefined open questions were asked during the interviews. Five items were used to gather information on social interaction with the robot Flobi. These were:

- Name three to five adjectives for the Robot.
- What features of the robot would you improve?
- Did you notice any changes throughout the study?
- Did you adapt your behavior throughout the study?
- Please order all tasks in terms of how much you enjoyed them.

3.3.3. Data Collection and Logging

Throughout the study, approximately 13 TB of data were collected. From these data, approximately 1 TB belonged to the gaming interaction with the robot head Flobi. The data from the interaction consisted of videos from the perspective of the robot (camera image of the left eye), audio recordings of the interaction, and videos from the overview camera placed above the playing field (see fig. 3.12).



(a) Camera view on top of the scenario

(b) View from the robots camera

Figure 3.12.: Views of the cameras used to record the scenario with the robot Flobi. Image (a) shows the from view above the gaming field. Image (b) shows a view from the robot camera.

Additionally, system events for every interaction were recorded. By combining timestamps of the different recordings, a whole corpus on human-robot interaction could be built. Two recordings containing observations of the entire living room, could also be used (see fig. 3.13).

A handwritten log was kept by the examiners in the control room to collect information on system failures and problems throughout the interactions. This log contains all requests by the participants regarding misunderstanding and problems with the robotic system. Also recorded were technical failures recognized from observing the components used by the system.

(a) Camera observing the dinning table (b) Camera observing the living room

Figure 3.13.: Views from the cameras used during the entire study. Image (a) shows the view of the dining table and the robot Flobi. Image (b) shows a view of the whole living room.

3.3.4. Results

For the robotic condition in the isolation study, eight participants (Age M = 23.625, SD = 4.307) were recruited. Each participant was a pilot candidate applying for a position at the German Aerospace Center. The main part of the isolation study was carried out during a total of 18 days. During this period, a total of 127 social interaction sessions took place. A questionnaire was completed for each session. Only one session was skipped due to a power failure on the compound that resulted in a system breakdown.

Results of the Ratings based on the Questionnaires

Each questionnaire completed after the interaction (see section 3.3.2) was analyzed for effects occurring over time. To evaluate the questionnaires, the mean values for each item were computed for each day. The resulting values were not normally distributed. A non-parametric Friedman test of differences for repeated measures was conducted for each item in the question groups (see table A.4, page 99).

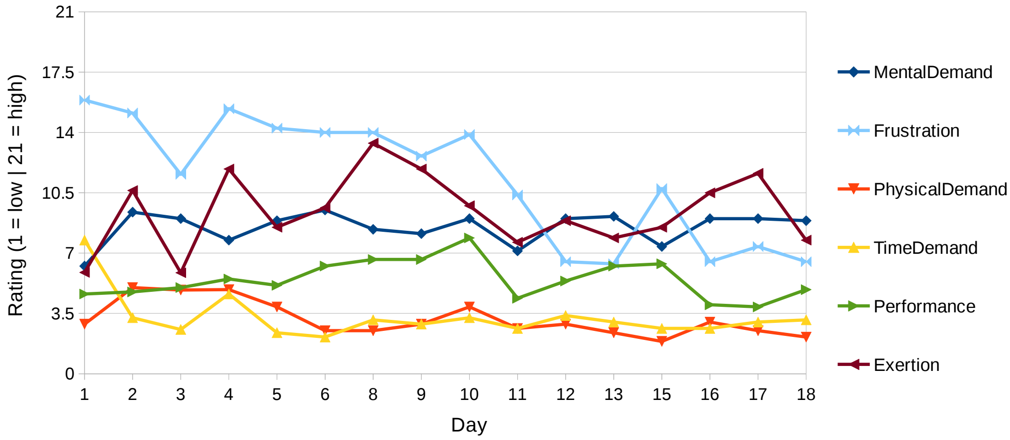


Figure 3.14.: Mean values for each day for the question group on *physical and mental requirements* of the interaction. A Friedman Test showed a significant difference only for the item *Frustration*.

Based on the mean values for the group on the *physical and mental requirements of the interaction*, only the item *Frustration* showed a statistically significant difference with a $\tilde{\chi}^2(15) = 38.941$, $p = .001$ (see fig. 3.14 and table A.1, page 97). For all other items, the test showed no significant differences over time. For the group on *attitudes towards the interaction* (see fig. 3.15 and table A.2, page 98) and the group on *physical and mental conditions* (see fig. 3.16 and table A.3, page 98) the results showed no significant differences over time.

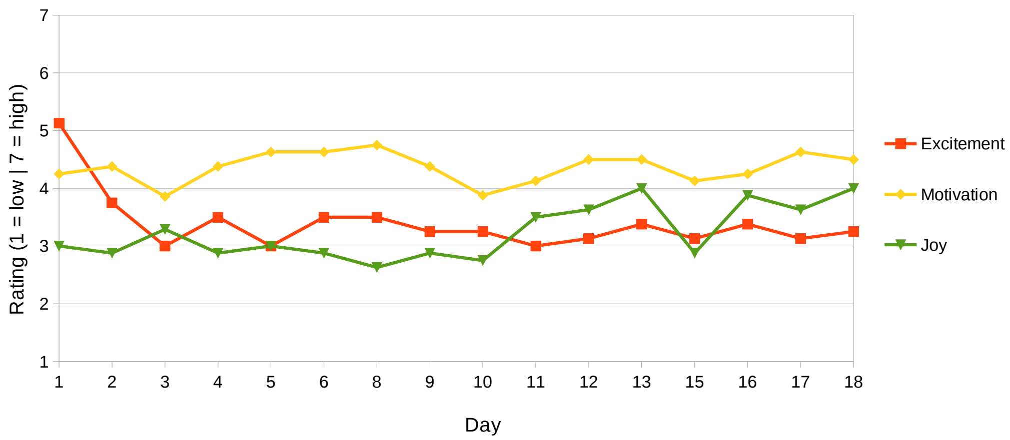


Figure 3.15.: Mean values for each day of the question group *attitudes towards the interaction*. A Friedman Test showed no significant differences.

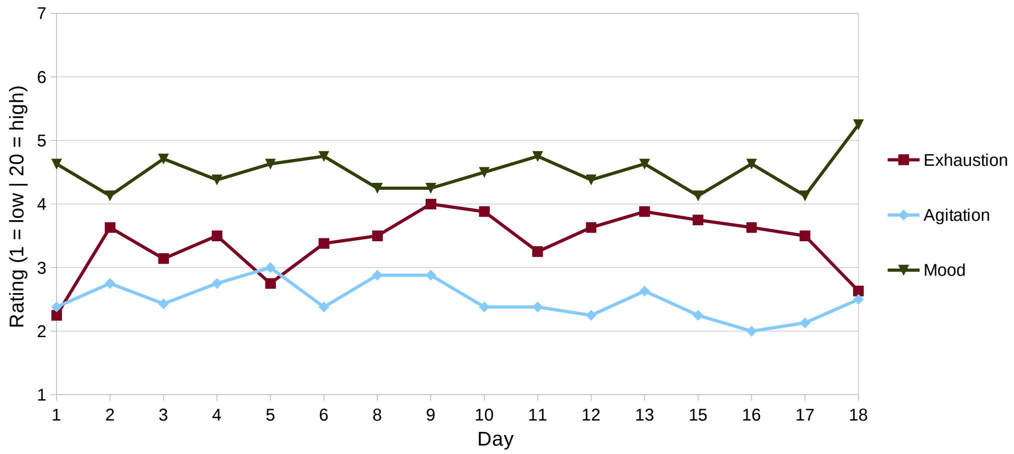


Figure 3.16.: Mean values for each day of the question group *physical and mental conditions* of the participants. A Friedman Test showed no significant differences.

Results of the Additional Ratings on the Final Day

On the last day of the isolation study, the participants rated how they perceived the interaction. The results of the ratings were not normally distributed. A one-sample Wilcoxon signed-rank test was conducted for each item.

For items covering how the robotic system was perceived (see fig. 3.17), the items *Suitable* (W(8) = 28, $p = .011$), *Annoying* (W(8) = 34, $p = .023$) and *Pleasant* (W(8) = 28, $p = .014$) showed significant differences compared to a value of one (the value one refers to *not at all*). Results are presented in table A.12 (see page 107).

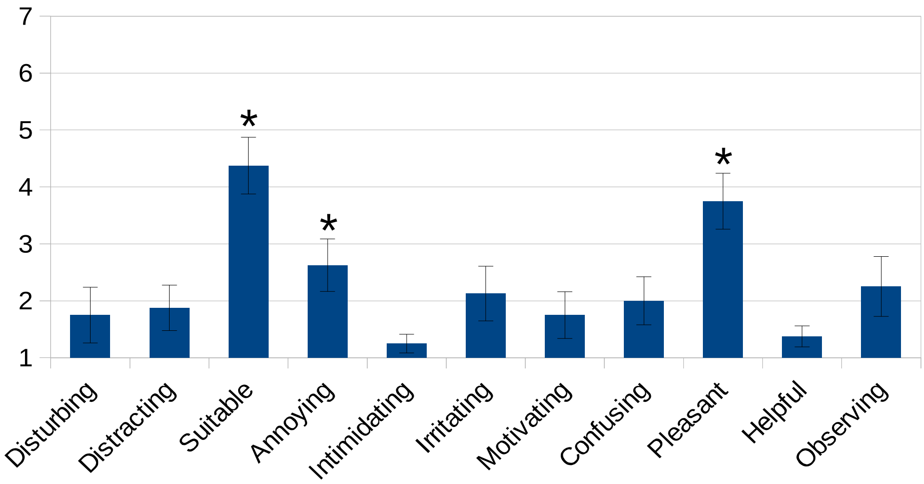


Figure 3.17.: Ratings covering how participants perceived the robotic system. The item *Suitable, Annoying and Pleasant* showed significant differences from the value one (not at all). $^{*}p < .05$.

General questions asking items *I felt well prepared for the interaction* and *I would prefer playing with the robot instead of a human* were found to be not significantly different from a value of one (see fig. 3.18 and table A.13, page 108).

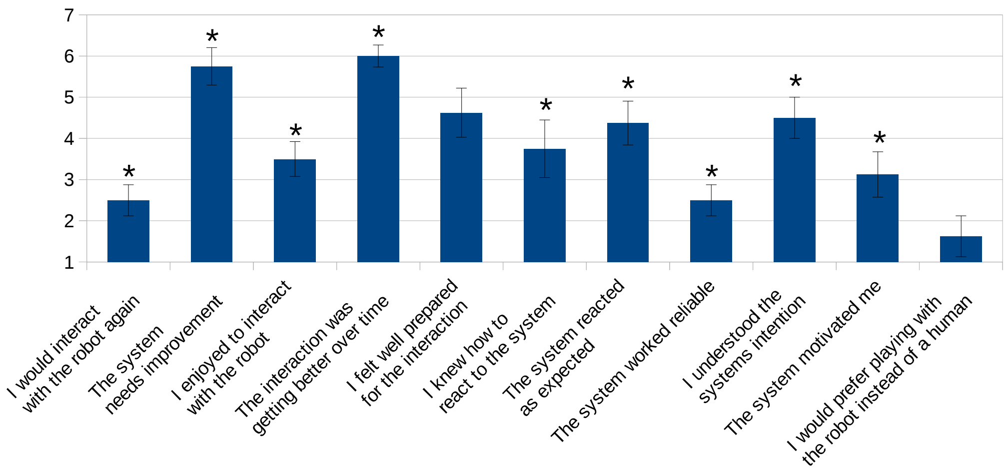


Figure 3.18.: Ratings for the general questions asked the final day of the study. *p < .05.

For the free form text, the remarks for the first item (*Please describe your impressions of the interaction throughout the last days.*) showed that the participants felt that the interaction improved over time (see table A.5, page 100). Still the interaction showed deficiencies due to several problems. This was also remarked in the second item (*Have technical difficulties throughout the interaction occurred over time? Please describe in brief.*) with comments on issues with misunderstandings in the dialog, as well as the faulty detection of cards (see table A.6, page 101). The item for concluding remarks for the entire interaction showed that it was perceived as interesting, although improvements were needed (see table A.7, page 102).

Number of Rounds Played

For every session played, the participants marked the number of rounds played. In total, 226.5 rounds were played (M = 2.098, SD = 0.052). A non-parametric Friedman test of differences among repeated measures was conducted on the mean count of rounds played per day (see table A.14, page 109). Results showed no significant differences over time ($\tilde{\chi}^2(15) = 13.648$, p = .552).

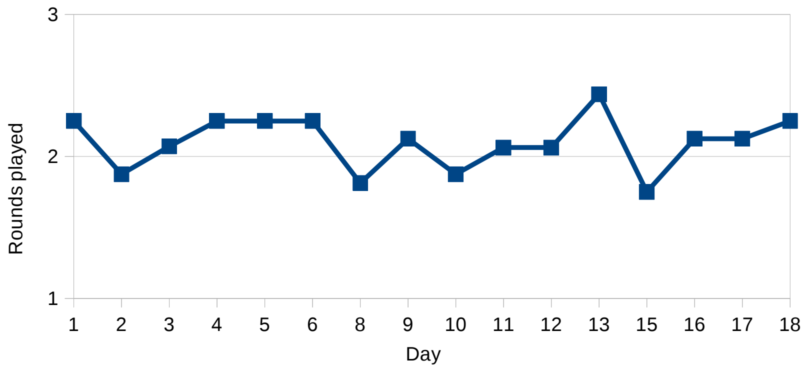


Figure 3.19.: Mean number of rounds played per session. A non-parametric Friedman test showed no differences over time.

Results of the NARS questionnaire

Before and after the study, participants rated items on negative attitudes towards robots (see section 3.3.2). A paired t-test was conducted for each item (see fig. 3.20). The results showed no significant effects for the different groups (see table A.15, page 109). The group *Negative Attitudes towards Situations of Interaction with Robots (NARS1)* showed a marginally significant difference between the ratings before and after the study [t(7) = -1.664, p = .07]. The eta^2 statistic (.315) indicated a large effect size. The group *Negative Attitude toward Social influence of Robots (NARS2)* and the group *Negative Attitude towards Emotions in Interactions with Robots (NARS3)* showed no significant effects.

Results of the Interviews

All participants were interviewed after the isolation part of the study. Each participant had to answer predefined questions (see section 3.3.2), five questions concerned the gaming interaction with the robotic system.

The first question involved adjectives describing the robot (see table A.8, page 103). From the handwritten notes, the most remarked adjectives concerned the problems that occurred throughout the interaction ("faulty", "need of improvement", "imperfect", "incomprehensibly").

For the items to be improved, the most remarked elements were speech recognition as well as card perception (see table A.9, page 104). Additionally, some participants suggested, that the robot should use more facial features throughout the interaction.

Question number three measured system changes noted by the participants. The written notes indicate that all participants thought the system improved over time (see table A.10, page 105).

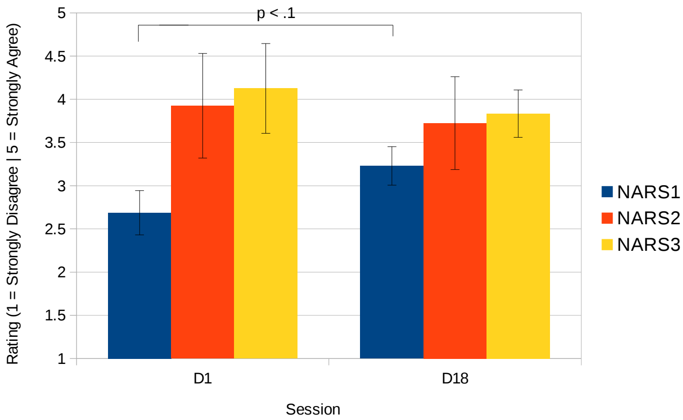


Figure 3.20.: Mean values for the items on *Negative Attitudes towards Robots*. Only the item *Negative Attitudes towards Situations of Interaction with Robots (NARS1)* showed a marginally significant difference for the ratings before and after the study.

For the forth question, participants indicated that they changed their behaviors over time by speaking more directly to the robot and by using shorter commands (see table A.11, page 106). In terms of the ranking of tasks by preference carried out throughout the study, the gaming interaction placed 6th out of 9.

Performance of the Robotic System

The isolation part of study was run for a total of 18 days (432 hours). Throughout each day, the participants attended to different tasks for nine hours per day. The robot Flobi was active 24/7. When the robot was not playing in a predefined session, it looked around focusing on faces found by the face tracking components. Throughout the study, the participants interacted with the robot on a total of 127 occasions for a total time of approximately 94 hours (around 45 minutes per session). One interaction was interrupted due to a power failure in the laboratory on day three. Log entries indicate seven faulty instances (8.89%) were recorded. On these occasions, the system became non-responsive, and a remote counteraction was needed. For each faulty session, the corresponding components were restarted, whereby participants could end their tasks.

3.4. Discussion

This section will discuss the results from the isolation study in 2013. To this end, the requirements formulated in the introduction (see section 1.1) will be discussed based on the results of the different questionnaires and the interviews.

R1 - Create a Social and Entertaining HRI

One requirement proposed was to create a social and entertaining interaction. As such, implemented was a one to one gaming interaction scenario with a robot.

The proposed system was successfully tested for the main study in 2013. In total, the robot was active and playing games on a total of sixteen days. The robot was also active on both weekend days, although unfortunately, the participants did not interact with the system on these days. Reasons for this may have included that the participants did not enjoy playing with the robot or that they favored the opportunity to use the internet and communicate with the outside world.

The number of rounds played, as indicated by the questionnaires, showed that these did not significantly differ from the number of required rounds. The participants were asked to play a total of two rounds, and only a few participants played more rounds when there was time left until the next task. Despite the problems, these additional rounds indicate that entertainment may be a factor in fostering greater commitment to interaction with a robot, even when no longer required.

The results from the questionnaires indicate, that the initial stress (leading to frustration), decreased by the end of the study. The analysis for the item *Frustration* found that the ratings changed significantly over time. Reasons for this may include that the participants had to learn how to interact with the robot. For some of the participants, the coordinate system was not clear. In addition, placing cards on the playing field seemed to be counterintuitive. From the feedback given by the robot, participants could not figure out how to add extra space between cards. The other participants helped the player by giving advice on how to place cards. Advice of this type helped to solve most problems in the later part of the study. Another point of frustration was the faulty speech recognition. The system often misunderstood commands given by the player. Instead of reacting as expected, the system would then just repeat the last instruction. This was mentioned in the questionnaire provided on the final day and explained in terms of the misunderstandings resulting in unnecessary time extensions for the game. During the subsequent interviews, the participants mentioned that the surrounding noise also negatively influenced these situations.

Data on the attitudes towards the interaction found that participants' excitement for the upcoming interactions decreased. This represents the novelty effect resulting in a more exciting interaction in the beginning and a decrease over time. Nevertheless this effect was not found to be statistically significant. It shows that the participants were still looking forward to the interaction. A reason for the decrease in the rating could be the nature of static interactions that did not change over time. Every game was the same, except that each game could lead to different results based on how lucky one player was in finding a pair. A system acting in a linear and static manner favors the novelty effect, and this can lead to a less positive perception of the robot in later interactions. This explanation was supported by the results of the interviews and the final questionnaire. The participants remarked that a more varied dialog would have been preferred, to make the interactions more interesting. Participants als commented on how predictable the robot's responses became. The participants said they would play with the robot again if the game-play became faster and more dynamic. Also mentioned was the lack of a wide range of facial expressions throughout the gaming interaction. More dynamic facial feedback in case of special situations throughout a game (e.g. finding three pairs in a row) could make the interactions more exciting.

The physical and mental conditions found no major impact on the participants ratings. The rating for exhaustion was average over the whole study. The ratings for mood where higher than average, and did not change significantly over time. Results also showed that the interaction itself had no major influence on the participants conditions.

From all items rated in the questionnaires for the gaming interaction only the item *Frustration* showed a significant decrease over time. None of the other measures changed over time. The novelty effect was only found for the item *Excitement*, as a tendency towards the effect. This indicates that a gaming interaction can help to cope with novelty effects, and may offer an interesting activity over time. Nevertheless, it appears that static interactions favors these effects, and should be avoided.

R2 - The Entertaining Interaction should provide a Cognitive Benefit.

The questionnaires showed no significant effect for cognitive load, although the values did increase over time. This may indicate a tendency that such a system might be challenging and therefore interesting over time. Comments from the interviews indicated that the system needed enhancements to the card detection parts. The robot sometimes reacted to fast, when participants were still manipulating cards, resulting in wrong classification. Errors lead to slow game-plays, suggesting that learning effects may be influenced. On a more positive note, participants noted that these errors decreased over time, although the system itself was not altered through the course of the study. The decrease my be attributed to participants learning how to better interact with the robot. Some participants mentioned that by using shorter commands when addressing the robot, better results could be obtained. Another reason for this rating may have been a better understanding of how to interact with the system. By understanding how to place cards and when to manipulate them, a more fluent game-play could take place. A more stable game-play may result in a greater cognitive benefit. The participants also mentioned in their conclusions, that they had expected the robot to be a perfect playing opponent. Because of the problems mentioned, the participants focus lay on how to handle the system instead of focusing on the memorizing part of the game. It seems that at this stage, the participants understood the game in terms of playing with a child that did not learn or was difficult to handle. The participants pointed out that by solving these issues a more interesting and therefore challenging game-play could occur.

R3 - Create a Robust Interaction for a Long-Term Study

The third requirement concerns the robustness of the proposed system. Under isolation and without proper knowledge, technical systems are difficult to maintain. Because of this, the interaction and the robot should allow the possibility to be remotely maintained. The interaction itself should allow the participants to react and handle such problems by themselves.

From the hardware side, the system was active for nearly the entire isolation period. Only on one occasion was the robot system shut down for an hour and a half. The shut down was due to a power blackout on the compound of the German Aerospace Center. As a result, the interaction in this occasion was stopped, and the robot restarted. After the incident, all subsequent interactions were performed without further disturbance. The isolation part of the study ran for a total of 432 hours. The robot was active for approximately 430 hours and played the game of pairs for approximately 94 hours.

From the system side, the log showed that 91.11% of all sessions were begun and terminated correctly by the participants. Faulty sessions resulted from various components malfunctioning. On each occasion, the participants informed the examiners using radio communication. The correction and repair of the errors was handled remotely from the control room. Within several minutes, the components were restarted and the participants could go on with their task and complete the interaction. This shows that the system works most of the time and that remote maintenance works in an acceptable way.

Conclusion to the Long-Term Study

The designed system worked very well throughout the study. The interaction system was successfully implemented and integrated. Throughout the study, the system interacted on 127 occasions, and was active for nearly the entire time. The massive amount of data collected will allow for further long-term research. The proposed system showed that a gaming system could be designed to act as an interesting interaction partner, even after repeated interactions. However a static dialog structure seems to favor the novelty effect, resulting in less interesting interactions. The results and the remarks by the participants suggest that a more dynamic solution would be preferred, and therefore should be sought out. The errors occurring on the system side can unsettle the participants, possibly leading to a lower regard for the system. Additional problem handling protocols in such cases would be a solution. Nevertheless, these findings offer the proposed system as a baseline for further research in the field of long-term interaction, and should be enhanced to cope with the difficulties discovered throughout the study.

3.5. Summary

This chapter described a system for a gaming interaction between a human and a robot. Presented were the different components implemented to realize the game-play and to structure the interaction. The system was evaluated in a long-term study conducted in cooperation with the German Aerospace Center. In section 3.3 the study design and the results were presented. The results were discussed in section 3.4. It was shown that the requirements could be fulfilled in principle. From the interviews conducted after the study it was remarked that a more correct working card perception as well as the speech recognition would be preferred by the participants. Another point mentioned was the static structure of the interaction leading to a less interesting game-play over time. Nevertheless the system performed well throughout the study. The hardware and software worked stable, showing that the designed system can be used in long-term interactions. The following chapter 4 will look at possibilities to integrate enhancements for the proposed system based on these findings and present a study conduced to evaluate them.

4. Long-Term Interaction with Contextual Knowledge

The system implemented here was reviewed based on results from the SoziRob project (see section 3.4). The goal was to enhance interaction capabilities and enrich the long-term interaction.

One point mentioned by the participants of the long-term study involved the static dialog. The interviews contained references to the lack of a dynamic and varying dialog. The static structure of the interactions appears to promote the novelty effect. To cope with it, the system was extended to collect information about past interactions. Based on this, the system should create additional feedback to be utilized in the dialog component. Providing feedback on the current and past interaction may help to create more interesting interactions. Moreover, this data can be used to make decisions based on how well the participants performed in the previous games (see listing 4.1 for an example). This could lead to a greater commitment by the participants, and therefore may result in a more interesting game-play, even in later sessions.

Listing 4.1: Example of how the robot could provide decisions based on game statistics.

```
<Query context: Player has lost more games>
Robot: "You have lost more games in the last session,
        will you work harder this time?"
Human: "Yes, this time I will win."
```

Also mentioned in the interviews was the problem of misclassification errors. This resulted in unnecessary situations in which the participants tried to repair the system's decisions. For the system used throughout the SoziRob study, the vision components and the speech components did not work perfectly. To evaluate if errors affected the perception of the system, a remote control interface was created. A more realistic and accurate interaction could be simulated by allowing a human to control the system in the background. Errors of detection and classification could be removed. If the participants are not aware that the system is remotely controlled, such a system can influence the perception of the robot, as well as the interaction.

The following sections describe the components implemented for the context knowledge part. Also presented are some extensions applied to the existing components. Additionally, the remote control component will be specified, needed to simulate a more realistic robot. Following a description of the components, a study conducted at Bielefeld University will be presented. The results section will show the findings of the study. The chapter will close with a discussion of the results and concluding comments.

4.1. Extensions Made for the Gaming Interaction

Using information about past interactions can enhance human-robot interaction. With this capability, a robot used for social interaction may be perceived less like an object, even if the interaction takes place on repeated occasions.

The gaming scenario and the system described in chapter 3 were enhanced to allow the robot to utilize information about past events. This extension was designed to be attached to the current system, without a major reworking of the existing components. This section will describe the enhancements to existing components and the components implemented additionally.

4.1.1. The Context Manager Component

The context manager component handles all the information collected between the human player and the robot throughout the interaction. The component stores data for each player individually. The data can be manipulated by external components, and results get forwarded to the dialog component. To distinguish between the statistics of the currently played round, those of prior games, and information about the states of the current game and external components, the component handles three scopes:

- **current-round**:
  The current-round scope only contains information about statistics for the actual ongoing round played between the robot and the player. This scope gets reset at the start of a new game.
- **game-summary**:
  The game-summary contains information about the statistics of all games played between the robot and the player. This scope also includes information for the current round.
- **game-state**:
  The game-state scope contains only information about the current state (e.g. whose turn is it, how many cards are visible). It also represents information from the external components (e.g. what are the results of the card classification).

The scope for the current round and the scope for the game-summary consist of a total of 16 items each. Presented are the statistics for the player and the robot. For each party, statistics are stored for the following items: *games won, games lost, drawn games, turns made, games started, pairs found and pairs found in a row.* Additionally the number of *games played* and the *current time of the round* is stored. For the current implementation, only the summed up values for the game-summary were exported to the dialog component. The implementation offers an extension to export the chronologically output of the recorded information.

The game-state scope contains data about the information received from external components. Saved are variables for the information about visible cards (e.g. information on turned cards, classification results), the state of the interaction (e.g. system idle, playing a game) or who leads in the current game. Like the current-round scope, the game-state scope is cleared at the beginning of a new round.

The context manager allows external components to update information in the scopes. A context update contains four elements:

- **Variable:**
  The variable to be updated by the value using the given action.
- **Action:**
  The action to be performed on the data. Possible actions include updating, setting or resetting a value.
- **Value:**
  The value that alters the current value for the given scope. In case of an update action this value is added to the current value. The set action sets the value for the given variable.
- **Scope:**
  Defines the scope to be changed for the given variable.

Based on the action the variable for the given scope will be updated with the value provided. After a given action is applied, the updated information is forwarded to the dialog component.

For the study (see section 4.2), the context manager component was provided with a graphical user interface (GUI). This interface allows manually setting the currently active player. Additionally, the GUI allows choosing between two different methods of forwarding the collected information. The first method publishes all information stored throughout all interactions with the given player. The second method forwards only the information of the game-state scope, leaving all statistics out that were stored for the current game and for the games played in the past. This second method was integrated for evaluation purposes. It allows for comparison of the extended system utilizing context information, with a static system. This static system behaves like the system used for the SoziRob project and provides no additional information on past games.

4.1.2. Enhancements for the Game-Control Component

The implementation of the game-control component (see section 3.2.1) allowed forwarding of data about the current state of the game using a data-model. This information were only used for local knowledge on the games state. No information about results or statistics was provided. The state machine used in the SoziRob was quiet complex. In total, 227 states (some in parallel) with 565 transitions were defined. The integration of statistics for different players directly inside the state machine could result in many more complex sub states. An extension inside the game-control component would unnecessarily expand the component and could create additional overhead, and might result in delayed reactions in the system. Also adding further players could be difficult to handle due to the simple structures provided to represent data.

To reduce the complexity of handling different players, and to allow external components to collect information about ongoing games, an additional context-update state was introduced (see fig. 4.1). The context update state performs alongside the main states. This minimizes effects on the main states and the game-play. The context-update state listens for events of the main states (e.g. switching of the currently active player). Based on these events, statistics can be published for external components. Upon detection of an event, the context state forwards the information to the context manager using the predefined update information (see section 4.1.1)

In total, the context updates state listens on eleven events that occur in the main states. These events result in updates for information on pairs found, games played and ended, results of the current game, turns made by the robot and the player, as well as who started how many games.

4.1.3. Enhancements to the Dialog System

The integrated dialog framework used variables presented in a data model that can be manipulated by external components. The values from variables can be used for pattern definition, and utilized as text for an output phrase. The default pattern of the dialog framework can not be defined to select an output phrase based on the values of this data model. To achieve this, the phrase interpretation needs to be extended to allow conditions to be set for each phrase.

The pattern definition of the dialog component was extended to use information from past interactions. The extension was implemented as part of a student's project in the applied informatics group. The implementation was enhanced for use during gaming interaction. The extension allows adding conditions to the output phrases (for a sample extension, see listing A.1, Page 123). It provides an operator, a variable and a comparison value. The defined operators make it possible to check whether the values are equal, greater or less then the comparison value. The value for the variable is derived from the data model, and is compared to the provided value. When the condition is met, the phrase is selected as output. Additionally, a reference for the selected phrase is stored. This reference ensures that in case a given pattern is recalled, the next possible phrase is selected. This helps to minimize repetitive usage of selected phrases. In case the

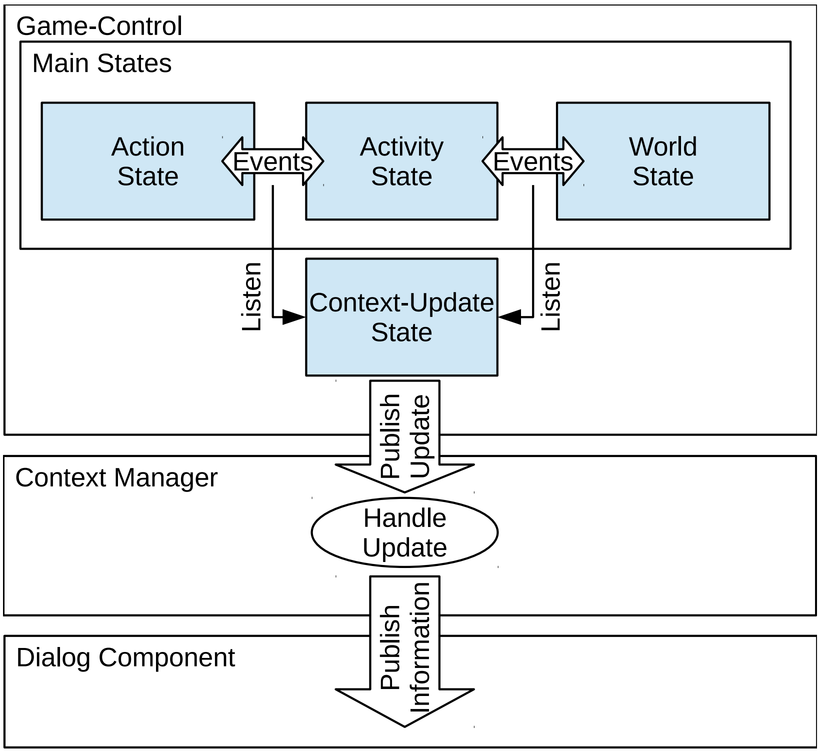


Figure 4.1.: Extension to the Game-Control Component. What is applied is the Context-Update State listening on the main states for events used for statistics. Received events are forwarded to the context manager.

condition of a phrase is not met, the next phrase is examined.

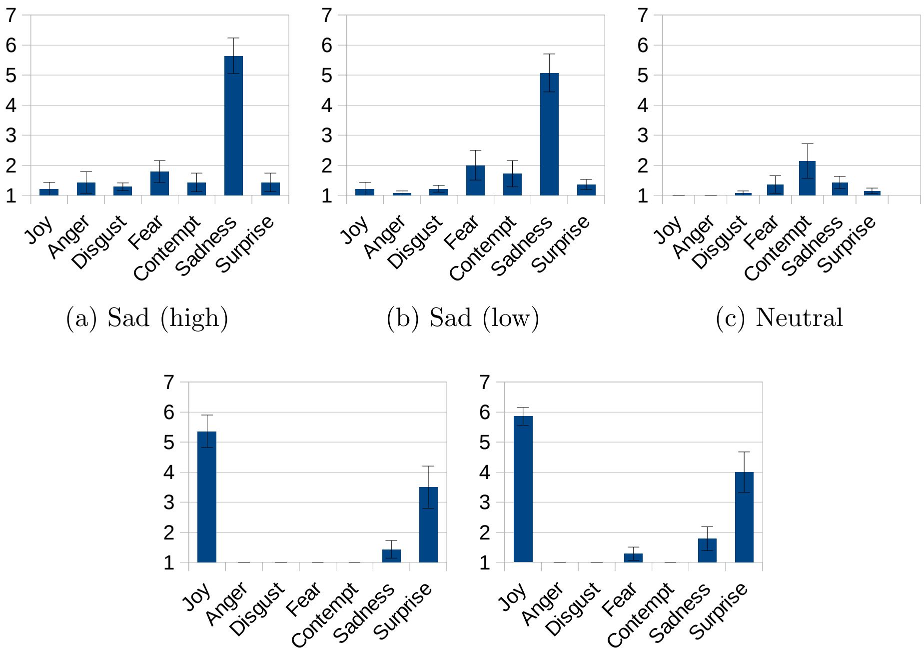
To better use the additional data provided by the context manager, 22 out of the existing 61 dialog pattern were extended. The conditions for each extended pattern were based upon the stored statistics provided in the data model of the context manger.

4.1.4. Enhancement for the Head-Control Component

It was suggested in the interviews after the isolation study, that the robot should make use of more facial expressions throughout the interaction (see section 3.3.4), because this could lead to a more positive perception of the system.

The emotional system was reused for the study on contextual knowledge. Although the same facial states (sad, neutral, happy) were used, their final configuration was reworked to reduce ambiguity.

These new configurations were evaluated using an online study. Participants rated the transitions from a neutral face to one of the five emotional faces (sad high, sad low, neutral, happy low, happy high). The videos were rated by 14 participants (4 female / 10 male, Age M = 28.64, SD = 5.99). A non-parametric Kruskal-Wallis Test showed significant ratings for all five configuration corresponding to the intended expression (see table A.16, page 110). The results indicated that the new configurations were perceived as expected by all participants (see fig. 4.2).



(a) Sad (high) (b) Sad (low) (c) Neutral

(d) Happy (low) (e) Happy (high)

Figure 4.2.: Ratings for the new configurations of the facial expressions. Ratings matched the intended expression.

Additionally, a blinking behavior was introduced to the head-control component. Human blinking occurs every four to six seconds and one blink takes 300 to 400 milliseconds [33]. A similar behavior was integrated for the robot. The blinking action is performed by a fast closing and opening of the four lid actuators. Due to the mechanical sound of the movement, the delay between two blinking events was defined to be longer than a human blink (15000ms + random offset of 1500 ms). This longer period was also selected to reduce the effect of a disturbed image in case there were classifications by corresponding components. The blinking part of the head control stores the configuration before the blinking event and restores the values after the event. This ensures that the lids correspond to the configuration before the blinking event.

4.1.5. Remote Control for the Scenario

The participants of the study in the SoziRob project mentioned the systems vulnerability to misclassification (see section 3.3.4). These problems were reported with regards to the visual detection and classification of cards. Most errors for the classification components occurred when the participants were still manipulating cards while the robot tried to evaluate them in its field of view. The other problem concerns speech recognition. Due to the noisy environment, the robot did not always understand commands and therefore repeated sentences. The assumption was made that a perfect robot would never lose, and perform better than a human player. To test this assumption, data was gathered from the interaction and the robot itself using a remote control component. This component allows an examiner to control every action performed by the robot. The component mimics the game-control component used for the autonomous setup. The structure used for the autonomous state machine was represented inside a controllable interface using GUI elements. Each task forwarded to the dialog component was provided by a given GUI control element. To create a control sequence equal to the autonomous system, the GUI follows the same predefined game structure (see section 3.2.1). The examiner controlling the interaction is forced to select the same actions at the same points throughout a game. Selecting an action results in switching to the next state of the game represented by its view. Triggering an action also results in sending a corresponding task to the connected components (e.g. trigger a dialog task, request the head to focus on the playing field).

For the vision part, the examiner must select the column and row for the card to be manipulated by the player (see fig. A.1, page 121). To this end, predefined fields can be selected with the GUI for a given card. To control the processing of speech output, the examiner can select answers for questions based on a set of predefined answers.

Additionally, it is possible to trigger restarts of rounds or to stop the interaction using an independent information view (see fig. A.2, page 122). This view also allows reactions to be performed by the controller, for questions from participants that do not match the current action (e.g. asking who leads while waiting for the robot turn). If the participant asks questions not relevant for the game-play, the controller can trigger a notify pattern, explaining that the robot did not understand the command.

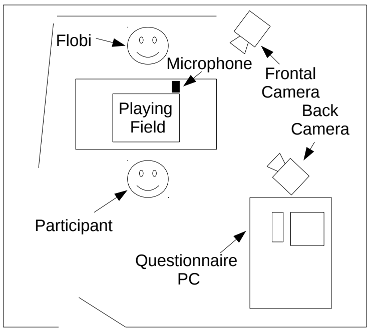
4.2. Evaluation of the Scenario Extensions

The system proposed in the preceding section (see section 4.1) was tested in a study at Bielefeld University. Due to the complexity of the study from the SoziRob project, a whole study under isolation could not be performed. Therefore the system from the the first study was reused and tested again for better comparability. To compare the different extensions, the following conditions were evaluated:

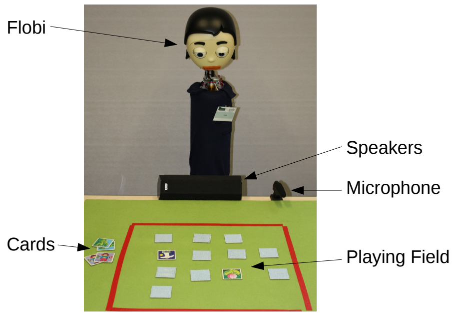
- **Condition C1:**
  Interaction with a fully autonomous system. No context knowledge is provided throughout the interaction
- **Condition C2:**
  Interaction with a fully autonomous system. The system uses the components for context knowledge from past interactions.
- **Condition C3:**
  Interaction with the remote controlled system. The system uses the components for context knowledge from past interactions. Participants are not aware that the system is controlled.

First, the autonomous system without context knowledge (C1) is compared to the extended system using information from past interactions (C2). The third condition (C3) compares effects from a perfect playing system with the autonomous system using context knowledge (C2). For the third condition the participants will not be aware that the system is remotely controlled. This information will be revealed after the last session.

Before the first condition, the participant was advised that the robot knew the name of the participant. Their name was given to the context manager by the examiner before each game.



(a) Laboratory Setup



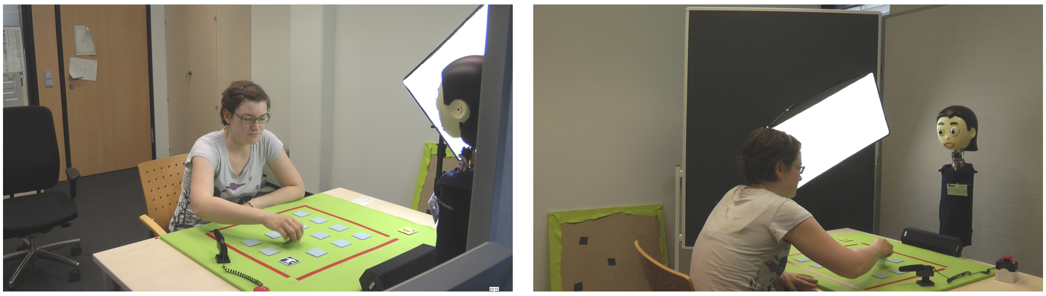
(b) Game Setup with Flobi

Figure 4.3.: The setup of the study for the context scenario. The game field is placed in front of the robot providing markers for the playing area.

The studies for the different conditions were conducted in a single room laboratory. The study was designed to collect data for each participant over four sessions. Participants were recruited using posters and directed emails to students of Bielefeld University. Participants received a monetary reward and a chance to win a monetary voucher.

During the interaction, the robot-head Flobi was placed next to a table in the laboratory (see fig. 4.3). In front of the robot was a board with a green blanket. On the blanket a square was marked to distinguish the possible playing field for the cards to be placed. The participant was placed opposite Flobi. A microphone for speech recognition was placed next to the player.

The entire interaction was recorded using two high definition cameras s(see fig. 4.4). One camera took a frontal view of the scene. The second camera was directed at the setup from behind the participant with a focus on the robot. Additionally, one of the robot cameras, as well as the input of the microphone and the system logs (component data, events, tasks, etc.) were recorded for every session.



(a) Frontal camera view

(b) Camera filming from behind

Figure 4.4.: Views from the cameras used during the study. Image (a) shows the view from the frontal camera filming the participant. Image (b) shows a view of the camera placed behind the participant.

A session with the robot was divided into three parts. In the beginning of each session, the participants completed the first part of the questionnaire, rating their expectations of the upcoming interaction. After this, the participants interacted with the robot. When they finished, they completed the second part of the questionnaire.

The questionnaires were presented on a PC placed in the same room. While completing the questionnaires, the robot was active and focuses on detecting faces. The blinking behavior was also active. To prevent the robot from speaking and thereby disturbing the participants while they worked on the questionnaire in session one, the microphone was muted for conditions C1 and C2 until game-play began. For the remote controlled condition (C3) the microphone was only a placebo and not active at all.

Before the first interaction the examiner explained the coordination system used by the robot throughout the game. The explanation was constructed from phrases similar to those used for the dialog component to address cards on the playing field. While the examiner explained the interaction, the robot Flobi was focusing on faces. During the explanation, the microphone remained turned off. After the explanation, the examiner indicated to the participant how to initiate the interaction, how to restart the system and how to stop the interaction. The participant was advised to play a minimum of one round, but was allowed to play as many rounds as desired. When the participant acknowledged all the information and no questions were left, the examiner activated the microphone and left the room.

For the sessions two, three and four, the examiner activated the microphone before the first step. In these sessions, each participant was advised to start the interaction directly after filling out the first questionnaire.

4.2.1. Setup for the Remote Control Condition

For the third condition (C3), the state-engine, vision and speech recognition were controlled by a human examiner. The examiner was placed in a room next to the experimental laboratory (see fig. 4.5). The game was controlled using the remote control component. A web-cam placed on Flobi's neck allowed the examiner to see the cards on the field as they were manipulated by the participant. The image of the camera was mirrored to match the view point of the human player. The examiner used the same set of cards as the participants. Each card turned by the participant was placed on a corresponding field in front of the examiner. Cards once placed were not removed until the pair was taken from the table by the participant, or a new round began. This procedure allowed the examiner to mimic perfect memory.

The participants verbal communications with the robot were forwarded to the control room by using a direct telephone connection. This allowed the examiner to hear the participants questions and answers to the robot without any delay. The microphone on the examiner side was deactivated.

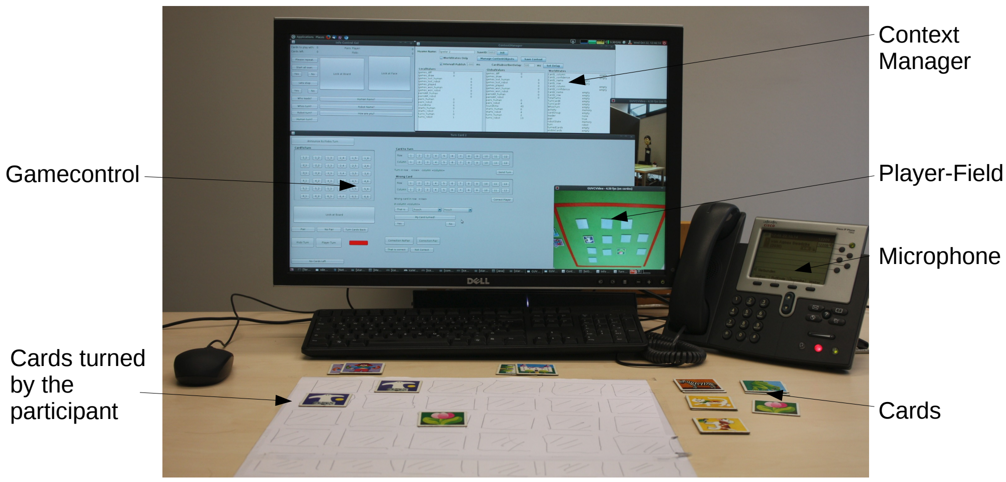


Figure 4.5.: The remote control setup. The controller can observe the field by using a web-cam and place identical cards in front of the display.

4.2.2. Questionnaire Design

For each session, a digital questionnaire was provided on a PC placed in the experimental room. The questionnaire for session one started with questions used to categorize the participants by gender, age and experience with intelligent systems.

The main part of each questionnaire was divided into two sub parts. The first part checked on the participants expectations upon the upcoming interaction. For the expectations 24 items were provided using a Likert-scale from 1 to 7. After this part the participants interacted with the robotic system. After the interaction was over the participants continued with the second part of the questionnaire. In this part the participants rated how they perceived the robot and the interaction while playing. The perception was rated by 24 items with a Likert-scale of 1 to 7. Additionally four questions concerning the technical performance of the system (e.g. if technical problems occurred) and three items on the reaction of the robot (e.g. was the reaction of the robot delayed) were rated. For each of these seven items a Likert-scale from 1 to 7 was used.

The questionnaires for session one and session four provided additional questions. These questions were designed to collect information on how engaged the participant felt throughout the interaction. The items were adapted from the items of the questionnaire provided by Leite et al. [27]. Used were the groups for engagement, social presence and the perceived support. The group for engagement was extended with ratings for the quality of the game-play (e.g. rating if the robot played fair). In total this part provided 37 items with a scale from 1 (strongly disagree) to 7 (strongly agree). Throughout all four sessions a total of 294 items were filled by each participant.

4.2.3. Indices for Questionnaire Items

To evaluate the effects of the interaction, several items rated in the questionnaires were grouped together, and presented by creating indices's. For each group, an index was created by computing the mean values. Each index was checked by evaluating its reliability using the Cronbach alpha coefficient. To show internal consistence, an index for a coefficient should be above the value of .700. In total eight indices's were created combining a total of 73 items (see table 4.1).

The first index (*Machine-Like vs. Human-Like (Pre)*) rated items on the participants expectations of their interaction with the robot (see table A.17, page 110). The second index (*Machine-Like vs. Human-Like (Post)*) rated items on the perception of the robots behavior after the interaction (see table A.17, page 110). These items rated whether the robot was perceived as more machine-like or more human-like. Each index combined ten items.

Index three (*Unlikable vs. Likable (Pre)*, see table A.19, page 112) and four (*Unlikable vs. Likable (Post)*, see table A.20, page 113) represent expected likability before the interaction, and perceived likability of the robot after the interaction. These indices's grouped fourteen items each.

Index number five (*Social Presence*) represented the social presence of the robots (see table A.21, page 113). This index combined six items about how the participant rated their insights on the robots feelings. This index also covered how the participant felt the robot would likely perceive the participant.

Index six (*Gaming Support*, see table A.22, page 114) and seven (*Social Support*, see table A.23, page 114) grouped items about how the robot's support was perceived. Index

Description	**Items**	**Reliability W1**	**Reliability W4**
Machine-Like vs. Human-Like (Pre)	10	.784	.933
Machine-Like vs. Human-Like (Post)	10	.925	.922
Unlikable vs. Likable (Pre)	14	.823	.940
Unlikable vs. Likable (Post)	14	.901	.934
Social Presence	6	.748	.774
Gaming Support	6	.693	.655
Social Support	8	.897	.874
Interaction Complexity	5	.902	.925

Table 4.1.: This table shows the indices created for the analysis. It includes the numbers of items per index and the computed reliability for session one and four. All indices are reliable with a Cronbach alpha coefficient of .700 or higher. Only the item *Gaming Support* showed a marginal reliability.

six combined six items and index seven combined eight items. One should remark that the result of the index for the item *Gaming Support* was below .700. This indicated a marginal reliability, nevertheless it was nearly reliable and will be used for further analysis.

Index eight (*Interaction Complexity*) grouped items concerning the complexity of the interaction (see table A.24, page 115). This index was computed using five items.

All proposed indices showed an acceptable value for the coefficient. Based on these findings, the following analysis is conducted using them.

4.2.4. Results for the Evaluation

For the study, a total of 39 participants (25 females and 14 males, age M = 26,54, SD = 5,684) were recruited. The study was conducted from June to October 2014. All participants were randomly assigned to the different conditions: C1 (9 females, 4 males), C2 (9 females, 5 males) and C3 (7 females, 5 males).

For all participants, an independent t-test was computed for ratings on their knowledge about robots, and intelligent systems (see table 4.2). No significant differences were found.

Table 4.2.: Results of the participants ratings for the items *Knowledge on Intelligent Systems* and *Knowledge on Robotic Systems.*

Item	**C1 M(SD)**	**C2 M(SD)**	**C3 M(SD)**	**C1 vs. C2**			**C2 vs. C3**		
				t	**df**	**Sig.**	**t**	**df**	**Sig.**
Knowledge on Robotics	2.540 (1.450)	1.860 (1.292)	1.830 (1.337)	-1,291	25	.209	0.046	25	.964
Knowledge on Intelligent Systems	3.540 (1.898)	3.140 (1.657)	2.500 (1.508)	-0.578	24	.568	1.027	24	.314

To evaluate possible effects between the proposed systems, a total of 21 items were analyzed, based on the ratings from the different questionnaires. These items rated 13 points on engagement, as well as the eight computed indices's (see section 4.2.3) involving the robots social behavior.

The different items analyzed were structured into the following groups:

- **Social Perception (5 Indices's)**
  Items concerning how the robotic system and its social capabilities were perceived.
- **Robotic Support (3 Indices's)**
  Ratings on how supportive the robot was perceived to be in case of gaming and social support.
- **Rating of Gaming (8 Items)**
  Items rating how the gaming nature of the interaction was perceived.
- **Ratings on Engagement (5 Items)**
  Ratings about the participant's perception of the robot as an interaction partner and vice versa.

The following sections will present the results of the elements structured by each group. To analyze the data of the interaction and to perform tests for each item the difference between the mean values of the first session (W1) and the mean values of the last session (W4) was computed. Based on the result of the difference an independent samples t-test between the conditions was conducted. The tests were performed between condition C1 and C2 and between condition C2 and C3.

A paired t-test was used to test for effects within each condition. This test was conducted using the mean values of session one (W1) compared to the mean values of session four (W4).

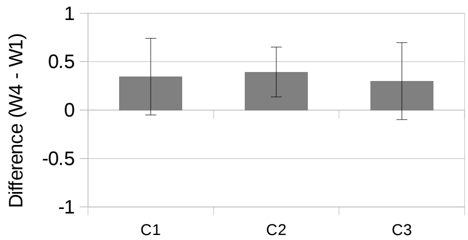
4.2.5. Social Perception

The group for social perception (see table A.25, page 116) consists of the following indices:

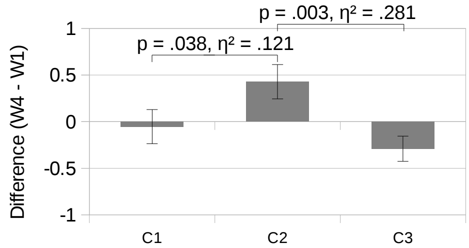
- Human-like vs. Machine-like (Pre)
- Human-like vs. Machine-like (Post)
- Unlikable vs. Likable (Pre)
- Unlikable vs. Likable (Post)
- Social Presence

Results for the Ratings between Conditions

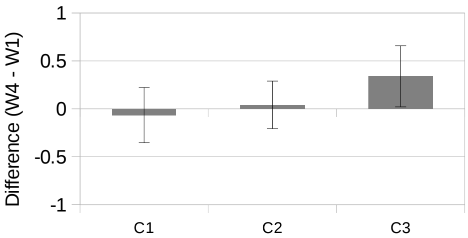
From the ratings a significant difference between condition C1 and C2 for the item *Human-Like vs. Machine-Like (Post)* could be found (see fig. 4.6 (b)). There was a significant difference between condition C2 and C3. For the item *Unlikable vs. Likable (Post)* (see fig. 4.6 (d)) there was also a marginally significant difference between condition C2 and C3.



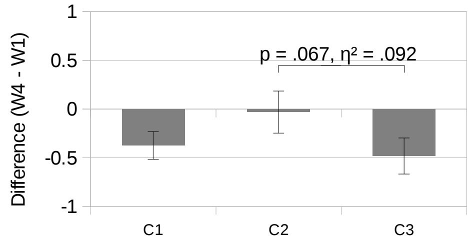
(a) Human-like vs. Machine-like (Pre)



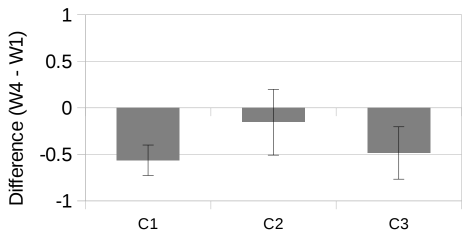
(b) Human-like vs. Machine-like (Post)



(c) Unlikable vs. Likable (Pre)



(d) Unlikable vs. Likable (Post)

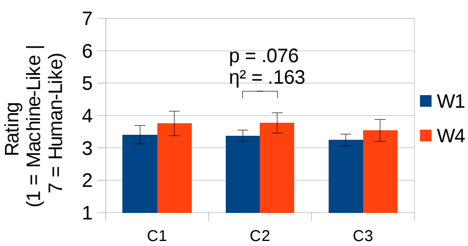


(e) Social Presence

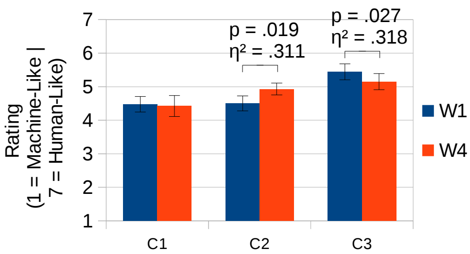
Figure 4.6.: Differences for the items in the group *Social Perception*.

Results for the Ratings within Conditions

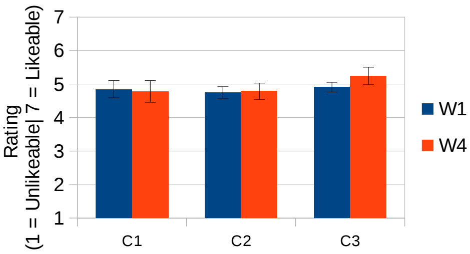
For the within effects of the different conditions the paired t-test showed a marginally significant increase for the item *Human-Like vs. Machine-Like (Pre)* (see fig. 4.7 (a)) in condition C2. The item *Human-Like vs Machine-Like (Post)* (see fig. 4.7 (b)) showed a significant increase for condition C2 and a significant decrease for C3. The item *Unlikable vs. Likable (Post)* (see fig. 4.7 (c)) showed a significant decrease for condition C1 and for condition C3. The item *Social Presence* ((see fig. 4.7 (e)) showed a significant decrease for condition C1 and a marginally significant decrease for C3.



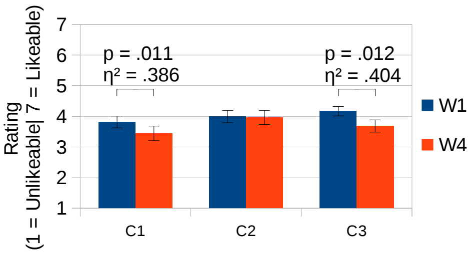
(a) Human-like vs. Machine-like (Pre)



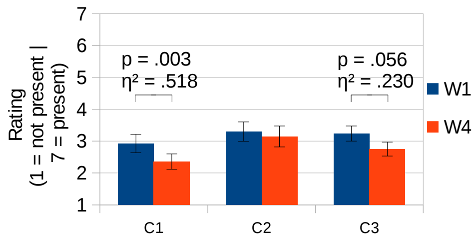
(b) Human-like vs. Machine-like (Post)



(c) Unlikable vs. Likable (Pre)



(d) Unlikable vs. Likable (Post)



(e) Social Presence

Figure 4.7.: Mean values for the items in the group *Social Perception*.

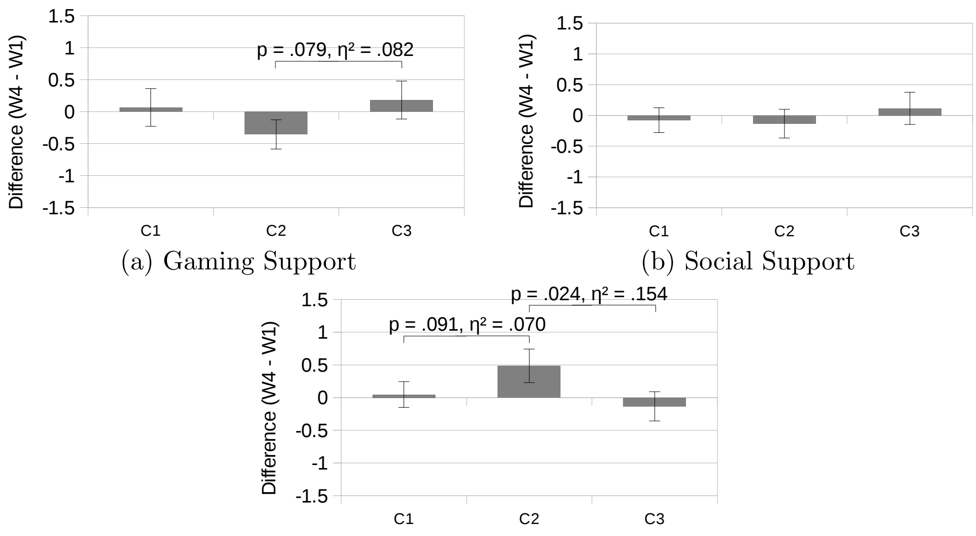
4.2.6. Robotic Support

The group for social support (see table A.26, page 117) consists of the following indices:

- Gaming Support
- Social Support
- Interaction Complexity

Results for the Ratings between Conditions

From the ratings for the item *Gaming Support*, a marginal effect between conditions C2 and C3 was found (see fig. 4.8 (a)). Based on the mean values for the index *Interaction Complexity*, the t-test showed a marginally significant difference between C1 and C2. There was also a significant difference between C2 and C3 (see fig. 4.8 (c)).



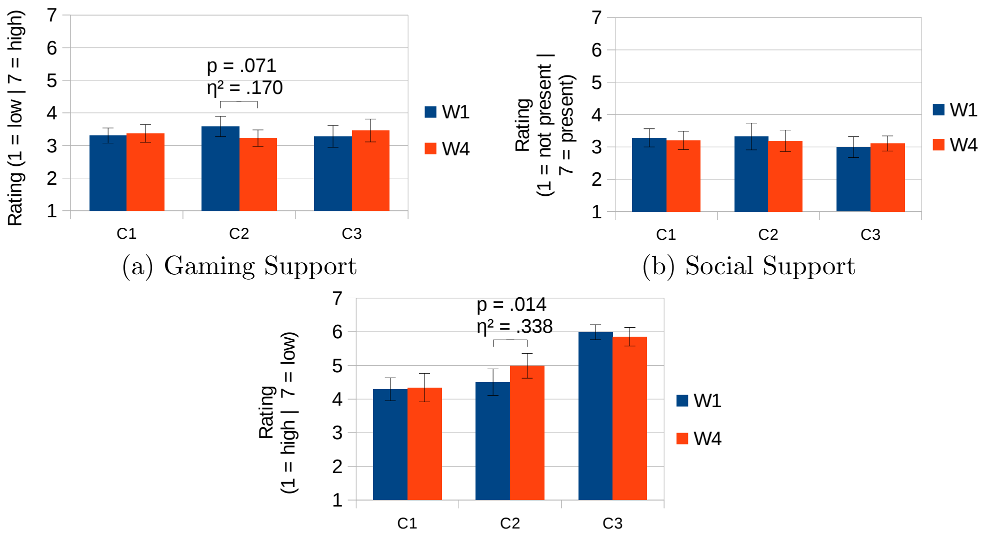
(a) Gaming Support (b) Social Support

(c) Interaction Complexity

Figure 4.8.: Differences for the items in the group *Robotic Support*.

Results for the Ratings within Conditions

For the within effects of the different conditions, the paired t-test between session one and four showed a marginally significant decrease for the item *Gaming Support* for condition C2 (see fig. 4.9 (a)). For the item *Interaction Complexity* there was a significant decrease towards a less complex interaction for condition C2 (see fig. 4.9 (c)).



(a) Gaming Support

(b) Social Support

(c) Interaction Complexity

Figure 4.9.: Mean values for the items in the group *Robotic Support*.

4.2.7. Ratings on Gaming

The rating items for the gaming features of the interaction (see table A.27, page 118) consist of the following:

- EN1 - I like to play pairs
- EN3 - I enjoyed playing with the robot
- EN6 - I wanted to beat robot
- EN7 - The Robot wanted to beat me
- EN8 - The Robot played fair
- EN9 - I played fair
- EN10 - The robot plays better than a human
- EN11 - I would like to play again with the robot

Results for the Ratings between Conditions

The ratings for *EN1* indicated a significant difference between the conditions C2 and C3 (see fig. 4.10 (a)). For the item *EN3* the t-test showed a marginally significant difference between conditions C1 and C2 (see fig. 4.10 (c)). The t-test for item *EN6* showed a marginally significant difference between conditions C1 and C2 and a significant difference between conditions C2 and C3 (see fig. 4.10 (d)). Based on the mean values for item *EN7*, the t-test showed a marginally significant difference between conditions C2 and C3 (see fig. 4.10 (e)). The t-test for item *EN9* showed a significant difference between conditions C1 and C2 and between conditions C2 and C3 (see fig. 4.10 (g)).

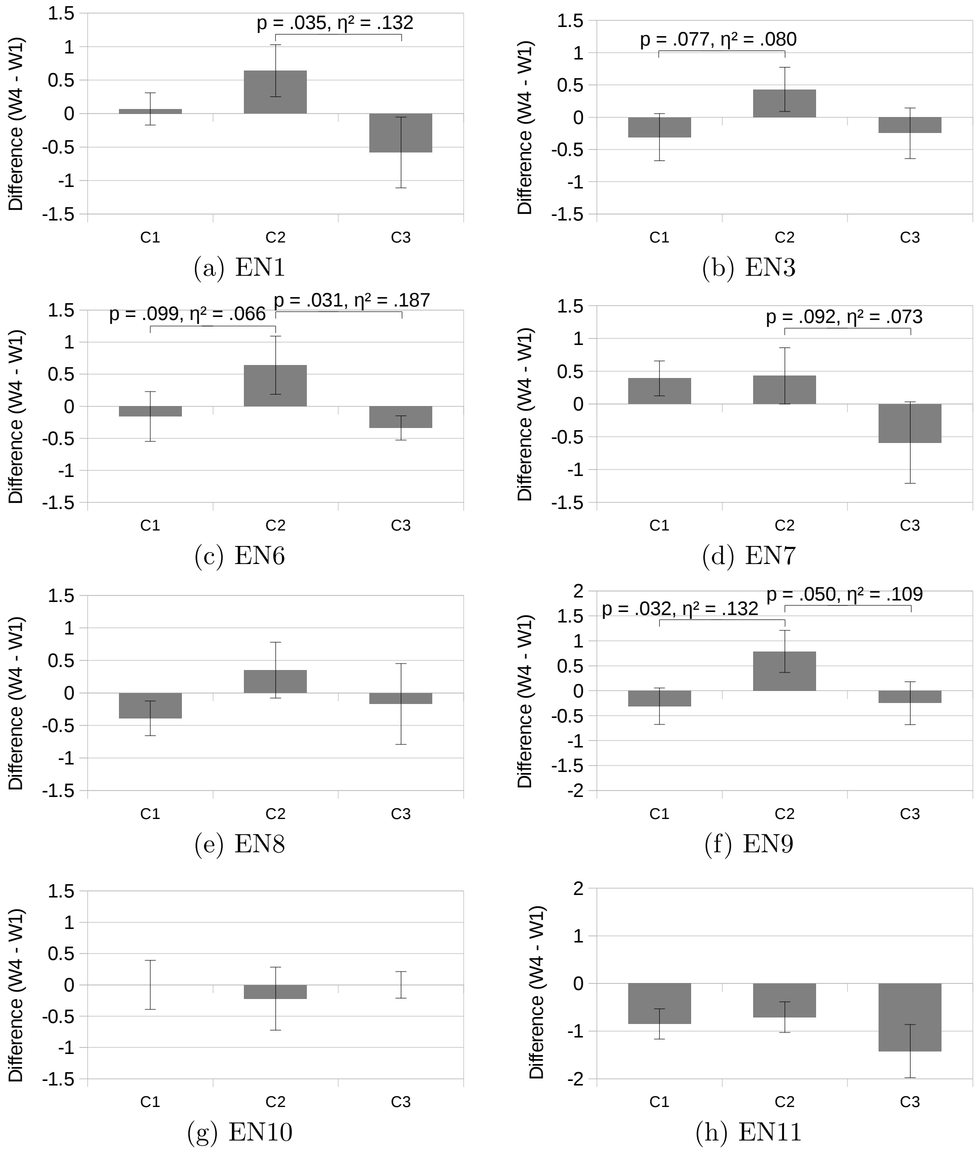
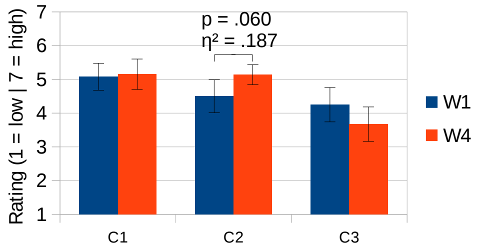


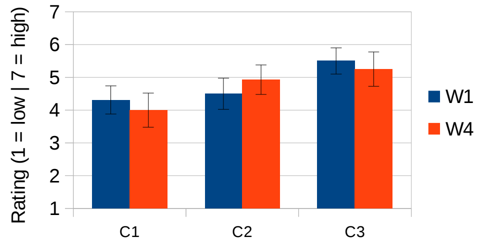
Figure 4.10.: Differences for the items in the group *Ratings on Gaming*.

Results for the Ratings within Conditions

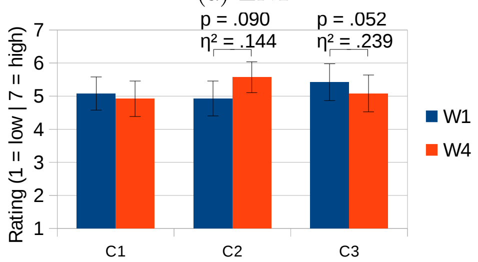
For the within effects of the different conditions, the paired t-test between session one and four showed a marginally significant increase for item *EN1* for condition C2 (see fig. 4.11 (a)). For item *EN6*, a marginally significant increase for condition C2, and a marginally significant decrease for condition C3 was found (see fig. 4.11 (d)). The item *EN7* showed a marginally significant increase for condition C1 (see fig. 4.11 (e)). The paired t-test for the item *EN9* showed a significant increase for condition C2 (see fig. 4.11 (g)). The paired t-test for the item *EN11* showed a significant decrease for conditions C1, C2 and C3 (see fig. 4.11 (i)).



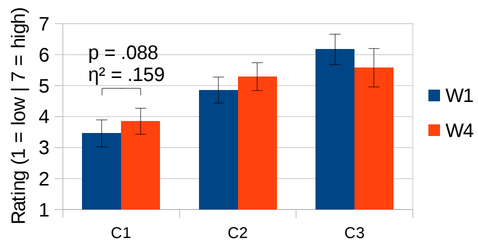
(a) EN1



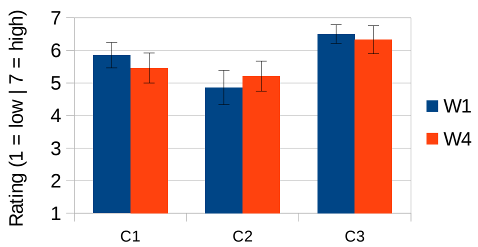
(b) EN3



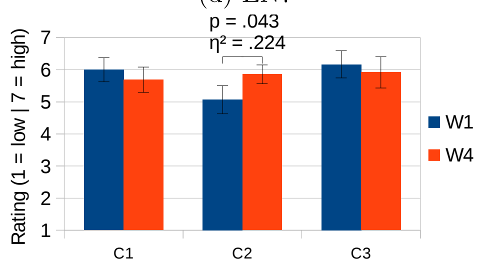
(c) EN6



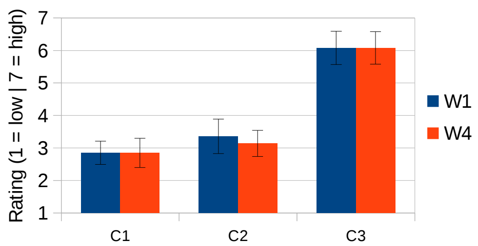
(d) EN7



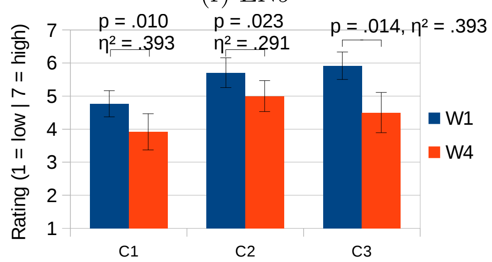
(e) EN8



(f) EN9



(g) EN10



(h) EN11

Figure 4.11.: Mean values for the items in the group *Ratings on Gaming*.

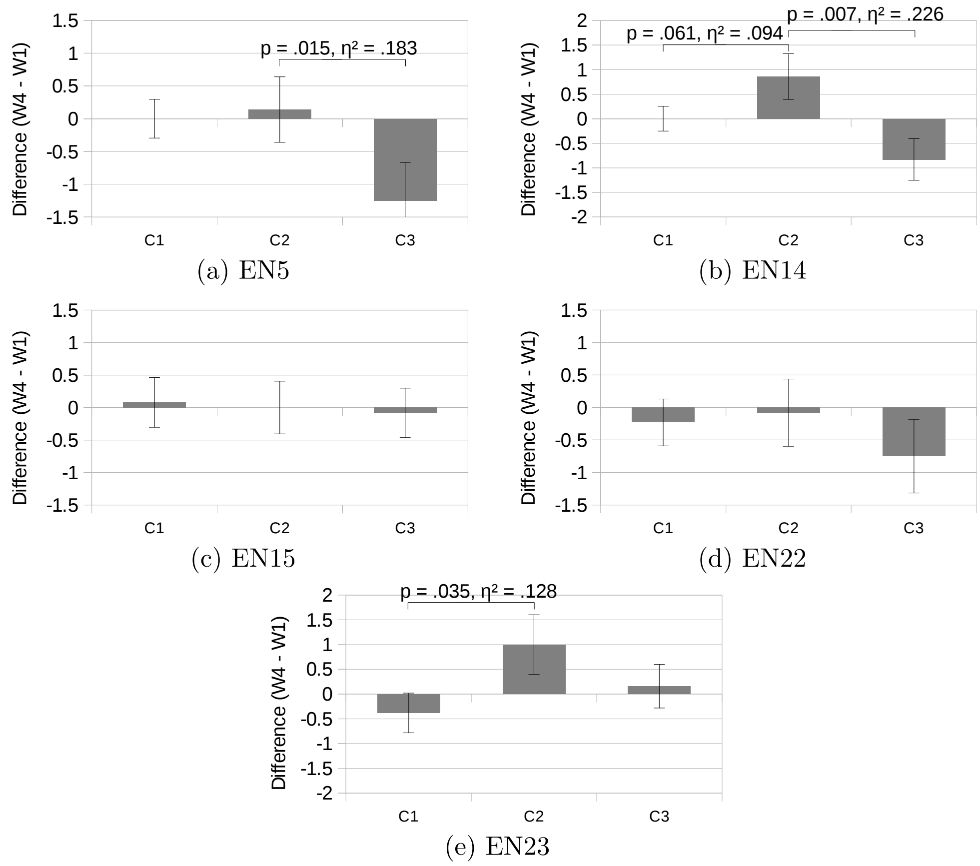
4.2.8. Ratings on Engagement

Ratings on engagement (see table A.28, page 119) consist of the following items:

- EN5 - I lost track of time while playing
- EN14 - I was concentrated while playing
- EN15 - The robot was concentrated while playing
- EN22 - The robot adapted its behavior to my behavior
- EN23 - I adapted my behavior to the robots behavior

Results for the Ratings between Conditions

The ratings for item *EN5* indicated a significant difference between conditions C2 and C3 (see fig. 4.12 (b)). Based on the mean values for item *EN14*, the t-test showed a marginally significant difference between conditions C1 and C2. There was also a significant effect between conditions C2 and C3 (see fig. 4.12 (e)). For item *EN23*, the t-test showed a significant difference between conditions C1 and C2 (see fig. 4.12 (h)).

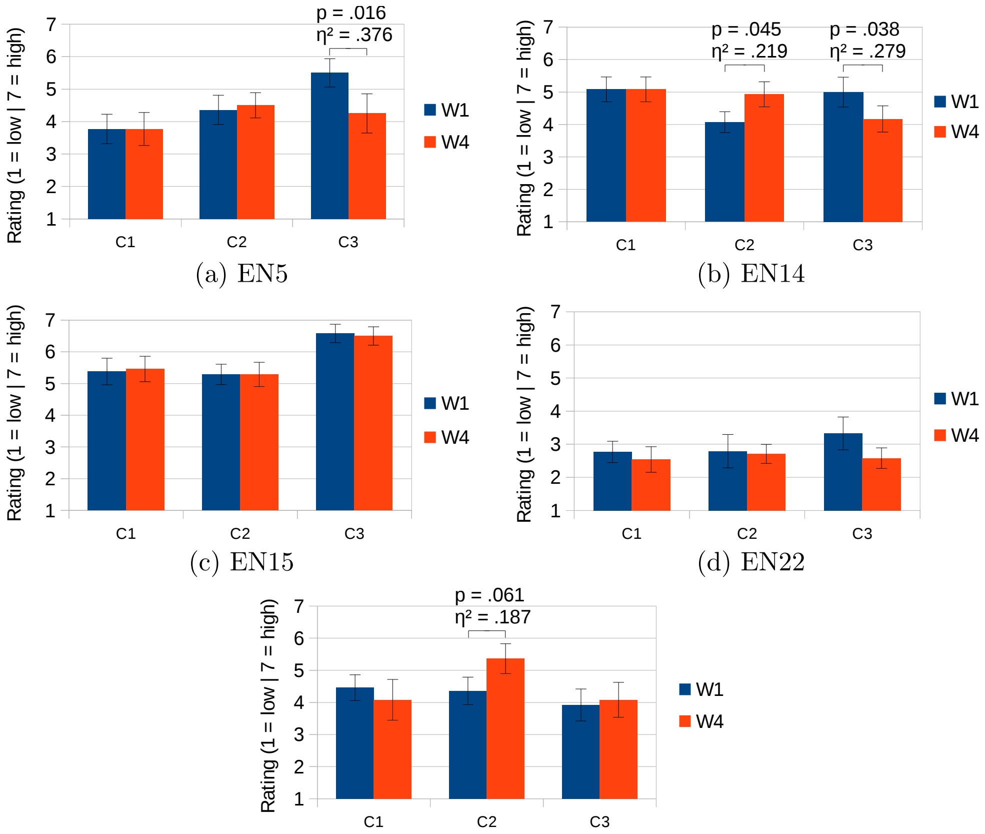


(a) EN5
(b) EN14
(c) EN15
(d) EN22
(e) EN23

Figure 4.12.: Differences for *Ratings on Engagement*.

Results for the Ratings within Conditions

In the within effects, the paired t-test between session one and four, item *EN5* indicated a significant decrease for condition C3 (see fig. 4.13 (b)). The paired t-test for item *EN14* showed a significant increase for condition C2, and a significant decrease for condition C3 (see fig. 4.13 (e)). The paired t-test showed a marginally significant increase in condition C2 for item *EN23* (see fig. 4.13 (h)).



(a) EN5 (b) EN14 (c) EN15 (d) EN22 (e) EN23

Figure 4.13.: Mean values for the items in the group *Ratings on Engagement*.

4.2.9. Video-Analysis of the Interaction

The recorded sessions were annotated by research assistants for a qualitative analysis of the conditions (see table A.29 (a), page 120). They coded the rounds and the time played per session. Additionally, the status of the round was coded as *completed*, *restarted* or *stopped*. *Completed* rounds were successfully started by the robot and played until all pairs were discovered. A round was coded as *restarted* when the player stopped the current round (e.g. due to system failures) and started a new one. A round was coded as *stopped* when the player entirely stopped the interaction for the session. Also, all occurrences of cheating behaviors by the participants were coded in the annotation. Examples off cheating behaviors included players peeking under the cards, or players exchanging cards when the robot was not looking.

In total, 156 sessions were annotated from a total of 80 GB of video data. The following items were analyzed from the annotated data: number of rounds played per session, time played per round, time for a complete session and occurrences of cheating.

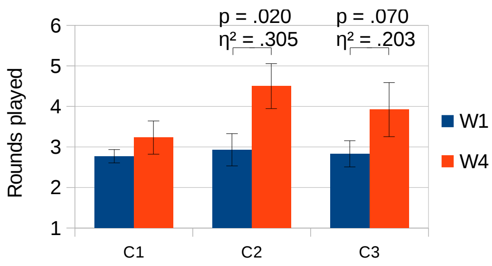
Some rounds played by the participants were restarted very early on. This was mostly due to technical issues or misunderstandings between robot and participant (e.g. not enough space between cards). Because these restarts occurred in the first minutes of the interaction, rounds that lasted less than 120 seconds were excluded from the analysis.

For each item, an independent t-test was computed between conditions C1 and C2, and between conditions C2 and C3 (see table A.29 (b), page 120). Tested was the difference between the mean values of session one and those of four. A paired t-test was used to evaluate within effects for each condition using results from the mean values of session one and those of four (see table A.29 (c), page 120).

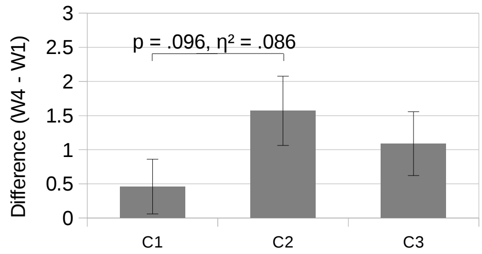
Rounds played per Session

Based on the mean values from *Rounds Played per Session*, the t-test showed a marginally significant difference between conditions C1 and C2 (see fig. 4.14 (b)) . There was no significant effect between conditions C2 and C3.

Conducted was a paired t-test for session one and session four, and found a significant increase in the amount of rounds played for conditions C2 and C3 (see fig. 4.14 (a)).



(a) Mean values for each condition



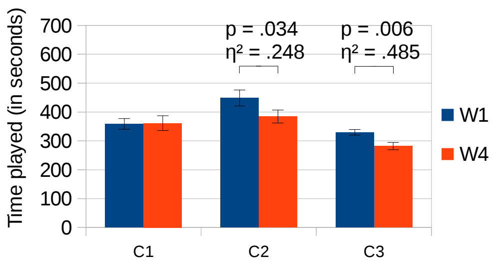
(b) Differences Between Session 1 and 4

Figure 4.14.: Mean values (a) and differences (b) for the *Rounds Played per Session*. There was a marginally significant effect between C1 and C2. C2 and C3 showed a significant effect between session 1 and 4.

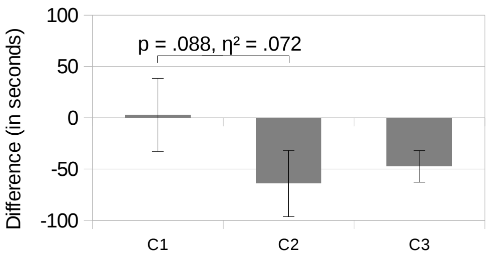
Mean Time for Rounds played per Session

Based on the mean values for the *Mean Time per Round Played per Session* the t-test showed a marginally significant difference between conditions C1 and C2 (see fig. 4.15 (b)).

A paired t-test for sessions one and four showed a significant decrease for conditions C2 and C3 (see fig. 4.15 (a)).



(a) Mean values for each condition



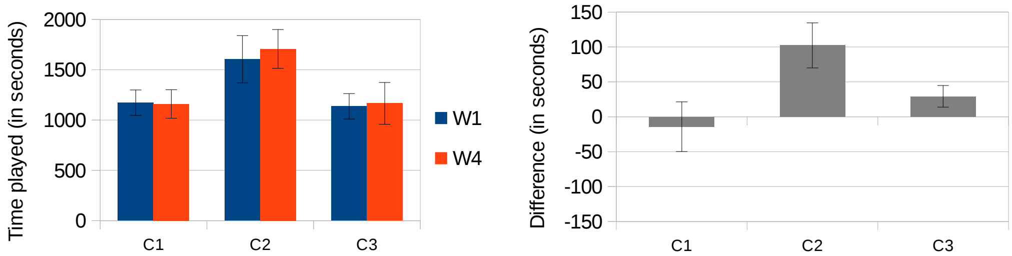
(b) Differences Between Session 1 and 4

Figure 4.15.: Mean values (a) and differences (b) for the *Mean Time per Round Played per Session*. There was a marginally significant effect between C1 and C2. C1 and C3 showed a significant effect between session 1 and 4.

Total Time played per Session

Based on the mean values for the *Total Time Played per Session* the t-test showed no significant difference between conditions C1 and C2 or between conditions C2 and C3 (see fig. 4.16 (b)).

A paired t-test for session one and four showed no significant difference within conditions C1, C2 or C3 (see fig. 4.16 (a)).



(a) Mean values for each condition

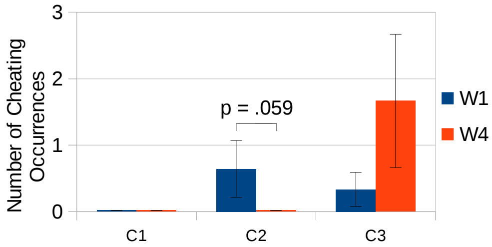
(b) Differences Between Session 1 and 4

Figure 4.16.: Mean values (a) and differences (b) for the *Total Time Played per Session*. There was no significant effect between or within conditions.

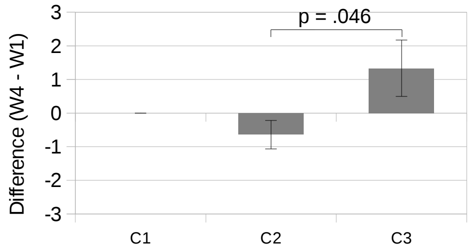
Occurrences of Cheating Behavior per Session

The data for the cheating behaviors did not show a normal distribution (mean values see table A.30, page 120). Therefore, a non-parametric test was used. Based on the mean values for the *Occurrences of Cheating Behaviors*, a Mann-Whitney Test showed a significant difference between C2 and C3 (see fig. 4.17 (b)).

The Wilcoxon Signed Rank Test between sessions one and four showed a marginally significant decrease for conditions C2 (see fig. 4.16 (a)).



(a) Mean values for each condition



(b) Differences Between Session 1 and 4

Figure 4.17.: Mean values (a) and differences (b) for the *Occurrences of Cheating Behaviors*. A Mann-Whitney test showed a significant difference between condition C2 and C3.

4.3. Discussion

This section discusses the findings of the study for the extended scenario. Each of the hypothesis's defined in the introduction (see section 1.2) will be examined.

H1 - Positive Ratings of a System using Context Knowledge

The first hypothesis proposed, that applying information about past interactions to the system would result in a more positive rating on the perception of the robot and the system. This rating would still be positive after repeated interactions.

For the study, the participants rated how human-like they found the robot to be, as well as how likable it was. The ratings indicating a machine-like or a human-like inclination showed nearly average values after the first session. For condition (C3), the ratings pointed towards human-like. Comparing the expectations for these items, the ratings increased for all conditions between session one and four. In all conditions the robot was perceived as more human-like compared to the expectations. Only condition C2 showed an increase in the ratings over time. Nevertheless, the ratings for condition C3 were still higher compared to condition C2. Comparing the ratings before and after each session, it appeared that the interactions promoted a more human-like perception. Using a system that behaves too perfectly, even if contextual information is used, results in a decrease over time in case of human-likeness. A system that plays well, but allows the human to keep up with the robot may be the best solution.

When examining the robot's likability, a decrease was found between the expectations and the perception for all conditions. The expectations before each session were higher than average and increased over time for the conditions providing context knowledge. Compared to perception, rated after each session, a decrease was found for all conditions. The largest decrease was found for the remote controlled system, and was marginally significantly higher compared to condition C2. Reasons for this decrease can be the greater performance of the controlled system, resulting in an opponent that is harder to win against. Because the robot did not forget cards and won more rounds compared to the autonomous system, the robot appears to promote dislike. In addition, reminding participants of how many rounds the robot leads by, can foster this dislike. Letting the human win a gaming interaction can help to positively influence the perception of the robot.

The results for the ratings on social presence decreased over time, showing that the robot was not truly perceived as socially present during all conditions. These ratings were consistent from the beginning to the end of the study. However the autonomous system using context knowledge (C2) showed no significant decrease between the first and the last sessions. It seems, the implemented interaction focused more on the gaming part instead of on social communication (e.g. speaking about how the participants feels throughout the game).

The data on how supported the participants felt throughout the interaction showed less than average ratings for gaming and social support. The ratings did not significantly change over time, showing that the robot was somehow perceived as supportive, however providing context knowledge had no impact on the ratings.

The data for the item interaction complexity were more interesting. The autonomous system used in condition C2 showed a significant increase towards less complex interactions over time. Data for this condition were significantly different from the other conditions. The participant had to learn throughout the study how to interact with the robot, resulting in better game-play. Providing context information seemed to promote the learning effect. Nevertheless condition C3 again showed ratings that moved towards less complex interactions compared to conditions C2 and C1, even though the ratings decreased through the end of the study. The reasons for these high ratings included the nearly non-existent problems of card perception and the correct handling of speech recognition.

The item EN1 rated how participants enjoyed playing the game of pairs in general. The findings indicate that the ratings significantly increased for condition C2 over time. For the autonomous system, contextual information did not influence the overall enjoyment of playing the game of pairs. Furthermore, a decrease was found for condition C3. There was also a significant difference between C2 and C3. In examining how the participants enjoyed playing the game with the robot (EN3), it was found that context knowledge promoted a more positive rating. The difference between conditions C1 and C2 was nearly significant, suggesting that the context system had a positive effect when playing with a robot. Even though the ratings for the remote controlled system (C3) were higher compared to the other conditions, a decrease was found over time, suggesting that a too perfect playing system might lead to less enjoyment when playing with a robot.

Based on all these findings, it appears that applying information about past interactions results in a more positive perception of the robot and the gaming interaction. Effects of novelty, resulting in a decrease in the ratings, were not found for an autonomous system providing context knowledge (C2). Interestingly, a perfect playing system with the same capacity to remember past interactions, resulted in a less positive rating on enjoyment. In contrast to a human player, the robot never forgets cards and will take every pair possible. This advantage leads to more rounds won by the robot, and also more pronouncements about how the robot plays better than the human. It seems that the advantage of memorizing cards, and announcing this drawback, negatively influenced the interaction.

Similarly to the system proposed by Kasap and Magnenat-Thalmann [20, 21] it was found that providing information about past interactions helped to keep the participant interested in further interactions. By interacting every time in the same ways, the participants interest on the interaction decreased as well as the ratings on the different items. This corresponds with Leite et al. [27], who found that a more static interaction can lead to a decrease in ratings. In the gaming study with the iCat robot, the proposed system varied the game-play by using different difficulties. This may help to keep the player interested in later interactions. In the study described throughout this chapter, the difficulty of the game was not altered. Nevertheless compared to the difficulty element, providing contextual information can lead to more interesting game-play, and therefore a more positive perception.

As such, the results support the given hypothesis. By using contextual knowledge, an entertaining interaction can be interesting even in later interactions. Providing information based on game statistics may lead to a more positive perception.

H2 - More Player commitment for a System with Context Knowledge

The second hypothesis proposed that providing contextual information to the player would result in more commitment to the gaming interaction and therefore to more games to be played.

From the findings on items covering whether the participant wanted to win against the robot (EN6), it was found that for condition C2 the ratings increased between sessions one and four. Conditions C1 and C3 showed a significant decrease compared to C2. Reasons for the decrease in condition C3 include the robot's perfect memory, which resulted in it winning many more games. It seems the participants surrendered after loosing several rounds and lost their interest in winning. In condition C2, the robot performed less than perfectly (48% losses in session one and 64% losses for session four). It seems the participants in this condition liked the possibly of outsmarting the robot and therefore wanted to win even more rounds. Providing information about how the participants performed seemed to promote the desire to win. A reason for the decrease in condition C1 could also be the imperfectly playing robot (56% losses for session one and 59% losses for session four) and the lack of a more dynamic dialog.

Of interest was the rating for item EN7, concerning whether the robot wanted to win against the human player. Here, it was found an increase for the autonomous system and decrease for the controlled system. Reasons for the increase in C1 and C2 could be that the players learned how to play with the robot over time. This coincides with finding from the SoziRob project. This learning effect may lead to a faster game-play and therefore more rounds played. Nevertheless, the ratings for condition C3 were higher due to more rounds won by the robot. Throughout the interaction, the robot asked if the player wanted to play again and to give the robot a chance to win in case the robot has lost before. This pronouncement may have challenged the participants to play more. For the decrease in condition C3, one possible reason could be that the participants did not play many rounds due to their more challenging opponent. Losing more rounds lead to fewer games played, and therefore fewer remarks made by the robot about whether the player wanted to catch up. This finding was supported by the ratings comparing the robot's to the human's performance (EN10). For conditions C1 and C2, the system was rated as less competent, and for condition C3 the perfect game-play was rated as very competent.

When looking at the rating for whether the participants wanted to play again with the robot (EN11), the results showed a significant decrease for all conditions, nevertheless condition C3 began with the highest rating in session one, and decreased the most until the end of session four. Here it seems the game lacked elements to keep the player interested, such as even more entertaining dialog about things not just concerning manipulating cards and announcing results.

Regarding how concentrated the participants were while interacting with the robot (EN14), and how concentrated the robot performs (EN15), conditions C2 showed that the rating increased for participant concentration. A reason for this may have been that the additional information, and therefore more dialog in the first sessions, distracted the player. In the later games, it was more clear to the player when the robot would announce information, and therefore they could focus on the game. For the perfect playing system in condition C3, the ratings decreased. A reason for this may have been the fast and direct reaction by the robot during the game. The robot communicated its intentions quickly compared to the other conditions, and therefore used his communication capabilities more frequently. This may have distracted some of the participants. For the first interaction in condition C3, it was found that more participants lost track of time compared to the other conditions (EN5). This shows that the perfect playing game was more interesting in the beginning resulting in more commitment.

Both conditions using the autonomous system (C1 and C2) required time for the participants to learn how to interact. In looking at the items rating whether the participants adapted their behavior to the robots behavior (EN23), higher ratings were found for condition C2. This was a significant difference compared to condition C1, even though the main structure of the game and the reaction times were the same. This suggests that providing contextual information may promote commitment and the interest of the player to play games without faults by adapting to the systems needs. Ratings on whether the robot adapted its behavior to that of the participants showed no effects. Changes in the robot's behaviors, such as those in the SoziRob system were not found.

The annotations for the rounds played per session showed that in the first session of all conditions the participants played nearly equal numbers of rounds. Through the end of the study, the participants played marginally significant more rounds in condition C2 compared to C1. There were also more rounds played for condition C2 compared to condition C3, nevertheless this result was not significant. These findings suggest that giving more feedback about statistics throughout the game may keep the player interested. The contextual information probably created a more competitive perception of the interaction, resulting in the participants desire to win more rounds than the robot.

For the time spent per round, it was found that the time played decreased for conditions C2 and C3. For condition C2, a reason for the decrease could be that the system itself spoke more in the beginning, resulting in longer periods for each dialog section. In later interactions, the participants learned how to interact with the robot, resulting in faster and shorter rounds. Interestingly here, even though learning was needed in condition C1, the time played per round was nearly constant for this condition. Further analysis of the video recordings is needed to better understand this issue. One reason for the decrease in round time for condition C3 could be that the participants learned how fast the robot could react, and therefore reacted faster themselves.

Analysis of the total time played per session found that in condition C2 the participants invested more time as a result of playing more rounds. This difference was not significant, nevertheless it suggests a trend towards more time spent in the condition providing contextual information.

These findings support the second hypothesis. By applying information about past interactions, the participants were more interested in the game and therefore played more rounds and invested more time. The participants adapted their behaviors, thus promoting faster game-play. The results for the perfect playing robot suggest that some participants did not like to play with a more difficult opponent.

H3 - Less Player Commitment for a Perfect Playing System

The third hypothesis posited that a perfect playing system would be perceived less positively, resulting in fewer games played in later interactions. Also such a system would be perceived more like a tool than a human-like partner.

The qualitative analysis of the rounds played per session found that the participants played nearly equal rounds in the first session for condition C2 and C3. In the last session there were fewer rounds played in condition C3 compared to C2, nevertheless this difference was not significant.

Data on the gaming nature of the interaction indicated that the participants liked to play the game of pairs in general more when interacting in condition C1. Condition C2 showed an increase that was nearly equal to the last session compared to C1. Providing contextual information helped to promote more enjoyment of the game. For condition C3, lower ratings and a significant decrease was found (EN1). The perfect playing opponent and its memory advantage, influenced how the participants enjoyed playing the game in a negative way. The results of the ratings for how the participants enjoyed playing with the robot showed a decrease for condition C3, as well. Nevertheless, these ratings were not significantly different, although higher compared to condition C2. The findings on enjoying the interaction with the robot were supported by the ratings on the likability of the robot. The ratings for condition C3 significantly decreased compared to condition C2. For social presence, a significant decrease in condition C3 can be found, however there was no significant difference compared to condition C2.

Items regarding whether the system was perceived as more machine-like than human-like showed that condition C3 was perceived as more human-like after the first session. After the last interaction, the ratings decreased for C3. For condition C1, the ratings significantly increased, however they were still lower compared to C3. Reasons for this could be that a system working not perfectly, given the case of a scenario where memorizing is important, becomes unpredictable. A system playing perfect without significant deficits becomes predictable and therefore interacts as expected. This promotes a more machine-like perception.

An interesting finding concerns the occurrence of cheating throughout the study. Different styles of cheating were found throughout the interactions. Some participants peeked under cards when the robot was not looking at the playing field. Others switched cards when turning cards back, resulting in incorrect knowledge about pair positions by the robot. Also used was the slow reaction time of the robot. For the current implementation, the robot needs to confirm that cards are turned back. Some participants played too fast, turning cards immediately after handling the robots cards. The robot had to see all cards covered and therefore asked the participants to turn cards back. Some participants used this to select different cards after they turned cards back. The quantity of cheating occurrences was coded using video data. The results it indicated that for condition C3, significantly more participants cheated compared to condition C2. For condition C2, only some participants cheated in the first session and no cheating was found in the last sessions. One reason for the greater amount of cheating could be the perception of the perfect playing system. Due to its nearly perfect memory, no pairs were forgotten. Some participants seemed to intent to disrupt this advantage. Another reason cheating may not have occurred often in condition C2 is the systems performance. In the later interactions it seemed that participants tried to keep the interaction alive in the case of problems. Focusing on this, and therefore the need to correctly handle cards resulted in reduced concentration on the game and who was winning. The ratings for whether the robot was perceived as playing fair (EN8) or if the participant played fair (EN9) showed that the robot was rated to be very fair, even in the remote controlled condition with the perfect playing robot. No participant seems to criticize its perfect memory and therefore the systems advantage. For both conditions involving fairness, results showed that the participants rated themselves as playing very fair. The results confirm that the participants in condition C2 increased in fairness, although for condition C3 a decrease was found. The difference between both conditions was significant. Even though the participants rated themselves as more fair, it was found that a remote controlled and therefore more competent system promoted more unfair behavior to cope with the disadvantages.

These findings support the third hypothesis. For interactions in a gaming context, a perfectly playing robot receives less positive ratings compared to an autonomous system. For an interaction with the goal of memorizing objects, a system that allows a human player to win because of less than perfect perception, is preferred. Humans tend to perceive a perfect playing system less positively, and to decrease the amount of interactions over time. Nevertheless, an accurate system was preferred over a system that makes mistakes and therefore consumes more time per round.

Conclusions for the use of Contextual Knowledge

The results in this chapter found that interactions using a more dynamic dialog, that utilizes knowledge about past interactions, was perceived more positively compared to a system played in a static manner. Context knowledge also helped to enhance the social perception of the robotic system. A system providing feedback based on how the interaction partner performed, was perceived to be more interesting even in later interactions. Also, the information could be used to steer the interaction, for instance to challenge the participants in case of an advantage by the robotic system.

For the findings on a remotely controlled and therefore perfect playing interaction partner, it could be argued, that errors did not necessarily negatively influence the interaction. A system that was not perfect was rated as more likable, although errors should not occur most of the time. This might lead to annoying situations, and negative ratings by the participants.

A positive aspect of a remote controlled system is the possibility of observing unusual behaviors in the participants. Here, the participants started to play unfairly in an attempt to catch up with their stronger opponent. Humans seemed to dislike being disadvantaged while playing with a robot, and some felt challenged to compensate for this by playing unfairly.

4.4. Summary

This chapter described the extensions made to the gaming interaction. The goal of the different extensions was to keep the human player interested even during later interaction. The extensions applied the capability to use gaming information about past interactions and to utilize them in reoccurring interactions. The first part of the chapter described the technical realization of the extensions (see section 4.1). Described were the new elements implemented, as well as the alterations to existing components. In section 4.2, a study was described that evaluates effects between a system providing knowledge about past events and a system without such information. Also evaluated was a perfect playing system in contrast to an autonomous system. Presented was data from the study based on questionnaires and the qualitative analysis. The last section of the chapter discussed the findings in relation to the hypothesis stated in the introduction (see section 4.3). Chapter 5 will provide a conclusion to the system described in the previous chapters, and the different studies conducted throughout this thesis. The final section will present an outlook on further research.

5. Conclusion

The introduction discussed difficulties experienced by humans when living and acting alone or in isolation (see chapter 1). In these conditions, without social support, effects such as boredom and depression can become more pronounced, adversely affecting health.

The field of social robotics investigates how robots can be employed by people in a wide variety of social situations. Several approaches have been used to implement these systems for long-term interaction. An autonomous robot can become a partner, supporting people in their daily tasks, while also creating a link to the outside world. In space missions, where support by humans is limited, communication with other individuals is often delayed and sparse. Using social robots can provide direct feedback and additional interaction opportunities.

Nevertheless, current robotic systems often lack the ability to remain interesting over longer periods of time. The literature has shown that approaches are often limited by the environment where they are applied. In addition, the unpredictability of human beings makes it difficult to understand every situation and its corresponding proper reaction.

As a contribution to the field of social robotics, this thesis describes a system that implemented and evaluated a robot as game playing partner intended for long-term interaction. The system was designed to allow for autonomous interaction enjoyable by humans and usable over longer periods of time. The system was expected to help cope with the effects of isolation, such as boredom and tedium. The system was also expected to provide cognitive benefits by challenging the human player, resulting in greater commitment throughout repeated interactions. To accomplish these goals, the system needed to be robust to failures, to break downs in communication between the interacting parties. Additionally, the system needed to be remotely maintainable, allowing for fast reactions upon system failures, thus ensuring long-term systems functionality.

The system described throughout this thesis was designed for one to one game-play with the robotic head Flobi. The selected game was the children game of pairs, where two players search for pairs in a spread of different cards, and try to outsmart each other. Due to the lack of manipulators, the robot provided dialog for communication with the player, and structured the interaction with a variety of verbal announcements and requests. Chapter 3 presented the structure of the interaction and the system with the different implemented components required to realize the interaction. The final system was capable of interacting autonomously with a human partner, addressing elements by using a dialog, and using visual elements to detect and classify cards.

The system was successfully integrated into the daily work routine of a long-term isolation study, and was conducted in cooperation with the German Aerospace Center. Throughout the study, the robot Flobi interacted in several sessions, playing many rounds per participant throughout the whole study. When system failures arose, corrections were performed remotely by examiners, thus allowing interactions to continue. In total, the robot interacted for more then 90 hours and played more then 100 games. One outcome of the study was the large quantity of recordings intended for further analysis of human robot interactions.

From the different ratings made by the participants before, throughout and after the study, as well as through interviews conducted at the end, it was found that in general, the idea of providing a system for gaming interactions was well accepted. Each participant understood how to play with the robot, and that by adapting to the system's behavior more games could be played through the end of the study. Data from the interviews supported the assumption that the game was interesting and could help to cope with boredom and tedium. Nevertheless, participants complained that the system's performance while classifying cards was error-prone. These problems interfered with several games, resulting in the interactions being perceived largely as frustrating. Also remarked was the faulty speech recognition, further increasing frustration due to misunderstandings, and therefore repeated and unnecessary dialog. That said, regardless of the system's problems, the participants remarked that the system learned over time, becoming gradually more competent. This finding was interesting given that the system itself did not change over time. It seems the participants considered their adaption to the system's behavior, resulting in a more fluid interaction, to be a learning process undertaken by the system. Another point, was the static structure of the interaction, with some participants suggesting that it needed improvement. Providing the same dialog over and over again leads to predictable behaviors, and therefore less interesting game-play in the later part of the study.

The results and findings from the long-term isolation study showed that a gaming interaction can help to cope with the effects of isolation. Regardless, the system was judged as needing to be more dynamic when communicating with the human player on more than one occasion.

The second part focused on possibilities to enhance the system based on findings from the long-term isolation study (see section 3.3). The current implementation was perceived as less positive due to the static structure of the dialog. A more dynamic dialog would be preferred, and could possibly lead to more interesting interactions. It can be assumed that more versatile dialog could be produced by providing feedback about past interactions. This would help to keep later games interesting and would also create a more comparative perception of the system.

Another point, was the error-proneness of the proposed system, resulting in fewer games played, and a greater frustration biased rating of the interaction. To evaluate whether a system that played more perfectly resulted in more positive ratings, a remote controlled version of the interaction was designed. By remotely controlling the system, a robot companion could be created that mimicked an opponent that played perfectly.

The first part of chapter 4 presented the components enhanced and added for the extension. The different components collected, stored and distributed information on games played during previous sessions. Also described was the remote control component that allows a human examiner to control the interaction by substituting vision and speech recognition with a controllable GUI.

The extensions were tested through a study at Bielefeld University. The first part evaluated effects of a system without knowledge of past interactions, as compared to a system that did provide such information. For this part, both systems played completely autonomously against the human player. The second part of the study compared the autonomous system with the remotely controlled system. For this condition, both systems utilized the information collected by the context knowledge components. Throughout the last condition, the participants were not aware that the system was being controlled in the background.

In the introduction, three hypothesis's were given (see section 1.2). The first hypothesis proposed that providing context information from past interactions resulted in a more positive perception of the system. The findings support this hypothesis. The rating showed that participants preferred the system with additional feedback. The system was rated as more human-like and also more likable. In addition, the gaming nature of the system providing context information was rated more positively.

The second hypothesis proposed that a context system resulted in greater commitment by the player in later interactions. This hypothesis was supported by the different ratings, and as shown by more games being played in the system using contextual knowledge. Participants also seemed to adapt better to this system compared to the one without this information.

The third hypothesis proposed that a system that played too perfectly, would be less positively rated, and therefore not preferred by the participants. Out findings support this claim. A perfectly playing robot was rated less positively due to its memory advantages. Despite having high initial ratings, rating for the condition with the remote controlled system decreased through the last session. For this condition it was also found that participants tried to cope with the robot's advantage by playing unfairly. On several occasions, cheating behaviors were found that allowed the human player to win against the robot. It seems that humans tend to reduce their disadvantages by cheating during competitive interactions with a robot.

Summarizing, one can argue that selecting gaming situations for human-robot interactions promotes the positive perception of the interaction. Also, a gaming interaction is often more familiar and natural, compared to an artificially created two-participant task. Providing a gaming system for long-term interactions may help to keep the interaction interesting, even when interactions take place repeatedly over time. Providing feedback about past interactions appears to help promote a more positive perception of the system. Additionally, it seems to create more dynamic interactions. Using information about interactions that occurred in the past, seemed to lead to more informative interactions, and appeared to keep the player more committed to the game. It also seemed to motivate the human player to perform better and beat the robot. Nevertheless, the system must perform in a competent manner, and be able to cope with problems that occur throughout the interaction. This also promotes the perception of a believable and reliable interaction partner. On the other hand, providing a too perfect acting system appears to lead to a less positive perception, because of its advantages compared human capabilities. In a gaming interaction between a human and a robot, the robot should not dominate over the human. Instead, it be designed to be a more equally capable companion.

6. Perspective

The system presented throughout this thesis offers a basis for further research in the context of gaming situations for long-term human-robot interaction. Throughout the studies conducted, a very large data set was acquired, representing different phases of the interaction. As a first step towards more insight on the data, it must first be annotated. These annotations should cover the verbal dialog between the human players and the robot. This can then be used for further research on how participants address the robot and what information needs to be understood and exchanged for gaming interactions.

Additionally, the behaviors of the participants should be annotated for body language and gaze changes over time. These findings may shed light on if and when boredom occurs, possibly resulting in reduced concentration on the interaction itself. From a gaming perspective, it will be interesting to better understand how the participants adapted to the robots needs. As noted before, the participants attributed a learning process to the robot. Analyzing when this adaptation occurs may help clarify issues in how participants understand the robot's behaviors.

In terms of the technical aspects of the interaction, several elements can be addressed and enhanced to make the system more robust. The current implementation sometimes misclassified cards whenever the humans hands are still over cards during manipulation. To make the system more robust detection of manipulation using 3D information can help improve reactions. Also important would be the autonomous detection and classification of the corresponding contextual information with relation to a human player.

In the current implementation, the capabilities of the robot platform were used in a simple manner, by looking at the board or the player, and using facial configurations to display emotions. Emotional feedback could be improved based on contextual knowledge. A more fluid change between the different emotions could help to promote a more human-like perception of the robot. The current emotional control component is designed to be rule based, and to change on the basis of the value for pleasure. The system developed in a previous study [24] found that an emotional model could be combined with the facial configurations of the robot head flobi. Based on this approach and game statistics, a more fluid emotional feedback can be implemented.

The presented system focused on one type of game and currently does not offer a more general setting to provide other types of games and interactions. For further research, it would be helpful to offer different games to the opponent, letting the player choose. This could also allow for testing different settings, such as reaction based games, more logical structured games or even games where the parties must cooperate.

Regarding the findings on a perfectly playing robot, it would also be of interest to apply different stages of difficulty to the robot's behaviors and to then analyze how such systems are perceived.

Finally, it should be mentioned that both studies presented here provided a fixed interaction on predefined occasions. Each participant acted voluntarily, but was inwardly driven by the monetary reward. To facilitate more voluntary, unscheduled interactions, the system should be placed in an open space, accessible at will by passing persons. This could help to evaluate whether the interaction was truly interesting, especially in the case of long-term interactions. To achieve this, the system must be made more robust to errors, and offer safe mechanisms that can be reinitialized by naive users.

Bibliography

[1] Becker-Asano, C. and Meneses, E. (2014). The hybrid agent MARCO: a multimodal autonomous robotic chess opponent. *Proceedings of the 2nd Intl. Conf. on Human-Agent Interaction*, pages 173–176. 2.3

[2] Berger, I., Kipp, A., Lütkebohle, I., Riether, N., Schneider, S., Süssenbach, L., and Kummert, F. (2012). Social robots for long-term space missions. In *63rd International Atronautical Congress*. 1.1

[3] Berkman, L. and Syme, S. (1979). Social networks, host resistance, and mortality: a nine-year follow-up study of Alameda County residents. *American journal of Epidemiology*, 109(2):186–204. 1

[4] Brom, C., Lukavský, J., and Kadlec, R. (2010). Episodic Memory For Human-Like Agents And Human-Like Agents For Episodic Memory. *International Journal of Machine Consciousness*, 02(02):227–244. 2.2

[5] Canny, J. (1986). A computational approach to edge detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, (6):679–698. 3.2.3

[6] Castellano, G., Aylett, R., Dautenhahn, K., Paiva, A., McOwan, P. W., and Ho, S. (2008). Long-term affect sensitive and socially interactive companions. In *Proceedings of the 4th International Workshop on Human-Computer Conversation*. Citeseer. 1.2

[7] Cattan, M., Newell, C., Bond, J., and White, M. (2003). Alleviating Social Isolation and Loneliness among Older People. *International Journal of Mental Health Promotion*, 5(3):20–30. 1

[8] Ceci, S. J. and Bruck, M. (1993). Suggestibility of the child witness: a historical review and synthesis. *Psychological bulletin*, 113(3):403. 2.1

[9] Dodd, W. and Gutierrez, R. (2005). The role of episodic memory and emotion in a cognitive robot. In *IEEE International Workshop on Robots and Human Interactive Communication*, pages 692–697. Ieee. 2.2

[10] Forlizzi, J. (2007). How robotic products become social products: an ethnographic study of cleaning in the home. *Proceedings of the ACM/IEEE international conference . . .*, pages 129–136. 2

[11] Ganster, T., Eimler, S. C. S., von der Pütten, A. M., Hoffmann, L., Krämer, N. C., and von der Pütten, A. (2010). *Methodological considerations for long-term experience with robots and agents*. na. 2

[12] Gockley, R. and Bruce, A. (2005). Designing robots for long-term social interaction. *International Conference on Intelligent Robots and Systems*, pages 2199–2204. 1.2

[13] Ho, W. C., Dautenhahn, K., Burke, N., Saunders, J., and Saez-Pons, J. (2013). Episodic memory visualization in robot companions providing a memory prosthesis for elderly users. In Encarnacao, P., Azevedo, L., Gelderblom, G. J., Newell, A., and Mathiassen, N.-E., editors, *Assistive Technology*, volume 33, pages 120–125. IOS Press. 2.2

[14] Ho, W. C., Dautenhahn, K., Lim, M. Y., Vargas, P. a., Aylett, R., and Enz, S. (2009). An initial memory model for virtual and robot companions supporting migration and long-term interaction. In *Robot and Human Interactive Communication, 2009. RO-MAN 2009. The 18th IEEE International Symposium on*, pages 277–284. Ieee. 1.2, 2.2

[15] Janssen, J. J., van der Wal, C., Neerincx, M., Looije, R., and van der Wal, C. (2011). Motivating Children to Learn Arithmetic with an Adaptive Robot Game. In Mutlu, B., Bartneck, C., Ham, J., Evers, V., and Kanda, T., editors, *Social Robotics*, volume 7072 of *Lecture Notes in Computer Science*, pages 153–162. Springer Berlin Heidelberg. 2.3

[16] Jost, C., Grandgeorge, M., Le Pévédic, B., and Duhaut, D. (2043). Robot or tablet: users' behaviors on a memory game. In *The 23rd IEEE International Symposium on Robot and Human Interactive*. 2.3

[17] Kahn, P. H., Friedman, B., Perez-Granados, D. R., and Freier, N. G. (2004). Robotic pets in the lives of preschool children. *Extended abstracts of the 2004 conference on Human factors and computing systems - CHI '04*, page 1449. 2

[18] Kanas, N. and Manzey, D. (2008). *Space Psychology and Psychiatry (Space Technology Library)*. Netherlands: Springer. 240 p. 1

[19] Kanda, T., Hirano, T., Eaton, D., and Ishiguro, H. (2004). Interactive Robots As Social Partners and Peer Tutors for Children: A Field Trial. *Hum.-Comput. Interact.*, 19(1):61–84. 1.2, 2.1

[20] Kasap, Z. and Magnenat-Thalmann, N. (2010). Towards episodic memory-based long-term affective interaction with a human-like robot. In *RO-MAN, 2010 IEEE*, pages 452–457. Ieee. 2.2, 4.3

[21] Kasap, Z. and Magnenat-Thalmann, N. (2012). Building long-term relationships with virtual and robotic characters: the role of remembering. *The Visual Computer*, 28(1):87–97. 1.2, 2.2, 4.3

[22] Kidd, C. D. and Breazeal, C. (2008). Robots at home: Understanding long-term human-robot interaction. In *Intelligent Robots and Systems, 2008. IROS 2008. IEEE/RSJ International Conference on*, pages 3230–3235. Ieee. 2.1

[23] Kim, M. M.-G. and Suzuki, K. (2011). Analysis of Bluffing Behavior in Human-Humanoid Poker Game. In Mutlu, B., Bartneck, C., Ham, J., Evers, V., and Kanda, T., editors, *Social Robotics*, volume 7072 of *Lecture Notes in Computer Science*, pages 183–192. Springer Berlin Heidelberg. 2.3

[24] Kipp, A., Carlmeyer, B., and Damm, O. (2014). Demonstration of emotion modeling using Flobi. *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, page 3391. 6

[25] Kipp, A. and Kummert, F. (2014). Dynamic dialog system for human robot collaboration: playing a game of pairs. *Proceedings of the second international conference on Human-agent interaction*, pages 225–228. 3.2.4

[26] Koay, K. L., Syrdal, D. S., Walters, M. L., Dautenhahn, K., and Lee Koay, K. (2007). Living with Robots: Investigating the Habituation Effect in Participants' Preferences During a Longitudinal Human-Robot Interaction Study. In *Robot and Human interactive Communication, 2007. RO-MAN 2007. The 16th IEEE International Symposium on*, pages 564–569. Ieee. 2

[27] Leite, I., Castellano, G., Pereira, A., Martinho, C., and Paiva, A. (2014). Empathic Robots for Long-term Interaction. *International Journal of Social Robotics*, pages 1–13. 2.1, 4.2.2, 4.3

[28] Leite, I., Martinho, C., and Paiva, A. (2013). Social Robots for Long-Term Interaction: A Survey. *International Journal of Social Robotics*, 5(2):291–308. 2.1, 2.2

[29] Lim, M., Aylett, R., Ho, W., Enz, S., and Vargas, P. (2009). A Socially-Aware Memory for Companion Agents. In Ruttkay, Z., Kipp, M., Nijholt, A., and Vilhjálmsson, H., editors, *Intelligent Virtual Agents*, volume 5773 of *Lecture Notes in Computer Science*, pages 20–26. Springer Berlin Heidelberg. 2.2

[30] Louie, W. G., McColl, D., and Nejat, G. (2012). Playing a memory game with a socially assistive robot: A case study at a long-term care facility. In *RO-MAN, 2012 IEEE*, pages 345–350. 2.3

[31] Lütkebohle, I. (2011). *Coordination and composition patterns in the "Curious Robot" scenario.* PhD thesis. 3.2.1

[32] Lutkebohle, I. and Hegel, F. (2010). The bielefeld anthropomorphic robot head "Flobi". In *2010 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3384–3391. 1.1

[33] Moses, Robert A and Adler, F. H. (1975). *Adler's Physiology of the Eye: Clinical Application*. Mosby. 4.1.4

[34] Naseem, I., Togneri, R., and Bennamoun, M. (2010). Linear regression for face recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(11):2106–2112. 3.2.3

[35] Nomura, T., Kanda, T., Suzuki, T., and Kato, K. (2004). Psychology in human-robot communication: an attempt through investigation of negative attitudes and anxiety toward robots. *RO-MAN 2004. 13th IEEE International Workshop on Robot and Human Interactive Communication (IEEE Catalog No.04TH8759)*, pages 35–40. 3.3.2

[36] Peltason, J. and Wrede, B. (2010). Pamini: A Framework for Assembling Mixed-initiative Human-robot Interaction from Generic Interaction Patterns. In *Proceedings of the 11th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, SIGDIAL '10, pages 229–232, Stroudsburg, PA, USA. Association for Computational Linguistics. 3.2.2

[37] Salem, M., Eyssel, F., Rohlfing, K., Kopp, S., and Joublin, F. (2013). To Err is Human(-like): Effects of Robot Gesture on Perceived Anthropomorphism and Likability. *International Journal of Social Robotics*, 5(3):313–323. 1.2

[38] Short, E., Hart, J., Vu, M., and Scassellati, B. (2010). No fair!! An interaction with a cheating robot. In *Human-Robot Interaction (HRI), 2010 5th ACM/IEEE International Conference on*, pages 219–226. Ieee. 2.3

[39] Sobel, I. and Feldman, G. (1968). A 3x3 isotropic gradient operator for image processing. *a talk at the Stanford Artificial Project in*, pages 271–272. 3.2.3

[40] Tulving, E. (2002). EPISODIC MEMORY: From Mind to Brain. *Annual Review of Psychology*, 53(1):1–25. 2.2

[41] Tulving, E. and Donaldson, W. (1972). Episodic and semantic memory. *Organization of memory*. 2.2

[42] Wada, K. and Shibata, T. (2006). Robot Therapy in a Care House - Results of Case Studies. In *Robot and Human Interactive Communication, 2006. ROMAN 2006. The 15th IEEE International Symposium on*, pages 581–586. 2.1

[43] Walters, M., Marcos, S., Syrdal, D. S., and Dautenhahn, K. (2013). An Interactive Game with a Robot: Peoples' Perceptions of Robot Faces and a Gesture-Based User Interface. In *ACHI 2013, The Sixth International Conference on Advances in Computer-Human Interactions*, number c, pages 123–128. 2.3

[44] Wrede, S., Fritsch, J., Bauckhage, C., and Sagerer, G. (2004). An XML Based Framework for Cognitive Vision Architectures. *ICPR (1)*, pages 757–760. 3.2

[45] You, Z.-j., Shen, C.-Y., Chang, C.-w., Liu, B.-j., and Chen, G.-d. (2006). A Robot as a Teaching Assistant in an English Class. *Sixth IEEE International Conference on Advanced Learning Technologies (ICALT'06)*, pages 87–91. 1.2

A. Appendix

A.1. Tables

Table A.1.: Mean values and standard deviation for the question group on *physical and mental requirements* for all days of the study.

Day	Mental-Demand	Physical-Demand	Time-Demand	Performance	Exertion	Frustration
1	6.250 (3.919)	2.880 (2.800)	7.750 (6.714)	4.630 (3.021)	5.880 (5.303)	15.880 (3.796)
2	9.380 (5.317)	5.000 (6.256)	3.250 (1.753)	4.750 (3.059)	10.630 (6.675)	15.130 (5.303)
3	9.000 (4.619)	4.860 (6.256)	2.570 (1.272)	5.000 (3.416)	5.860 (3.976)	11.570 (6.852)
4	7.750 (5.574)	4.880 (6.151)	4.630 (4.033)	5.500 (3.423)	11.880 (5.817)	15.380 (5.423)
5	8.880 (5.540)	3.880 (3.834)	2.380 (0.916)	5.130 (2.642)	8.500 (5.928)	14.250 (7.186)
6	9.500 (5.477)	2.500 (0.926)	2.130 (1.356)	6.250 (4.097)	9.630 (4.809)	14.000 (7.483)
8	8.380 (4.104)	2.500 (1.773)	3.130 (1.246)	6.630 (5.423)	13.380 (6.022)	14.000 (7.368)
9	8.130 (5.027)	2.880 (1.885)	2.880 (1.458)	6.630 (4.438)	11.880 (4.734)	12.630 (7.050)
10	9.000 (4.536)	3.880 (4.581)	3.250 (1.389)	7.880 (5.249)	9.750 (4.892)	13.880 (5.890)
11	7.130 (3.227)	2.630 (1.506)	2.630 (2.722)	4.380 (2.387)	7.630 (4.173)	10.380 (7.671)
12	9.000 (5.806)	2.880 (1.885)	3.380 (1.847)	5.380 (3.068)	8.880 (4.086)	6.500 (6.845)
13	9.130 (3.834)	2.380 (1.188)	3.000 (1.927)	6.250 (4.892)	7.880 (3.871)	6.380 (4.689)
15	7.380 (6.479)	1.880 (0.835)	2.630 (1.685)	6.380 (4.207)	8.500 (5.529)	10.750 (8.190)
16	9.000 (4.243)	3.000 (1.414)	2.630 (1.188)	4.000 (2.138)	10.500 (5.014)	6.500 (6.655)
17	9.000 (5.806)	2.500 (0.756)	3.000 (2.619)	3.880 (1.808)	11.630 (5.041)	7.380 (6.760)
18	8.880 (5.515)	2.130 (0.641)	3.130 (1.808)	4.880 (3.137)	7.750 (4.950)	6.500 (3.586)

Table A.2.: Mean values and standard deviation for the question group on *positive attitudes towards the interaction* for all days of the study.

Day	Excitement	Motivation	Joy
1	5.130 (0.835)	4.250 (1.035)	3.000 (1.604)
2	3.750 (0.463)	4.380 (1.061)	2.880 (0.835)
3	3.000 (1.528)	3.860 (1.345)	3.290 (1.604)
4	3.500 (1.069)	4.380 (0.916)	2.880 (0.641)
5	3.000 (1.195)	4.630 (0.744)	3.000 (1.414)
6	3.500 (1.195)	4.630 (1.188)	2.880 (1.553)
8	3.500 (1.414)	4.750 (0.707)	2.630 (1.188)
9	3.250 (0.886)	4.380 (0.744)	2.880 (0.835)
10	3.250 (0.886)	3.880 (0.641)	2.750 (0.886)
11	3.000 (0.756)	4.130 (0.835)	3.500 (1.309)
12	3.130 (1.246)	4.500 (1.195)	3.630 (1.188)
13	3.380 (0.916)	4.500 (0.926)	4.000 (0.926)
15	3.130 (0.991)	4.130 (1.126)	2.880 (1.642)
16	3.380 (1.302)	4.250 (1.389)	3.880 (1.458)
17	3.130 (1.126)	4.630 (0.916)	3.630 (1.188)
18	3.250 (1.282)	4.500 (0.756)	4.000 (1.069)

Table A.3.: Mean values and standard deviation for the question group on *physical and mental conditions* for all days of the study.

Day	Exhaustion	Agitation	Mood
1	2.250 (1.035)	2.380 (0.744)	4.630 (0.916)
2	3.630 (1.188)	2.750 (0.886)	4.130 (1.126)
3	3.140 (1.069)	2.430 (1.134)	4.710 (0.488)
4	3.500 (1.195)	2.750 (0.707)	4.380 (0.744)
5	2.750 (0.886)	3.000 (1.195)	4.630 (0.518)
6	3.380 (1.506)	2.380 (0.916)	4.750 (0.707)
8	3.500 (1.414)	2.880 (1.126)	4.250 (1.165)
9	4.000 (0.535)	2.880 (1.356)	4.250 (1.165)
10	3.880 (1.458)	2.380 (1.061)	4.500 (0.756)
11	3.250 (1.165)	2.380 (0.916)	4.750 (0.463)
12	3.630 (1.302)	2.250 (1.165)	4.380 (0.518)
13	3.880 (1.126)	2.630 (0.744)	4.630 (0.744)
15	3.750 (0.886)	2.250 (0.707)	4.130 (0.641)
16	3.630 (1.061)	2.000 (1.069)	4.630 (0.518)
17	3.500 (0.926)	2.130 (0.835)	4.130 (1.356)
18	2.630 (0.916)	2.500 (0.926)	5.250 (0.707)

Table A.4.: Results of the Friedman test for the three group of the questionnaires. The test showed a significant difference for the item *Frustration*.

Variable	**N**	$\tilde{\chi}^2$	**df**	**Sig**
MentalDemand	7	12.886	15	0.611
PhysicalDemand	7	19.813	15	0.179
TimeDemand	7	20.459	15	0.155
Performance	7	6.825	15	0.962
Exertion	7	16.573	15	0.345
Frustration	7	38.941	15	0.001
Excitement	7	23.777	15	0.069
Motivation	7	10.461	15	0.790
Joy	7	20.929	15	0.139
Exhaustion	7	23.918	15	0.066
Agitation	7	16.935	15	0.323
Mood	7	22.741	15	0.090

Table A.5.: Impressions on the gaming interaction throughout the study. Provided is the German original text and an English translation.

German	English
Hat sich im Vergleich zum Studienanfang stark verbessert, bleibt aber verbesserungswürdig	Strong improvement compared to the beginning, still needs improvement
Marsha hat sich noch weiter verbessert und macht immer weniger Fehler. Trotzdem läuft noch nicht alles rund. Alles in allem war es eine nette Abwechslung mit einem Roboter Memory zu spielen, aber deutlich anstrengender als mit einem echten Menschen.	Marsha improved her performance and makes less errors. Nevertheless the interaction is not perfect. Its a a nice diversion to play the game of pairs with a robot, but more exhausting compared to a human.
Es ist besser als zu Beginn, aber trotzdem noch sehr zäh und fehlerbehaftet. Es kommt nur selten richtiger Spielfluss zusammen.	Better compared to the beginning, nevertheless still tough and faulty. There is barely real game-play.
Marsha hat erhebliche Verbesserungen im Memory Spiel gezeigt. Sie spielt jedoch nicht, wie von einem Roboter zu erwarten wäre, perfect.	Marsha shows improvement for the game-play. Plays not perfect like it would be expected.
Marsha scheint sich besser an die Positionen der Karten erinnern zu können.	It seems Marsha improved her capabilities to remember card positions.
Hat Spaß gemacht.	Enjoyed to play.
Marsha scheint schlechte und gute Tage zu haben. Die Qualität des Spiels hängt stark vom Lärm in der Umgebung ab. Wenn es ruhig ist, klappt das Spiel gut und dann macht es auch Spaß.	It seems Marsha has good and bad days. The quality of the game depends on the surrounding noise. If it is quiet, the game-play is good and it makes fun to play.
Marsha schien sich verbessert zu haben. Klare Anweisungen konnte sie gut umsetzen, es war möglich zügig mit ihr zu spielen.	It seems Marsha improved. Clear commands were interpreted correctly. Fluid gamplay was possible.

Table A.6.: Problems mentioned throughout the study. Provided is the German original text and an English translation.

German	English
Hatte zeitweise kurze Aussetzer.	Sometime short blackouts.
Erkennt Karten nicht, erkennt Paare nicht. Versteht Befehle nicht, stürzt ab	No card detection, no pair detection, commands not understood, system crashes.
Sie erkennt Karten nicht, Sie sieht nicht alle Karten, Sie kennt keine Paare von mir oder ihr, Sie hört einfach auf zu spielen, Sie weiß nicht wer dran ist, Sie versteht mich oft nicht.	No card detection, not all cards seen, no pair detection of both parties, stops playing, did not know whose turn it is, did not understand commands.
Des öfteren werden Paare falsch erkannt, bzw. als solche erkennt die keine sind.	Fault card classification, names pairs that are no pairs.
Von Zeit zu Zeit scheint sich Marsha aufzuhängen. Sie reagiert dann minutenlang nicht.	Marsha sometimes crashes and is not responding.
Erkannte Paare nicht, sagte nichts mehr	No card detection, did not speak.
Paare werden nicht erkannt, bei Konversationen mit Marsha über Probleme im Spiel scheint mir die Technik schnell überlastet, Marsha reagiert dann oft gar nicht mehr.	No pair detection, through conversation about problems the system gets overloaded and stops responding.
ein Paar nicht erkannt, verschiedene Karten als Paar bezeichnet, flasche Anzahl von Karten auf dem Spielfeld gesehen, keine Reaktion bei Begrüßung	false detection of pairs, announces pair upon different cards, detects incorrect number of cards on field, no reaction upon greeting.

Table A.7.: Conclusions on the study drawn by the participants. Provided is the German original text and an English translation.

German	English
Macht zwar Spaß, muss aber verbessert werden	It was fun to play, needs improvement
Marsha war eine nette Ablenkung und eine interessante Erfahrung. Jedoch war das Spielen auch sehr anstrengend und manchmal nervig	Marsha was a nice diversion and an interesting experience. Nevertheless the game-play was exhausting and sometimes annoying.
Um richtig Spaß zu haben, müsste es viel schneller gehen und Marsha müsste besser werden. So hatte es eher den Character von einer Geduldsprobe für alle Probanden.	To enjoy the game-play, the interaction must be fast and Marsha needs to improve the game-play. It was more a trial of patience for all participants.
Anfangs interessant, zum Schluss sehr berechenbar und monton.	Interesting in the beginning, at the end foreseeable and monton.
Die meiste Zeit und Energie verwendet man nicht auf das Spiel selbst, sondern auf die Interaktion mit Marsha. Das ist ermüdend.	Most time and energy was not spend on the game, instead on the interaction with Marsha. That was exhausting.
Immer wieder gerne.	Always like to play again.
Wenn man ganz alleine isoliert wohnt ist es eine gute alternative mit ihr zu spielen. Einen Menschen ersetzen kann sie nicht. Mit 15 Minuten pro Spiel dauert alles einfach viel zu lang.	When you are alone and isolated it is a good alternative to play with the robot. It can not replace a human being. With 15 minutes per game it takes to long to play.
Marsha war für mich nie mehr als eine Maschine auf dem Niveau einer vierjährigen, mit der das Spielen sehr anstrengend und oftmals nervig war (obwohl sie sich in den letzten Tagen enorm gesteigert hat und ein Spielfluß zu Stande gekommen ist).	Marsha was never more than a machine for me on the level of a four year old child. Playing was exhausting and often annoying (nevertheless there was an improvement allowing a more fluent game-play).

Table A.8.: Adverbs describing the robot. Provided is the German original text and an English translation.

German	**English**
anspruchsvoll, cool, verbesserungswürdig	demanding, cool , need of improvement
nett, bemüht, fehlerhaft, stur, lernfähig, aber nur zu gewissem Grad	kind, willing, faulty, stubborn, adaptive
reizend(negativ), langsam, anstrengend, künstlich	provocative, slow, exhausting, artificial
interessant, sehr vorraussehbar, verbesserungsfähig, versteht nicht warum sie nicht perfect spielt	interesting, foreseeable, need of improvement, did not understand why it is not playing perfectly
ermüdend, fehleranfällig, umständlich	tiring, faulty, inconvenient
nervig, schlau, dumm, unverständlich	annoying, clever, dumb, incomprehensible
langsam, tagesform abhängig, geräuschempfindlich	slow, depending on day form, sound sensitive
anstrengend, frustrierend, nervig, lustig	exhausting, frustrating, annoying, funny

Table A.9.: Mentioned improvements for the system. Provided is the German original text and an English translation.

German	English
er hatte kaum Probleme, hat eig. Immer funktioniert, mehr Mimik, Mechanik verdecken, Hände+Gliedmaßen	barely any problem, system did work, more mimic, hide mechanic, hands+extremities
Kartenerkennung	Card detection
schneller spielen, fehler beheben(Erkennung), persönlicher einstellen	play fast, correct errors, add personality
Bilderkennung verbessern, abwechslungsreicher Dialog	correction for vision, more dynamic dialog
man musste lernen mit ihren fehlern umzugehen, Kartenerkennung und Anzahl	need to learn how to react upon errors, card detection and more cards
nix, hat immer gut geklappt	nothing, works fine every time
gesonderter Raum, Geräuschempfindlichkeit, Bilder bennen	extra room, noise sensitivity, names for pictures
war zu empfindlich, hat unfair gespielt, Kartenerkennung, Spracherkennung	to sensitive, played unfair, card detection, speech recognition

Table A.10.: Remarks on improvements noticed. Provided is the German original text and an English translation.

German	English
Spiel ist besser und flüssiger geworden, 2 Spiele in 30 Minuten	Games got better and more fluent, two games in 30 minutes
Fehler wurden weniger über die Zeit	Less errors over time
genereller trend in bessere richtung	generally a trend towards a better direction
Leistungszuwachs über die Zeit	Performance increase over time
im Lauf der zeit konnte sie sich besser zurückerinnern	better repetition of memory over time
wurde mit der Zeit schlauer, konnte mehr Paare entdecken	got smarter over time, found more pairs
ist besser geworden, hat fortschritte gemacht	got better, made progress
Spielfluss ist besser geworden, mehr Spiele waren möglich, am Ende hat es sogar ein bisschen Spaß gemacht	game-play got better, more games played, nearly enjoyed playing in the end

Table A.11.: Remarks on changes for the participants behavior. Provided is the German original text and an English translation.

German	English
dachte erst sie könne wie ein Mensch spielen, ist langsamer und geduldiger geworden	thought it could play like a human, played slower and more patient
kurze, knackigere Befehle	used shorter commands
deutlicher reden, Kommandos gelernt, bei Problemen neu starten	speak more clearly, learned commands, restart upon error
mehr Ungedult über die Zeit	more impatient over time
direktere Stichpunkte als Befehle, manchmal Ungehalten am Ende weil total genervt war, über sich selbst geärgert	more clear commands, impatient, galled over himself
ihre Fehler durchgehen lassen, damit Spiele weitergehen	accept errors to complete games
Sprechen vereinfacht, laut und deutlich reden, andere um Ruhe gebeten	speak slower and more clearly, ask others for silence
leiser geredet, weniger gesprochen, hat sie angeschaut	speak more quiet, speak less, look at the robot

Table A.12.: Results for the perception of the robot throughout the study. Presented are the results of a one-sample Wilcoxon signed-rank test.

Item	**Mean**	**SD**	**N**	**TestStat**	**StErr**	**AsymSig**
Disturbing	1.750	1.389	8	6.000	1.837	0.102
Distracting	1.875	1.126	8	10.000	2.716	0.066
Suitable	4.375	1.408	8	28.000	5.534	0.011
Annoying	2.625	1.302	8	34.000	7.062	0.023
Intimidating	1.250	0.463	8	3.000	1.061	0.157
Irritating	2.125	1.356	8	10.000	2.716	0.066
Motivating	1.750	1.165	8	6.000	1.871	0.109
Confusing	2.000	1.195	8	10.000	2.716	0.066
Pleasant	3.750	1.389	8	28.000	5.701	0.014
Helpful	1.375	0.518	8	6.000	1.732	0.083
Observing	2.250	1.488	8	10.000	2.646	0.059

Table A.13.: Results for the general questions asked on the final day of the study. Presented are the results of a one-sample Wilcoxon signed-rank test.

Item	**Mean**	**SD**	**N**	**TestStat**	**StErr**	**AsymSig**
I would interact with the robot again.	2.500	1.069	8	28.000	5.799	0.016
The system needs improvement.	5.750	1.282	8	36.000	7.045	0.011
I enjoyed interacting with the robot.	3.500	1.195	8	36.000	7.106	0.011
The interaction improved over time.	6.000	0.756	8	36.000	7.036	0.011
I felt well prepared for the interaction.	4.625	1.685	8	28.000	5.852	0.170
I knew how to react to the system.	3.750	1.982	8	34.500	7.098	0.020
The system reacted as expected.	4.375	1.506	8	36.000	7.089	0.011
The system worked reliable.	2.500	1.069	8	21.000	4.637	0.024
I understood the systems intention.	4.500	1.414	8	36.000	7.071	0.011
The system motivated me.	3.125	1.553	8	28.000	5.884	0.017
I would prefer playing with the robot instead of a human.	1.625	1.408	8	3.000	1.118	0.180

Table A.14.: Mean values and standard deviation for the rounds played for all days of the study.

Day	**Round Played**
1	2.250 (0.886)
2	1.875 (0.641)
3	2.071 (0.732)
4	2.250 (0.463)
5	2.250 (0.707)
6	2.250 (0.535)
8	1.813 (0.372)
9	2.125 (0.443)
10	1.875 (0.641)
11	2.063 (0.417)
12	2.063 (0.417)
13	2.438 (0.623)
15	1.750 (0.707)
16	2.125 (0.354)
17	2.125 (0.835)
18	2.250 (0.463)

Table A.15.: Results for the paired t-tests performed on the items for the *Negative Attitudes towards Robots*. For the item NARS1 a marginal significant difference was found.

Item	**Pre-Test**		**Post-Test**		**t(7)**	**Sig**	**Eta²**
	M	**SD**	**M**	**SD**			
NARS1	2.688	0.257	3.229	0.223	-1.664	.070	0.316
NARS2	3.925	1.710	3.725	1.519	0.798	0.226	0.096
NARS3	4.125	1.469	3.833	0.777	0.596	0.285	0.056

Table A.16.: Results of the Kruskal-Wallis Test for the ratings on the new facial expressions. The ratings match the expected expressions.

	HappyHigh	HappyLow	Neutral	SadLow	SadHigh
Chi-Square	65.098	65.508	13.272	31.696	43.934
df	6	6	6	6	6
Asymp. Sig.	0.000	0.000	0.039	0.000	0.000
Rank (Joy)	86.79	84.64	42.50	38.54	37.96
Rank (Anger)	35.00	37.00	42.50	37.86	41.14
Rank (Disgust)	35.00	37.00	45.79	43.57	45.79
Rank (Fear)	40.39	37.00	49.61	52.04	52.93
Rank (Contempt)	35.00	37.00	60.61	48.36	41.39
Rank (Sadness)	46.64	45.86	56.43	79.21	85.89
Rank (Surprise)	67.68	68.00	49.07	46.93	41.39

Table A.17.: Questions combined for the index *Machine-Like vs. Human-Like (Pre)*.

German	English
Bitte schätzen Sie auf der folgenden Skala Ihre Erwartungen an den Roboter ein:	Please rate your expectations for the robot using the scale:
[hat kein Bewusstsein—hat ein Bewusstsein]	[Mindless — has a Mind]
[künstlich—realistisch]	[Artificial — Realistic]
[bewegt sich steif—bewegt sich flüssig]	[Moves Stiffly — Moves Fluently]
[unecht—natürlich]	[Fake — Natural]
[tot—lebendig]	[Dead — Alive]
[unbewegt—lebhaft]	[Inactive — Lively]
[mechanisch—organisch]	[Mechanical — Organic]
[träge—interaktiv]	[Dull — Interactive]
[teilnahmslos—ansprechbar]	[Apathetic — Amenable]

Table A.18.: Questions combined for the index *Machine-Like vs. Human-Like (Post)*.

German	**English**
Ich habe den Roboter erlebt als:	I perceived the robot as:
[hat kein Bewusstsein—hat ein Bewusstsein]	[Mindless — has a Mind]
[künstlich—realistisch]	[Artificial — Realistic]
[bewegt sich steif—bewegt sich flüssig]	[Moves Stiffly — Moves Fluently]
[unecht—natürlich]	[Fake — Natural]
[tot—lebendig]	[Dead — Alive]
[unbewegt—lebhaft]	[Inactive — Lively]
[mechanisch—organisch]	[Mechanical — Organic]
[träge—interaktiv]	[Dull — Interactive]
[teilnahmslos—ansprechbar]	[Apathetic — Amenable]

Table A.19.: Questions combined for the index *Unlikable vs. Likable (Pre)*.

German	**English**
Bitte schätzen Sie auf der folgenden Skala Ihre Erwartungen an den Roboter ein:	Please rate your expectations for the robot using the scale:
[unfreundlich—freundlich]	[Unfriendly — Friendly]
[unhöflich—höflich]	[Impolite — Polite]
[unangenehm—angenehm]	[Unpleasant — Pleasant]
[furchtbar—nett]	[Awful — Kind]
[inkompetent—kompetent]	[Incompetent — Competent]
[ungebildet—kenntnisreich]	[Uneducated — Educated]
[verantwortungslos—verantwortungsbewusst]	[Irresponsible — Responsible]
[unintelligent—intelligent]	[Unintelligent — Intelligent]
[unvernünftig—vernünftig]	[Irrational — Rational]
[ängstlich—entspannt]	[Afraid — Relaxed]
[aufgewühlt—ruhig]	[Agitated — Calm]
[still—überrascht]	[Silent — Surprised]
[demotivierend—motivierend]	[Demotivating — Motivating]

Table A.20.: Questions combined for the index *Unlikable vs. Likable (Post)*.

German	English
Ich habe den Roboter erlebt als:	I perceived the robot as:
[unfreundlich—freundlich]	[Unfriendly — Friendly]
[unhöflich—höflich]	[Impolite — Polite]
[unangenehm—angenehm]	[Unpleasant — Pleasant]
[furchtbar—nett]	[Awful — Kind]
[inkompetent—kompetent]	[Incompetent — Competent]
[ungebildet—kenntnisreich]	[Uneducated — Educated]
[verantwortungslos—verantwortungsbewusst]	[Irresponsible — Responsible]
[unintelligent—intelligent]	[Unintelligent — Intelligent]
[unvernünftig—vernünftig]	[Irrational — Rational]
[ängstlich—entspannt]	[Afraid — Relaxed]
[aufgewühlt—ruhig]	[Agitated — Calm]
[still—überrascht]	[Silent — Surprised]
[demotivierend—motivierend]	[Demotivating — Motivating]

Table A.21.: Questions combined for the index *Social Presence*.

German	English
Für den Roboter waren Meine Gedanken klar.	My thoughts were clear to the robot.
Die Gedanken des Roboters waren für Mich klar.	The robots thoughts were clear to me.
Ich kann sagen wie der Roboter sich gefühlt hat.	I can state how the robot felt.
Der Roboter kann sagen wie Ich Mich gefühlt habe.	The robot can state how i felt.
Der Roboter reagiert auf meine Gefühle.	The robot reacts to my feelings.
Ich reagierte auf die Gefühle des Roboters.	I reacted to the robots feelings.

Table A.22.: Questions combined for the index *Gaming Support*.

German	English
Der Roboter hat Mir während des Spiels geholfen.	The robot was helpful while playing the game.
Die Ansagen des Roboters waren für Mich hilfreich.	Announcements from the robot were helpful to me.
Die Kommentare des Roboters waren für Mich hilfreich wenn Ich sie brauchte.	The comments of the robot were helpful when I needed them.
Ich hatte das Gefühl durch die Anwesenheit des Roboters besser gespielt habe.	I think I played better with the robot around.
Der Roboter hat Mich gelobt wenn Ich gut gespielt habe.	The robot praised me when I played well.
Der Roboter hat Mich gelobt wenn Ich nicht so gut gespielt habe.	The robot praised me when I did not play so well.

Table A.23.: Questions combined for the index *Social Support*.

German	English
Der Robter sorgt sich um Mich.	The robot cares for me.
Der Roboter gibt gute Tipps.	The robot offers good advises.
Der Roboter akzeptiert Mich wie Ich bin.	The robot accepts me the way I am.
Der Roboter unterstützt Meine Entscheidungen.	The robot supports my decisions.
Ich kann auf den Roboter zählen.	I can count on the robot.
Der Roboter ermutigt mich.	The robot encourages me.
Der Roboter versteht mich.	The robot understands me.
Der Roboter lobt Mich wenn Ich etwas besonders gut gemacht habe.	The robot praises me when I did something peculiar good.

Table A.24.: Questions combined for the index *Interaction Complexity*.

German	**English**
Die Interaktion mit dem Roboter	The interaction with the robot
[fiel mir schwer—fiel mir leicht]	[was difficult for me — was simple for me]
[war ineffizient—war effizient]	[was inefficient — was efficient]
[war verwirrend—war selbsterklärend]	[was confusing — was self explaining]
[war frustrierend—war angenehm]	[was frustrating — was pleasant]
[war kompliziert—war leicht zu verstehen]	[was complicated — was easy to understand]

Table A.25.: Results for the group on the *Social Perception*. Table (a) shows the mean values for the sessions of each condition. Table (b) shows the results for the independent t-test. Table (c) shows the results of the paired t-test conducted within conditions.

Item	C1		C2		C3	
	W1 M(SD)	**W4 M(SD)**	**W1 M(SD)**	**W4 M(SD)**	**W1 M(SD)**	**W4 M(SD)**
Machine-Like vs. Human-Like (Pre)	3.408 (1.037)	3.754 (1.359)	3.379 (0.639)	3.772 (1.177)	3.242 (0.626)	3.542 (1.180)
Machine-Like vs. Human-Like (Post)	4.477 (0.842)	4.424 (1.128)	4.500 (0.831)	4.929 (0.662)	5.442 (0.826)	5.150 (0.828)
Unlikable vs. Likable (Pre)	4.847 (0.934)	4.781 (1.168)	4.745 (0.697)	4.786 (0.899)	4.905 (0.498)	5.244 (0.902)
Unlikable vs. Likable (Post)	3.819 (0.694)	3.446 (0.847)	3.990 (0.743)	3.960 (0.857)	4.167 (0.531)	3.685 (0.688)
Social Presence	2.924 (1.04)	2.359 (0.869)	3.298 (1.139)	3.143 (1.228)	3.237 (0.825)	2.750 (0.764)

(a) Mean Value for each condition presenting session one and session four.

Item	C1 vs. C2	C2 vs. C3
Machine-Like vs. Human-Like (Pre)	t(25) = 0.017, p = 0.494	t(24) = -0.114, p = 0.455
Machine-Like vs. Human-Like (Post)	t(25) = -0.386, p = 0.352	t(24) = -1.486, p = 0.075
Unlikable vs. Likable (Pre)	t(25) = -0.409, p = 0.343	t(24) = 0.730, p = 0.237
Unlikable vs. Likable (Post)	t(25) = -1.426, p = 0.083	t(24) = -1.090, p = 0.143
Social Presence	t(25) = 0.100, p = 0.461	t(24) = 0.202, p = 0.421

(b) Results for the independent t-tests for effects between conditions.

Item	C1			C2			C3		
	t(12)	**Sig.**	**Eta²**	**t(13)**	**Sig.**	**Eta²**	**t(11)**	**Sig.**	**Eta²**
Machine-Like vs. Human-Like (Pre)	-0.874	0.200	0.065	-1.527	0.076	0.163	-0.755	0.233	0.054
Machine-Like vs. Human-Like (Post)	0.293	0.388	0.008	-2.328	0.019	0.311	2.159	0.027	0.318
Unlikable vs. Likable (Pre)	0.228	0.412	0.005	-0.165	0.436	0.002	-1.065	0.155	0.102
Unlikable vs. Likable (Post)	2.627	0.011	0.386	0.142	0.445	0.002	2.605	0.012	0.404
Social Presence	3.439	0.003	0.518	0.438	0.335	0.016	1.726	0.056	0.230

(c) Results for the paired t-tests for effects within conditions.

Table A.26.: Results for the group on the *Social Support*. Table (a) shows the mean values for the sessions of each condition. Table (b) shows the results for the independent t-test. Table (c) shows the results of the paired t-test conducted within conditions.

Item	**C1**		**C2**		**C3**	
	W1 M(SD)	**W4 M(SD)**	**W1 M(SD)**	**W4 M(SD)**	**W1 M(SD)**	**W4 M(SD)**
Gaming Support	3.308 (0.836)	3.372 (0.994)	3.584 (1.164)	3.227 (0.940)	3.278 (1.160)	3.459 (1.217)
Social Support	3.279 (1.024)	3.202 (1.010)	3.322 (1.546)	3.188 (1.231)	2.990 (1.123)	3.105 (0.810)
Interaction Complexity	4.293 (1.230)	4.339 (1.520)	4.500 (1.478)	4.986 (1.378)	5.984 (0.770)	5.850 (0.954)

(a) Mean Value for each condition presenting session one and session four.

Item	**C1 vs. C2**	**C2 vs. C3**
Gaming Support	t(25) = 1.851, p = 0.038	t(24) = 3.062, p = 0.003
Social Support	t(25) = 0.282, p = 0.391	t(24) = -0.750, p = 0.231
Interaction Complexity	t(25) = 1.303, p = 0.102	t(24) = 1.557, p = 0.067

(b) Results for the independent t-tests.

Item	**C1**			**C2**			**C3**		
	t(12)	**Sig.**	**Eta²**	**t(13)**	**Sig.**	**Eta²**	**t(11)**	**Sig.**	**Eta²**
Gaming Support	-0.216	0.416	0.004	1.568	0.071	0.170	-0.610	0.278	0.036
Social Support	0.330	0.374	0.010	0.660	0.260	0.035	-0.438	0.335	0.019
Interaction Complexity	-0.180	0.430	0.003	-2.474	0.014	0.338	0.597	0.282	0.034

(c) Results for the paired t-tests for effects within conditions.

Table A.27.: Results for the group for the *Ratings on Gaming*. Table (a) shows the mean values for the sessions of each condition. Table (b) shows the results for the independent t-test. Table (c) shows the results of the paired t-test conducted within conditions.

Item	C1		C2		C3	
	W1 M(SD)	**W4 M(SD)**	**W1 M(SD)**	**W4 M(SD)**	**W1 M(SD)**	**W4 M(SD)**
EN1	5.080 (1.441)	5.150 (1.625)	4.500 (1.829)	5.140 (1.099)	4.250 (1.765)	3.670 (1.775)
EN3	4.310 (1.548)	4.000 (1.871)	4.500 (1.787)	4.930 (1.685)	5.500 (1.382)	5.250 (1.815)
EN6	5.080 (1.801)	4.920 (1.935)	4.930 (1.979)	5.570 (1.742)	5.420 (1.929)	5.080 (1.929)
EN7	3.460 (1.561)	3.850 (1.519)	4.860 (1.562)	5.290 (1.684)	6.170 (1.697)	5.580 (2.151)
EN8	5.850 (1.405)	5.460 (1.664)	4.860 (1.956)	5.210 (1.718)	6.500 (1.00)	6.330 (1.497)
EN9	6.000 (1.354)	5.690 (1.437)	5.070 (1.639)	5.860 (1.099)	6.170 (1.467)	5.920 (1.676)
EN10	2.850 (1.281)	2.850 (1.625)	3.360 (1.985)	3.140 (1.512)	6.080 (1.782)	6.080 (1.730)
EN11	4.770 (1.423)	3.920 (1.977)	5.710 (1.684)	5.000 (1.754)	5.920 (1.443)	4.500 (2.111)

(a) Mean Value for each condition presenting session one and session four.

Item	**C1 vs. C2**	**C2 vs. C3**
EN1	t(25) = -1.373, p = 0.091	t(24) = -2.09, p = 0.024
EN3	t(25) = -1.261, p = 0.110	t(24) = -0.061, p = 0.476
EN6	t(25) = -0.303, p = 0.382	t(24) = 0.486, p = 0.316
EN7	t(25) = -1.471, p = 0.077	t(24) = -1.309, p = 0.102
EN8	t(25) = -1.185, p = 0.124	t(24) = -1.230, p = 0.116
EN9	t(25) = -0.287, p = 0.388	t(24) = -2.319, p = 0.015
EN10	t(25) = -1.325, p = 0.099	t(17.274) = -1.993, p = 0.031
EN11	t(25) = -0.086, p = 0.467	t(24) = -1.372, p = 0.092

(b) Results for the independent t-tests.

Item	C1			C2			C3		
	t(12)	**Sig.**	**Eta²**	**t(13)**	**Sig.**	**Eta²**	**t(11)**	**Sig.**	**Eta²**
EN1	-0.322	0.3765	0.009	-1.662	0.060	0.187	1.103	0.147	0.108
EN3	0.843	0.208	0.061	-1.249	0.117	0.115	0.638	0.268	0.039
EN6	0.395	0.350	0.014	-1.422	0.090	0.144	1.773	0.052	0.239
EN7	-1.443	0.0875	0.159	-1.000	0.168	0.077	0.939	0.184	0.081
EN8	0.686	0.253	0.041	-0.751	0.233	0.045	0.518	0.308	0.026
EN9	0.843	0.208	0.061	-1.863	0.043	0.224	0.583	0.286	0.033
EN10	0.000	0.500	0.000	0.425	0.339	0.015	0.000	0.500	0.000
EN11	2.668	0.010	0.393	2.219	0.023	0.291	2.545	0.014	0.393

(c) Results for the paired t-tests for effects within conditions.

Table A.28.: Results for the group for the *Ratings on Engagement*. Table (a) shows the mean values for the sessions of each condition. Table (b) shows the results for the independent t-test. Table (c) shows the results of the paired t-test conducted within conditions.

Item	C1		C2		C3	
	W1 M(SD)	W4 M(SD)	W1 M(SD)	W4 M(SD)	W1 M(SD)	W4 M(SD)
EN5	3.770 (1.641)	3.770 (1.833)	4.360 (1.692)	4.500 (1.454)	5.500 (1.508)	4.250 (2.094)
EN14	5.080 (1.382)	5.080 (1.382)	4.070 (1.207)	4.930 (1.439)	5.000 (1.595)	4.170 (1.403)
EN15	5.380 (1.502)	5.460 (1.450)	5.290 (1.437)	5.290 (1.541)	6.580 (0.996)	6.500 (1.000)
EN22	2.770 (1.166)	2.540 (1.391)	2.790 (1.888)	2.710 (1.069)	3.330 (1.723)	2.580 (1.084)
EN23	4.460 (1.450)	4.080 (2.290)	4.360 (1.598)	5.360 (1.737)	3.920 (1.730)	4.080 (1.881)

(a) Mean Value for each condition presenting session one and session four.

Item	C1 vs. C2	C2 vs. C3
EN5	t(21.448) = -1.245, p = 0.114	t(24) = -1.907, p = 0.035
EN14	t(25) = 0.332, p = 0.372	t(17.399) = 0.391, p = 0.350
EN15	t(25) = -0.291, p = 0.387	t(24) = -1.131, p = 0.135
EN22	t(25) = 0.366, p = 0.359	t(24) = -0.861, p = 0.199
EN23	t(25) = -0.101, p = 0.460	t(24) = -0.462, p = 0.324

(b) Results for the independent t-tests.

Item	C1			C2			C3		
	t(12)	Sig.	Eta²	t(13)	Sig.	Eta²	t(11)	Sig.	Eta²
EN5	0.000	0.500	0.000	-0.414	0.343	0.014	2.454	0.016	0.376
EN14	0.000	0.500	0.000	-1.835	0.045	0.219	1.968	0.038	0.279
EN15	-0.201	0.422	0.004	0.000	0.500	0.000	0.220	0.415	0.005
EN22	0.64	0.267	0.036	0.138	0.447	0.002	1.326	0.106	0.150
EN23	0.959	0.178	0.077	-1.661	0.061	0.187	-0.378	0.357	0.014

(c) Results for the paired t-tests for effects within conditions.

Table A.29.: Results for the annotation of the video recordings. Table (a) shows the mean values for the sessions of each condition. Table (b) shows the results for the independent t-test. Table (c) shows the results of the paired t-test conducted within conditions.

Item	**C1**		**C2**		**C3**	
	W1 M (SD)	**W4 M (SD)**	**W1 M (SD)**	**W4 M (SD)**	**W1 M (SD)**	**W4 M (SD)**
Rounds Played	2.770 (0.599)	3.230 (1.481)	2.929 (1.492)	4.500 (2.066)	2.830 (1.115)	3.920 (2.314)
Time per Round (in Sec.)	358.870 (66.305)	361.572 (92.778)	448.352 (103.458)	384.209 (83.123)	329.418 (34.038)	281.968 (43.991)
Time per Session (in Sec.)	1173.649 (455.574)	1159.569 (509.769)	1604.112 (879.220)	1706.516 (720.161)	1137.221 (438.670)	1166.542 (717.866)
Number of Cheating per Session	0.000 (0.000)	0.000 (0.000)	0.643 (1.598)	0.000 (0.000)	0.333 (0.888)	1.667 (3.473)

(a) Mean Value for each condition presenting session one and session four.

Item	**C1 vs. C2**	**C2 vs. C3**
Rounds Played	t(25) = 1.346, p = 0.096	t(24) = 0.502, p = 0.310
Time per Round (in Sec.)	t(25) = -1.395, p = 0.088	t(24) = -0.443, p = 0.331
Time per Session (in Sec.)	t(25) = 0.367, p = 0.359	t(24) = 0.213, p = 0.417

(b) Results for the independent t-tests.

Item	**C1**			**C3**			**C3**		
	t(12)	**Sig.**	**Eta²**	**t(13)**	**Sig.**	**Eta²**	**t(11)**	**Sig.**	**Eta²**
Rounds Played	-1.066	0.154	0.094	-2.294	0.020	0.305	-1.595	0.071	0.203
Time per Round (in Sec.)	-0.076	0.471	0.001	1.990	0.034	0.248	3.070	0.006	0.485
Time per Session (in Sec.)	0.088	0.466	0.001	-0.383	0.354	0.012	-0.147	0.443	0.002

(c) Results for the paired t-tests for effects within conditions.

Table A.30.: Mean values for the occurrences of cheating behaviors in conditions C2 and C3.

Condition	**W1 M (SD)**	**W4 M (SD)**
C2	0.643 (1.598)	0.000 (0.000)
C3	0.333 (0.888)	1.667 (3.473)

A.2. Figures

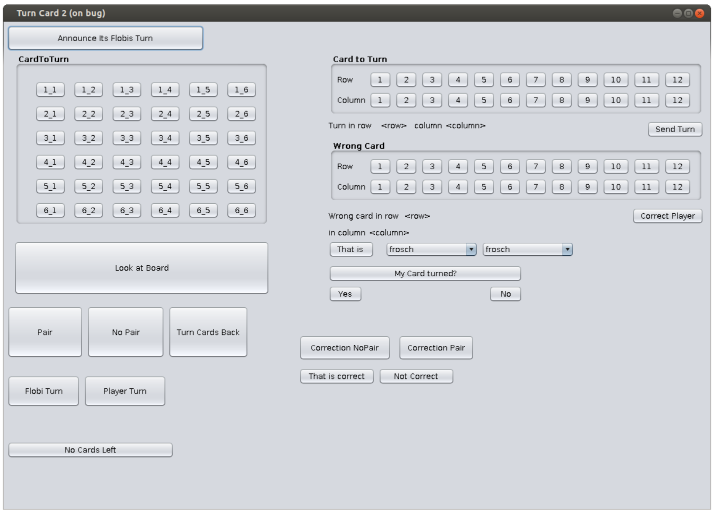


Figure A.1.: View of the Controller GUI to select cards for the robot to be turned by the participant. By directly selecting column and row a dialog task with the corresponding information gets triggered.

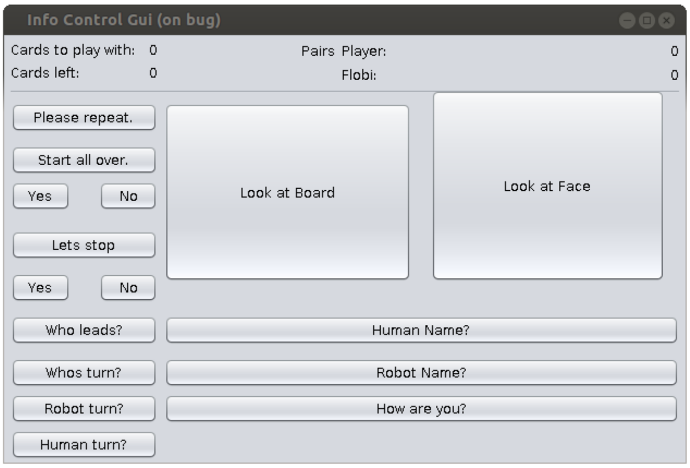


Figure A.2.: Independent information view of the Controller GUI. This view allows to react on questions of the participant and to restart the interaction.

A.3. Listings

Listing A.1: Example of the extended dialog pattern to use context information. Elements at the beginning of each phase item indicate a condition to be checked for selection of the output to use.

```
...
  <phrases choice="nextValid">
    <phrase>
      <!-- gloablgamesdiff >= 2 -->
      {gt-globalgamesdiff-2}
      Wieder gewinne ich, langsam musst
      du dich anstrengen!
    </phrase>
    <phrase>
      <!-- gloablgamesdiff >= 1 -->
      {gt-globalgamesdiff-1}
      Ich gewinne und liege mit einem Spiel vorne.
    </phrase>
    <phrase>
      <!-- gloablgamesdiff == 0 -->
      {eq-globalgamesdiff-0}
      Damit erziehle ich den ausgleich in unseren Spielen.
    </phrase>
    <phrase>
      <!-- gloablgameplayed >= 0 -->
      {gt-globalgamesplayed-0}
      Ich gewinne und damit steht es jetzt
      %globalgameswonrobot% zu %globalgameswonhuman%.
    </phrase>
  </phrases>
...
```