

**Comparison of Robots' and Embodied Conversational Agents' Impact on
Users' Performance**

Jakub Zlotowski

University of Tampere
Department of Computer Sciences
Interactive Technology
M.Sc. thesis
Supervisor: Markku Turunen
August 2010

University of Tampere

Department of Computer Sciences

Interactive Technology

Jakub Zlotowski: Comparison of Robots' and Embodied Conversational Agents' Impact on Users' Performance

M.Sc. thesis, 51 pages, 4 index pages

August 2010

Robots and Embodied Conversational Agents (ECAs) are two technologies that strive to make computers more accessible for their users by incorporating physical body. Both have been widely used in different domains, such as education or entertainment. However, relatively little attention has been paid to specific qualities of these technologies. The primary focus of this paper is on comparison of robots' and ECAs' impact on users' task performance. An experiment with 16 participants was conducted in between-subjects design. Subjects were asked to solve mathematical problems on a computer. Moreover, each participant was interacting with either a small robot-rabbit or its computer agent version, which provided them with feedback on their task performance. The data from a post-test questionnaire regarding the interaction and the task performance was analyzed using Mann-Whitney U test, and independent and dependent samples t-tests. The results show that both the robot and the ECA helped participants to focus on the task and the amount of time required to solve a problem decreased during the course of the experiment. Moreover, it took participants more time to solve a problem in the robot's than the agent's condition, but participants were more forgiving for the robot's than the ECA's repetitive feedback. Furthermore, the robot and the agent were liked, and the interaction with them was rated as entertaining and fun. These findings have important implications for a choice between these 2 technologies as educational and entertainment tools.

Key words and terms: robot, embodied conversational agent, human-robot interaction, social robot.

Acknowledgements

I would like to express my greatest gratitude to my supervisor, Markku Turunen, for his encouragement during preparations of the experiment presented in this paper and insightful feedback that helped to make it more interesting. I am also grateful to Jaakko Hakulinen for his help with organizing the experiment.

I would like to acknowledge the staff at the Department of Computer Sciences and the Faculty of Information Sciences for supporting me during the application process and giving me a chance to study at the University of Tampere.

I am grateful to Jaspreet Singh whose help with the development of Java based computer application allowed me to conduct the experiment few months earlier. I would like also to thank Joao Carlos Andrioli Machado for his help with the development of the original version of the application in VB, which was abandoned due to technical problems.

Finally, I want to thank my girlfriend Laura Jose Moreno Cidras for her patience and support during my work on this thesis.

Contents

1. Introduction	1
2. Literature Review	5
2.1 Social Technology	5
2.2 Embodiment.....	8
2.3 Physical presence	12
2.4 Social Facilitation	14
3. Methods.....	16
3.1 Participants.....	16
3.2 Equipment	16
3.2.1 Nabaztag	17
3.2.2 jNabServer	18
3.2.3 Embodied Conversational Agent.....	19
3.2.4 Computer application	20
3.3 Procedure	22
3.4 Variables	23
4. Results	25
4.1 Task perception.....	25
4.2 Social acceptance and User experience	26
4.3 Task performance.....	28
5. Discussion	32
5.1 Task performance.....	32
5.2 Perceived impact of Nabaztag	35
5.3 Entertainment.....	36
6. Conclusions	38
References	42
Appendix A.....	49
Appendix B.....	50
Appendix C.....	51

1. Introduction

Since prehistoric times humankind has glorified the living things of nature. Plants and animals have been worshiped and in many cultures they have been regarded as the representation of gods. Nevertheless, we, as human beings, tend to see ourselves on top of the hierarchical tree of species. This appreciation of nature and the desire to bend the surrounding world to serve our needs are perhaps the strongest motivations for giving technology “life” and making it resemble living creatures. Two of the most prominent examples of this technological attempt are Embodied Conversational Agents (ECAs) and robots. However, anthropomorphizing technology does not only refer to physical appearance, but also adds other human-like characteristics such as speech, facial expressions and emotional or social capabilities.

Many believe that anthropomorphic technology will provide benefits over faceless, text-based computer displays. Humanizing computers could make them easier and more comfortable to use [Laurel, 1997; Shneiderman and Maes, 1997]. Moreover, it would allow the user the use of various modalities, rather than forcing him to read a text, which may be disruptive to the main task [Catrambone et al., 2004].

In addition to their potential performance improvements, personified interfaces can also positively affect user experience and the social acceptance of technology. Koda and Maes [1996] reported that they are more engaging and well suited for the entertainment domain. This finding was further supported by Bickmore and Picard [2005] who found that people were willing to engage in a relationship with an ECA, and perceived that relationship more positively than when interacting with non-relational agents. Participants created an emotional-bond with an agent. Moreover, similar observations were reported for Human-Robot Interaction (HRI) [Breazeal and Scassellati, 2000]. This trend can be also noticed on the market as personal service robots such as Sony’s robot-dog Aibo, Violet’s robot-rabbit Nabaztag or Philip’s iCat have become commercially available. In addition, ECAs such as Microsoft Word’s Clip have been made available to a wide market.

On the other hand, Sproull et al. [1997] found negative aspects of embodiment, as participants in their experiment felt less relaxed and confident, and expressed higher arousal when interacting with a talking-face display in comparison to those interacting with

a text display. Nevertheless, they also reported that people attributed some of the personality features differently in these two conditions.

One of the areas that could benefit from the ability of ECAs and robots to engage people and improve their performance is education. Lester et al. [1997] found that pupils exhibited performance gains after interacting with an animated pedagogical agent and had a more positive perception of their learning experience. Furthermore, Rickel and Johnson [2000] implied that such an agent in virtual reality could provide interactive demonstrations, navigational guidance, nonverbal attentional guides and feedback, which would lead to better processing and memorizing of the studied material.

The benefits of robots as educators have been also extensively investigated. Robins et al. [2005] conducted a longitudinal study on children with autism, who during the course of the experiment showed improvements in social skills. Moreover, robots have also successfully worked as tour-guides in museums [Burgard et al., 1999; Shiomi et al., 2006].

Although considerable research has been devoted to the benefits of embodiment in education, rather less attention has been paid to the comparison of the special qualities of ECAs and robots. It could be interesting to see how they differ and in which situations one technology should be used over the other. Yamato et al. [2001] was the pioneer of such research and he indicated that an ECA had a bigger impact on human choices, but people felt closer to a robot. The area of comparison was expanded further by Powers et al. [2007]. Participants in their experiment were more engaged, enjoyed it more and felt more sense of presence when interacting with a robot. Moreover, they found a robot to be more lifelike and attributed to it a higher number of personality traits. On the other hand, they disclosed more information to the computer agent and were able to recall more information from a conversation with it.

However, it remains unclear whether ECAs and robots would affect user task performance differently. Yamato et al. [2001] and Powers et al. [2007] were interested only in social aspects of interaction. Nevertheless, it is possible that the social perception of the interaction reported in the papers described above would differ in conditions where people should focus on some tasks that require attention instead of having a relaxing conversation with an agent.

Moreover, considering the relative popularity of animated pedagogical agents and robots in education, it is important to explore their potential impact on people's quality of work. While there are numerous benefits of the embodiment, it is possible that robot's physical presence in the real world will work as a moderator and bring different qualities in comparison with an ECA.

It is well known that even the mere presence of others can affect one's task performance. People exhibit an improvement of performance on simple tasks and an impairment of performance on complex tasks when another person is present. This phenomenon is called "social facilitation" [Zajonc, 1965]. Zajonc explained it by stating that the presence of others serves as a source of arousal. In addition, from Yerkes-Dodson law [Yerkes and Dodson, 1908], we know that arousal increases the likelihood of an organism making habitual or well learned responses.

Baron [1986] proposed a cognitive explanation for the social facilitation effect. He suggested that the attention conflict between the task and observer can facilitate simple as well as impair complex tasks. The current view in social psychology is that both arousal and cognitive processes influence social facilitation [Aiello and Douthitt, 2001].

There is a vast body of research under the "Computers are Social Actors" (CASA) paradigm, which reveals that people show social responses to different types of media in a similar manner as when interacting with other humans [Fogg and Nass, 1997; Lee et al., 2000; Nass and Lee, 2000; Nass et al., 1997]. Nass et al. [1997] suggested that we can take any single theory about human-human interaction from Psychology and replace one human with a machine to test its validity in HCI.

Based on this paradigm, it was shown that ECAs produce a social facilitation effect [Hall and Henningsen, 2008; Park and Catrambone, 2007]. Since both Yamato et al. [2001] and Powers et al. [2007] reported that people felt a higher presence of and were more engaged by a robot than an ECA, we can expect that the social facilitation effect will be stronger for users interacting with the former.

This thesis reports an experiment to explore the differences in human interaction with ECAs and robots in the working or educational domain. The main goal of the study

was to find in what context these technologies provide optimal benefit for its users. The following aspects were analyzed:

- The impact of ECA and robot on the task performance
- The impact of ECA and robot on people's perception of the task
- Human social perception of ECA and robot

The experiment was conducted in the usability lab at the University of Tampere. Sixteen participants volunteered for this experiment. The design was between-subjects design, where each participant interacted either with a robot or an ECA. They were asked to solve a series of modular arithmetic problems on a computer. A small robot-rabbit or its computer agent version was used to provide feedback on the participants' task performance. After 10 minutes of doing this mathematical task, participants were asked to fill a post-test questionnaire. Subjects' task performance and answers to the questionnaire were recorded. The statistical software package SPSS was used to analyze the data using the Mann-Whitney U test, and independent and dependent samples t-tests.

The principal findings of this experiment show that on average it took participants more time to solve one modular arithmetic task in the robot's than in the agent's condition. Moreover, subjects showed higher forgiveness for the robot's repetitive feedback compared with the ECA. The most promising results for educational domain show that both the robot and the ECA helped participants to focus on the task and the amount of time required to solve a problem decreased during the course of the experiment. Furthermore, the robot and the agent were perceived as entertaining and liked by subjects, with the robot rated slightly higher on the former scale.

The thesis is structured with several chapters, sub-chapters and sub-sub-chapters. Chapter 2 introduces the theoretical background of social interaction with ECAs and robots. The methodology and experimental set-up are explained in Chapter 3. The results of the conducted experiment are presented in Chapter 4, while Chapter 5 is devoted to the discussion of the results obtained and their potential implications for Human-Computer Interaction (HCI). Finally, in Chapter 6, the conclusions are presented together with ideas for further research.

2. Literature Review

In this chapter I present literature related to this study. The main theoretical concepts are explained and by reviewing what has been previously done I indicate gaps in current knowledge about ECAs and robots. I begin in chapter 2.1 by presenting evidence that human interaction with technology is essentially social in a similar way to interaction with other human beings. Chapter 2.2 is devoted to the research on technology embodiment and its consequences for social interaction. In chapter 2.3 a detailed comparison of ECAs and robots is presented. Finally, in the last chapter, 2.4, I explain the social facilitation effect as an important factor for choosing between ECAs or robots in education.

2.1 Social Technology

People's social responses to technology were treated as an abnormality in the 1980s. It was thought at that time that this behavior when interacting with computers was the result of a person's dysfunction. Many researchers believed that only people who were young, had a lack of knowledge about technology or either a psychological or social dysfunction would respond socially towards a machine [Barley, 1988; Turkle, 1984; Winograd and Flores, 1987; Zuboff, 1988]. According to the same authors "normal", educated and mentally adjusted people would not express any social behaviors while interacting with a machine.

On the other hand, most computer users have experience of speaking to the computer themselves or have heard other people doing it in various situations, e.g., swearing at the computer when it suddenly crashed and important data was lost or muttering while playing computer games. It would hardly be possible to explain this type of behavior with the above mentioned theories considering that the majority of people are well adjusted individuals. Dennett [1987] proposed another explanation. He suggested that as technology is simply a proxy for the programmer [Searle, 1980], people's social behavior towards it is in fact not directed at technology itself, but rather at its creator.

However, those views were challenged in a series of experiments on the "Computers are Social Actors" (CASA) paradigm [Nass et al., 1997], which showed that social responses to different types of media are normal and common, and are not a result of

dysfunction or unconsciously directing those responses towards a human creator. In addition, people are not aware of engaging in this kind of relationship with a machine and very often it occurs in contradiction to their conscious declarations about not seeing machines as social.

In the same paper Nass et al. [1997] proposed that any single theory about human-human interaction from Psychology can be tested in HCI by replacing one human with a computer. Moreover, those findings impact interaction not only with computers, but also all types of media and machines, and are universal.

Furthermore, in the same series of experiments, participants behaved differently when even the smallest cues, such as text style in text based HCI or voice of machine, were modified. Moreover, people behaved politely towards the computers in a similar manner as they do with other humans, as if they did not want to hurt their feelings. Another interesting finding from this research shows that computers with similar personality to their users were preferred. Computers that expressed mismatching personality behaviors in different modalities were liked less than when they were consistent [Lee and Nass, 2003; Nass et al., 2000]. In addition, computers that flattered users, even insincerely, made them feel more positively about themselves, interaction and computers.

This research shows clearly that the sociality of HCI spans many areas and is more universal than people may think. Further investigations on this topic showed that just vocal cues alone are enough to elicit stereotypic responses from the users and to make them behave towards the machine as if it had gender [Lee et al., 2000]. Even when the voice was deliberately made to sound like that produced by a machine to remind users that they were interacting with a computer, they still showed a tendency to attribute personality to it [Nass and Lee, 2000]. Moreover, Fogg and Nass [1997] revealed that people were keener to help computers that had helped them earlier, which shows that based on the rule of reciprocity (tendency to help others who helped us before) computers are able to motivate people to change their behavior.

Finally, Klein et al. [2002] reported that users interact longer with a computer that had previously caused their frustration if the system is affectively supportive. This shows

clearly a need for systems that allow users to reach their goals, but also help them control their emotional state in case of failure.

However, not all researchers agree that human interaction with a machine is the same as human-human interaction. While research into the CASA paradigm investigated how socio-psychological theories apply to HCI, Shechtman and Horowitz [2003] proposed a different methodology. Instead of measuring non-conversational behaviors, they were interested to see whether there would be a difference in the way people communicate with other people and machines. Participants in their experiment were informed whether they would be talking with another person using a text-based computer mediated program or whether their conversational partner would be a computer. The quality of the conversation was analyzed afterwards.

Participants put more effort, used more words and spent more time on the conversation, when their conversational partner was a human. Moreover, those conversations differed in quality. Participants who knew that they were interacting with a human partner wrote more relationship statements by trying to build a connection or influence the partner. However, they also expressed more yielding and hostile behaviors towards their discussants. Finally, among the assertive participants, only those who had a human partner exhibited that trait and only an assertive human partner influenced participants.

These results imply that social HCI is not the same as interaction between people. It is crucial to remember that by simply replicating human interaction by a machine, its users will not be responding to it exactly in the same manner as they would while speaking with another person. However, if the machine is capable of pursuing a normal conversation, the users will probably notice it, learn to trust it and finally believe that the system can actually understand them, which will lead to a better quality of following conversations. It is possible that the participants in Shechtman and Horowitz's [2003] experiment used a simpler language as they assumed that the computer would not be able to understand longer and more complicated sentences, and its behavior could hardly be influenced by assertive statements. [Zlotowski, 2010]

The above findings should be taken into consideration when designing new systems. A thorough evaluation framework for Human-Robot Interaction was presented by Weiss et al. [2009] who postulate the importance of evaluating not only the usability of

robots, but also other factors: social acceptance, user experience and societal impact. Despite possible differences in interaction between human and machine, even the smallest changes of voice or physical appearance can result in a different perception of the system by the users and also cause their different behavior. Unconsidered decisions can have negative consequences for the usefulness and reception of the system. However, carefully designed systems that will utilize natural human social and emotional needs and skills can improve human machine interaction. In the next sub-chapter I will present how it can be achieved by anthropomorphizing the interface.

2.2 Embodiment

One of the ways of improving social communication with a machine is achieved by an embodiment of the system. Currently the two most prominent approaches are ECAs - virtual agents displayed on a computer display - and robots – physically-embodied systems. It is believed that humanizing computers would make them easier and more natural to use [Laurel, 1997; Shneiderman and Maes, 1997]. Moreover, it would allow the user the use of various modalities, such as speech or touch, rather than forcing him or her to read a text, which may be disruptive to the main task [Catrambone et al., 2004].

Apart from the potential performance improvements, personified interfaces can also positively affect user experience and the social acceptance of technology. Bickmore and Picard [2005] reported that people are willing to engage in a relationship with ECAs and perceived the relationship more positively with relational agents. Participants created an emotional-bond with an agent. Moreover, similar observations were reported for Human-Robot Interaction [Breazeal and Scassellati, 2000].

Further support for the proponents of ECAs come from King and Ohya [1996] who showed that human forms presented on a computer screen are assessed as more intelligent in comparison with simple geometric forms. However, as Dehn and van Mulken [2000] rightfully noticed, subjects did not interact with a working system and thus the only information available on particular objects was their physical appearance. It should not be surprising that in such a condition, based on the subjects' pre-existing world knowledge, a geometrical object is rated as less intelligent than an anthropomorphized one. A further study conducted by Koda and Maes [1996] indicated that when people have other sources

to base their judgment on, a humanization of physical appearance becomes less important. Participants in their experiment were asked to rate the intelligence of different faces of agents, such as human or animal. In this condition human faces were rated as more intelligent. However, when they were asked to do the same after playing poker against a virtual opponent who was represented by such an agent or when there was no visualization, they rated its intelligence equally.

Moreover, Sproull et al. [1997] reported that subjects rated textual output as having higher attractiveness and friendliness than when it was presented with an ECA face. However, Koda and Maes [1996] found that participants in their experiment liked a visualized poker player more than an invisible one. A possible explanation for these inconsistent findings might be the difference in the particular anthropomorphization chosen, as in the first experiment the agent was in 3D, but in the latter in 2D [Dehn and van Mulken, 2000]. It has been also reported that even small changes in the robot's appearance and voice can affect the user's mental models [Powers and Kiesler, 2006]. Furthermore, in Sproull's experiment, voice was used for communication with the user and they might have based their judgment on that.

Koda and Maes [1996] implied that ECAs are more engaging and well suited for the entertainment domain. Empirical evidence was provided by Takeuchi and Naito [1995], who showed that the virtual card matching game with animated faces is more entertaining than any system in which an arrow visualizes the opponent's moves. However, it is possible that this entertainment value of ECAs is domain-specific. While in the entertainment domain ECAs make the system more entertaining, in the educational field they did not have any effect [van Mulken et al., 1998]. On the other hand, as the author of that paper notes, it can be a result of a task used in an educational condition. The system presented subjects with information and pictures of fictitious employees of a research institute with a pointing arrow or animated face and they had to rate the entertainment of the interface. It is possible to assume that the display of a face is sufficient to improve the entertainment rating of an interface and therefore, since in the task there were pictures of employees presented to all subjects, the rating was already improved by that fact. This explanation would imply that the mere inclusion of a face rather than animation is sufficient for the entertainment advantage. In addition, Lester et al. [1997] did not find differences in entertainment rating

between an agent that gives fully animated advice and a muted agent. It seems that the entertainment domain has been also targeted by industry as personal service robots such as Sony's robot dog Aibo, Violet's robot rabbit Nabaztag or Philip's iCat have become commercially available. In addition, ECAs such as Microsoft Word's Clip have also become available on the public market.

Researchers have also been interested to see how much attention embodied interfaces will gather. Takeuchi and Naito [1995] tracked the eye movements and response times of two human player opponents in order to compare the effect of a face with a 3D arrow in a virtual card matching game. They found that the facial display gathered more eye contact, which presumably meant more attention. At the same time their attraction towards the face distracted them from the main task (card game) and they required more time for a reaction.

ECAs are also relatively popular in education. It is thought that their ability to engage people can be helpful. Pupils exhibited performance gains after interacting with an animated pedagogical agent and enjoyed the learning experience more [Lester et al., 1997]. Consistent results were provided by Takeuchi and Naito [1995], who state that once people are accustomed to synthesized faces they become more efficient and a long partnership provides further performance improvements. Furthermore, Rickel and Johnson [2000] proposed that a pedagogical agent in virtual reality could provide interactive demonstrations, navigational guidance, nonverbal attentional guides and feedback, which would lead to better processing and memorizing of material.

Moreover, an experiment conducted with "Herman the Bug", an animated pedagogical agent, supported these findings as students showed statistically significant gains in the test scores [Lester et al., 1997]. Lester and colleagues have also investigated how different types of agents (muted, verbally capable, animated or fully expressive) affect learning of the material. The biggest improvements in the test scores between pre-test and post-test were observed for animated and fully expressive agents. Nevertheless, all the agents had a positive impact on pupils' performance. Lester et al. referred to it as a 'persona effect' and gave two possible explanations for it. First, they thought that there might be a direct cognitive effect on superior knowledge acquisition – due to the agent's ability to engage students more actively in learning, agents can simulate reflection and self-

explanation. The second explanation relates to the agent's ability to motivate users for a more positive perception of a learning experience itself as part of an interaction with an agent. However, both hypotheses require further research. Moreover, participants in a post-test questionnaire reported that they liked working with Herman the Bug, his advice was accurate, his utility was high and they hoped to have a chance to use it for homework.

The benefits of robots as educators were also extensively investigated. Children with autism in a longitudinal study conducted by Robins et al. [2005] showed improvements in social skills. Moreover, robots have been successfully employed as tour-guides in museums. The visitors rated them highly and they improved the visitors' interest in science technology, as the visitors played more actively not only with the robots, but also with the exhibits [Burgard et al., 1999; Shiomi et al., 2006]. It shows that robots can positively impact pupils' interest in a taught topic.

On the other hand, there are also negative aspects of technology embodiment. It has been suggested that humanizing the interface may induce false mental models of the system. It is possible that human-like appearance or behavior in one area of interaction may lead people to believe that the agent will behave like humans in other areas as well and have human cognitive and emotional capabilities. Due to this generalization, a person might have wrong expectations about the agent's behavior and anticipate from it capacities that it does not possess. Moreover, problems can arise when animations to render an agent livelier do not map onto system behavior. This can happen when system inactivity is visualized by an agent's idle-time movements, such as tapping its foot or looking around. The user might think that the system is currently in the middle of some process as a result of mistaking the agent's activity for system activity. Such a situation would lead to less efficient interaction with the system. [Norman, 1994; Shneiderman and Maes, 1997; Wilson, 1997]

Finally, opponents of animated agents argue that attending to an eye-catching object, such as an agent, might be another source of distraction and require further already limited cognitive resources from a user [Walker et al., 1994]. Wright et al. [1999] noticed that animated graphics can hamper text retention and therefore have a detrimental effect on user performance. People feel less relaxed and confident, and express higher arousal while interacting with a talking-face display in comparison with a text display [Sproull et al.,

1997]. However, they also attributed some of the personality features differently in these two conditions.

Further discussion and a summary of research conducted on ECAs can be found in the article written by Dehn and van Mulken [2000]. While both ECAs and robots are relatively popular in domains such as entertainment and education, considering the advantages and disadvantages of system embodiment presented above one may ask how they differ. The obvious difference is the physical presence of robots in the real world, while ECAs are only displayed on a monitor. In the next sub-chapter I will present research that compares these two technologies and try to find special qualities that are specific for each of them. It is important to answer this question as it is possible that in some situations a physical presence could improve interaction with a user, but in others it could have a negative effect.

2.3 Physical presence

According to Kiesler and Hinds [2004], people perceive robots differently than most other computer technologies. People's mental models of robots are more anthropomorphic in comparison with other systems [Friedman et al., 2003]. It may be a consequence of Science Fiction movies and books, which have shaped human vision of what robots are [Khan, 1998] or a result of perception of autonomous movement [Scholl and Tremoulet, 2000]. Another significant difference is that the majority of the robots will be fully mobile and the interaction will take place in a rapidly changing human environment [Kiesler and Hinds, 2004]. While ECAs are also able to move, their movement is limited by the size of the display on which they are presented. Moreover, as Kiesler and Hinds [2004] state, robots make decisions, learn about themselves and the surrounding world, and impact the information they process and the actions they emit.

In addition, research shows that embodiment is not the same as presence. People are more engaged with others who are present in the real world in comparison with those who are projected [Schmitt et al., 1986]. It has been reported that the human brain processes 2D structures differently than embodied ones [Kawamichi et al., 2005]. If these findings are also valid for technology, it would suggest that ECAs will be perceived differently and they will be less engaging than robots.

Since ECAs and robots differ in certain aspects, it would be interesting to explore these differences and see how physical presence modifies the effect of the system's embodiment. There is relatively little literature focused on such a comparison.

Yamato et al. [2001] investigated how an agent and robot can affect a subject's decision. Users were presented with a series of color squares one at a time on a computer display. Most of the colors were unfamiliar to ordinary people. The subjects were asked to name each color while a computer agent-rabbit or small robot-rabbit gave recommendations using a selection of unfamiliar names. The results showed that the computer agent had a slightly bigger impact on the human selection of color names, but participants felt closer to the robot. The speech of the robot and that of the agent were recognized equally well.

Powers et al. [2007] broadened the area of comparison between embodied agents and robots to social influence, engagement and human disclosure of information during interaction as well as to conversational memory. Participants in their experiment had a conversation with the humanoid head of an agent/robot about their health habits. They communicated with it by typing the answers after each question on a keyboard. Moreover, from time to time the agent/robot asked some sensitive disclosure questions. The time of the conversation, the amount of the information disclosed to the robot about the participant's unhealthy habits, intentions for healthier future behaviors, memory of the conversation details, mental emotional state and attitudes towards the agent or robot were measured.

The results did not show any significant difference of social influence between the robot and its ECA equivalent. Participants were more engaged, enjoyed it more and felt more sense of presence when interacting with the robot. Moreover, they found the robot to be more lifelike and attributed more and stronger personality traits to it. However, they disclosed more information to the computer agent and were able to recall more information from a conversation with it. The results suggest that people were more entertained by the robot and did not pay that much attention to the content of the conversation. These findings are also consistent with Yamato et al. [2001] and give a strong indication that, apart from the social influence, human interaction with an embodied conversational agent differs from that with a robot.

Moreover, in another paper [Kiesler et al., 2008], the authors concluded the results of this experiment as proof of the higher anthropomorphization of a robot compared to an ECA. This might be an important notion, especially considering reports from Nowak and Biocca [2003], who observed that the users feel more highly the presence of low level anthropomorphic systems in comparison with no or high anthropomorphization – the latter when the system’s capabilities do not match its high anthropomorphic image. Since robots are perceived by people as more anthropomorphic, special care should be taken to ensure that their skills will be high enough to meet human mental models of them. It is possible that the robot in Powers et al. [2007] experiment met these conditions and therefore its presence was rated higher than the ECA’s.

In another study, Bartneck [2003] compared the usefulness of ECAs and robots in an ambient intelligent home. Participants played a negotiation game on a flat panel touch screen with eMuu – an emotional Muu robot and its screen based version. Due to the noise of the robot’s motors, participants had lower speech recognition in this condition compared to the agent. However, it did not affect the rating on usability and user control as in both conditions it was similar, which the author interpreted as higher forgiveness for the speech recognition errors in the robot condition. On the other hand, in contrast to Powers et al. [2007], there was no difference in enjoyability of the character between conditions.

Moreover, in the robot condition users acquired higher scores in the negotiation game. At the same time, the joint gain, which is a sum of the user’s and character’s scores, remained unaffected. This finding was interpreted as a social facilitation effect – performing easy tasks better and difficult tasks worse when another person is present. Since it has major implications in the educational domain for the choice of technology, I will discuss it in more detail in the following sub-chapter.

2.4 Social Facilitation

Zajonc [1965] in his social facilitation theory explained that the presence of others serves as a source of arousal. From the Yerkes-Dodson law [Yerkes and Dodson, 1908] we know that arousal increases the likelihood of an organism making habitual or well learned responses. It also explains the improvement of performance on simple tasks and the impairment of performance on complex tasks.

Baron [1986] proposed a cognitive explanation for social facilitation. He suggested that attention conflict between the task and observer can facilitate simple tasks and impair complex ones. This conflict can be triggered if the distraction is very interesting or hard to ignore, there is a pressure to complete the task quickly and accurately, and it is hard or impossible to attend the task and distracter at the same time. The current view in social psychology is that both arousal and cognitive processes influence social facilitation [Aiello and Douthitt, 2001].

From the CASA paradigm we know that humans tend to express the same social responses while interacting with technology as with other humans. Therefore, it should not be surprising that ECAs can produce a social facilitation effect [Hall and Henningsen, 2008; Park and Catrambone, 2007]. Moreover, since Yamato et al. [2001] and Powers et al. [2007] reported that people felt more highly the presence of a robot and were more engaged by it in comparison with an ECA, we can understand better why in Bartneck's experiment [2003] the robot induced a stronger social facilitation effect. It might have important consequences for the choice of technology in education as a robot may improve a pupil's performance with well trained and easy tasks, while an ECA may be a better choice for new and complicated tasks.

The research reviewed above clearly demonstrates that ECAs and robots may have an impact on the user's task performance and social perception of the interaction. Moreover, while embodiment of both technologies brings certain advantages and disadvantages, there is a gap in knowledge about the specific qualities of each of them. With a special focus on the educational domain, this study will try to answer the following hypotheses:

- H1: Participants will be more engaged, perform more, better and faster easy tasks during interaction with a robot than with an ECA.
- H2: Participants will perceive differently the difficulty and interest of tasks, and the comfortability of interaction when a robot is present compared to when an ECA is present.

3. Methods

The experiment on ECAs' and robots' impact on users' task performance was conducted in a usability lab at the University of Tampere. The design was between-subjects. Subjects were either interacting with a robot or an ECA. The participants' answers for a mathematical task and response times were recorded. Moreover, before the experiment began, informed consent and a questionnaire with demographic data were collected.

The details regarding the participants, equipment, procedure and variables used during the experiment are explained in this chapter. Each sub-chapter is dedicated to a different topic. In sub-chapter 3.1 detailed information about the participants is presented. Sub-chapter 3.2 is focused on apparatus and software used during the experiment. Sub-chapter 3.3 presents the experimental set-up. Variables and measures controlled during the experiment are presented in sub-chapter 3.4.

3.1 Participants

In total 16 participants (11 male and 5 female) volunteered for the experiment. Participants were 26 years old on average, with age range between 20-32 years. They had different origins, representing in total 9 countries (Finland being the most represented with 5 representatives) mainly from Europe. Nevertheless, all of them at the time of the experiment were living in Tampere (Finland). All but 2 participants were students of one of the educational institutions in the city. Since the instructions, post-test questionnaire and the robot's/agent's feedback were in English, all the participants were competent English language speakers and 2 of them were native speakers. However, as explained later in this paper, the user's main task was mathematical; therefore competent non-native speakers were not handicapped by the language barrier.

3.2 Equipment

Apparatus used in this experiment is presented in the following sub-sub-chapters.

3.2.1 Nabaztag

Robot Nabaztag was used in the robot's condition. Nabaztag is commercially available and sold by Violet The Smart Object Company. It is a robot-rabbit that can connect to the internet via Wi-Fi. It has 2 interchangeable magnetic ears. Its height, excluding ears, is 16 cm and, including ears 23 cm. Diameter of the base is 13.5 cm. It is equipped with a microphone for input and a speaker for output. Using wireless network, it can send and receive files. It is capable of playing audio files and has text to speech functionality. Moreover, it can move its ears and change color of 4 LED lights placed above the speaker (4 colors are available).



Figure 1. Nabaztag robot-rabbit.

For this experiment special audio files were created and they were played by Nabaztag and ECA. They included 6 feedback messages (Appendix A), which Nabaztag was providing to the users based on their task performance. There were 3 positive and 3 negative statements. All of them were short messages that tried to encourage the participants to continue the task. They were created using publicly available online text to speech technology from AT&T Labs (<http://www2.research.att.com/~ttsweb/tts/demo.php>). US English female's voice (Claire) was used in this experiment. Default wav audio files

were converted to mp3 files to decrease their size and improve their transfer time to Nabaztag.

Before an audio file was played, Nabaztag's one pulsating red light was signaling that an audio file is being transferred and it will soon start speaking. Usually it took approximately 2 seconds for an audio file to be received and started. To accelerate the speed of the robot's responses, Nabaztag did not move its ears before speaking as that would require additional time.

3.2.2 jNabServer

The communication with Nabaztag was done using jNabServer – open-source server software, which handles communication with Nabaztag. It was used instead of the default Violet's servers, which provide services for Nabaztag, due to shorter communication time required by jNabServer. It also allowed not having an internet connection. jNabServer is written in j2se 6.0 (Java Standard edition) and can be downloaded from: <http://www.cs.uta.fi/hci/spi/jnabserver/#>. It was developed by Juha-Pekka Rajaniemi and Ville Antila. Its version 1.01, used in this experiment, has been updated by Jaakko Hakulinen. Currently, there is a version 2.0 of jNabServer available. However, version 1.01 was chosen for this experiment since the latest does not have any documentation available.

jNabServer is built on top of a lightweight HTTP-server. jNabServer does not provide services, but it enables communication, which can be done by developing plugins. The communication between Nabaztag and jNabServer is done via cycles of requests. Each cycle of HTTP request starts from Nabaztag. It asks for one of possible files from jNabServer. [Rajaniemi, 2007]

More technical details and instructions on how to create plugins can be found at: <http://www.cs.uta.fi/hci/spi/jnabserver/#> under Documentation section and in the thesis of Rajaniemi [2007]. In this experiment, default BootPlugin was modified for the communication with Nabaztag. The modified BootPlugin was not changing current plugin into DefaultPlugin. To ensure fastest possible response times from it, the ping interval on which Nabaztag was sending requests to the server was set to its minimum – 1 second.

Moreover, since all the communication with Nabaztag was done using BootPlugin, the robot did not move its ears before speaking as fast feedback from the robot was desired. Furthermore, on first Nabaztag's connection with the server after its reboot or power on, the robot needs to receive a BootCode. Since it takes some time, it was done before the experiment began to ensure the same response times for all participants.

3.2.3 Embodied Conversational Agent

The ECA used in this experiment was an animated gif image created from a photo of Nabaztag. To display, in a similar manner as Nabaztag does, that an audio file is being received (what also meant that the agent is going to speak), a red flashing circle was displayed on the agent's body. The ECA was displayed on 19 inches LCD display. The area outside of the agent's window was all black. Since there is some literature which suggests that the size can be a factor in social perception [Huang et al., 2002; Judge and Cable, 2004], special care was taken to ensure that the size of the ECA was similar to the size of real Nabaztag robot.



Figure 2. Embodied Conversational Agent version of Nabaztag robot with blinking red light.

Due to the system's design, in the robot's condition it was Nabaztag that was requesting file from jNabServer on 1 second intervals. Therefore, the response times could have differed depending on when a request was sent by maximum of 1 second. Moreover, there was also time needed for an audio file to be received by Nabaztag. In total, the delay between the command sent to Nabaztag to speak and the time it actually started speaking was approximately 2 seconds. In the virtual agent's condition, the command and execution to play an audio file could have been processed instantly. Moreover, using loudspeakers could have resulted in different quality and volume compared to Nabaztag's speakers. That could have affected subjects' perception of the system as the agent would be more responsive than the robot. To make both conditions equal, Nabaztag was placed between loudspeakers and behind the LCD screen on which the agent was displayed, with the robot's speaker located under that display. The robot was not visible to the participants. Therefore, in reality Nabaztag controlled via jNabServer was used for playing audio files in both conditions.

3.2.4 Computer application

In the present study participants were doing a task on a computer. Java application was used to implement the task and the post-test questionnaire. The operating system used in the test was Windows XP. The size of the screen was 19 inches with screen resolution 1280x1024.

The user's task was to solve series of Gauss modular arithmetic statements, such as $50 \equiv 38 \pmod{4}$. To solve the problem, participant had to subtract a middle number (i.e. 38) from the first number (i.e. $50 - 38$), and then the result of this (i.e. 12) was divided by the last number (i.e. $12/4$). If the dividend was a whole number (as here 3), then the statement was true. On the contrary, if the dividend was a decimal number, then the statement was false. In the present study, all the numbers in statements were randomized, the first two from range 1-99 and the third 1-9.

Beilock et al. [2004] suggested modular arithmetic as advantageous for laboratory experiments as it is rather unusual and therefore its learning history can be controlled. Moreover, it was chosen for this experiment as it is a relatively easy task and, due to its

repetitiveness, also potentially boring task. That coupled with a possibility to end the task faster, allowed to see whether the robot or the ECA can motivate participants to continue doing the task despite lack of any benefits to do so.

The application for the task was written in Java Standard Edition 6. Components from Swing library were used in the GUI. The application had one field where modular arithmetic statements were displayed and 3 buttons: two for defining whether presented statement was true or false, and one for ending the task and going directly to the post-test questionnaire. When the subject pressed “True” or “False” button, new modular arithmetic statement appeared. After every three correctly solved problems, the buttons were temporarily disabled and a command was sent to Nabaztag to play on random one of 3 positive feedback audio files. Moreover, additional button “Refresh” appeared and subjects were asked to press it after Nabaztag finished speaking to see a following modular arithmetic statement. Similar logic was applied in case a user answered three times incorrectly, when instead of a positive statement, one of negative feedback was played. If a participant did not decide to end the task by pressing “End the task” button, after 10 minutes the task has automatically ended and the post-test questionnaire popped-up.

The questionnaire included ten 5-point Likert scale statements (Appendix B) regarding the task, interaction and the robot/agent, such as “I liked Nabaztag” or “The task was easy”. Subject had to choose how much does he agree or disagree with the statements.

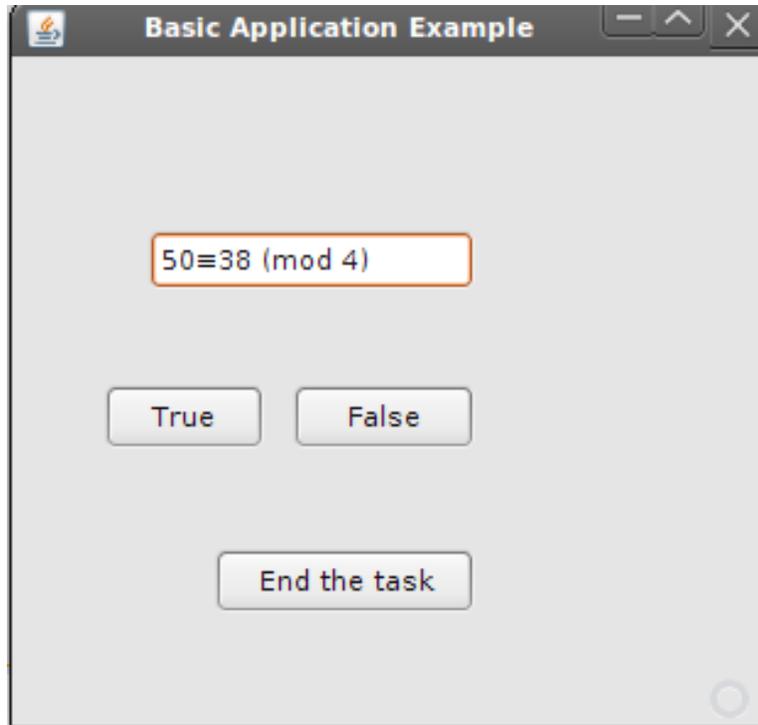


Figure 3. Main window of a computer application for modular arithmetic task.

Apart from the replies for the post-test questionnaire, the application recorded the time of a new statement appearing and a user's selection of statement's correctness, whether a user's answer was correct and type of feedback provided by Nabaztag.

3.3 Procedure

Each participant entered the laboratory together with an experimenter and was placed in front of a desk with a computer screen, keyboard and mouse. Moreover, above the computer screen, Nabaztag or another display with the ECA was placed, depending on the experimental condition. The robot or the agent was directed towards the participant.

At the beginning of the experiment, an informed consent and a demographic data were collected by the experimenter. After this procedure was done, the experimenter gave a printed version of instructions to the participant (Appendix C) and explained the task. Subjects were asked to perform, as fast and as accurately as possible, a series of mathematical tasks (modular arithmetic) on the computer. They were informed that during

the process, robot-rabbit Nabaztag (introduced the same way in both conditions) will regularly give them feedback on their task performance.

Moreover, participants received instructions on what is modular arithmetic statement and how to judge whether it is true or false. They were also informed how to use the application designed for this task by pressing “True” or “False” buttons. Furthermore, they were told that after 10 minutes the questionnaire will pop-up, but they can end the mathematical task earlier by pressing “End the task” button and go directly to the questionnaire. Finally, they were told that, after Nabaztag will end speaking, they will need to press “Refresh” button for a new statement to appear. Since subjects’ understanding of the mathematical task was crucial in this experiment, special care was taken to ensure that participants understood clearly how to solve it.

After subjects read the instructions of the experiment, they were told that they can start whenever they are ready by opening the application located on the desktop and the experimenter left the room to ensure that the potential social facilitation effect could not be attributed to his presence or affect participants’ willingness to end the task faster by going directly to the questionnaire. In addition, participants were asked to come out from the room whenever they completed the experiment.

3.4 Variables

Subjects were randomly assigned to the robot’s or the ECA’s condition. In table 1, dependent variables and measures that were recorded and analyzed, are presented.

Table 1. Measures in the experiment.

Variable	Measure
Task performance	Amount of mathematical problems solved
	Amount of mathematical problems solved correctly
	Speed of solving a problem after Nabaztag’s feedback
	Self-report: Confidence and focus on task
Social acceptance and User experience	Self-report: Conformability, entertainment and attachment

	Amount of time spent on the task
Task perception	Self-report: Task perception

The statistical software package SPSS was used for analyzing the data. The results obtained in this experiment are presented in the following chapter.

4. Results

Tests of the hypotheses were conducted using statistical analyses to determine the impact of the robot and the ECA on users. It was done in 2 steps. First the data was transferred to Microsoft Excel where the average response times were calculated for each user from a single response times collected during the experiment. Moreover, the average times of user's responses were also calculated for the first 5 and the last 5 statements. In the second step the data was transferred to SPSS statistical package together with the results of the post-test questionnaire for further statistical analysis.

Due to relatively small samples examined during this experiment, special care was taken to ensure that the statistical test assumptions were met. As Likert scale used in the post-test questionnaire cannot guarantee to be on interval level of measurement and Shapiro-Wilk test showed that most of the items do not have normal distribution, the self-report was analyzed using non-parametric test – Mann-Whitney U. Normal distribution of the data of users' response times displayed on histograms and in Shapiro-Wilk test allowed analyzing this part of the data using independent and dependent samples t-tests. All the collected data was regarded as valid and analyzed. The results of this analysis are presented below separately for each variable: Task perception, Task performance, Social acceptance and User experience. They are discussed in relation to a previous research in the following chapter.

4.1 Task perception

The chosen mathematical task was supposed to be easy and boring for some of the assumptions taken for hypotheses direction in this experiment. Therefore, it was important to see how participants perceived the task and whether this perception was affected by the experimental condition. Contrary to the experimental assumptions, participants disagreed that the task was easy. On a 5-point Likert scale where 1 was “Strongly disagree” to 5 – “Strongly agree” the mean was $M = 2.31$, $SD = 1.01$. Moreover, the condition did not affect perception of the task difficulty ($U = 25.5$, $N1 = N2 = 8$, $p = .43$) the robot's $M = 2.13$ ($SD = .83$) versus the agent's condition $M = 2.5$ ($SD = 1.2$). However, relatively high standard

deviation can suggest that some participants thought about the task as difficult when others as easy.

As expected, the task was perceived as rather boring ($M = 3.56$, $SD = 0.96$). Neither the robot ($M = 3.38$, $SD = 1.06$) nor the ECA ($M = 3.75$, $SD = .89$) influenced the results $U = 25.5$ $p = .43$.

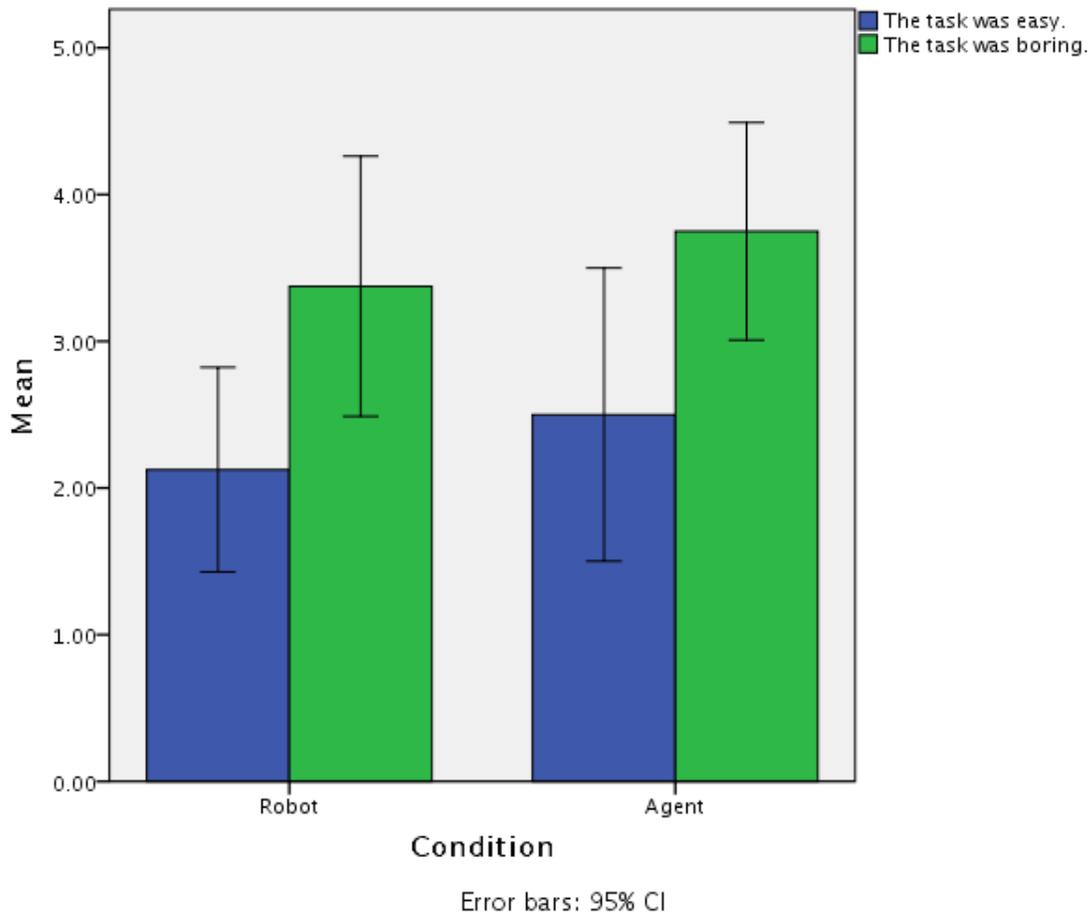


Figure 4. Ratings of perceived task difficulty and boringness in different experimental conditions.

4.2 Social acceptance and User experience

It was predicted that participants would be more engaged in doing mathematical task by the robot than the ECA. Engagement was measured by the amount of time spent by subjects on the task and a self-report question in the post-test questionnaire. Subjects were able to

continue modular arithmetic task for 10 minutes or stop it at any time without consequences. Since the task itself was not very attractive, it was expected that participants will end it sooner than the maximum time allowed. However, the results did not provide any support as all the participants in both conditions proceeded with the task for full 10 minutes.

In addition, there was no statistically significant difference ($U = 19.5$, $N1 = N2 = 8$, $p = .13$) in perceived entertainment between both conditions. Nevertheless, subjects showed a trend to be more entertained by the robot ($M = 4.13$, $SD = 0.99$) than the ECA ($M = 3.63$, $SD = 0.74$). Similarly while no difference in having fun was observed between the conditions $U = 4.25$, $N1 = N2 = p = .46$, a trend in the same direction could be noticed with means respectively $M = 4.25$ ($SD = .46$) and $M = 3.63$ ($SD = 1.06$).

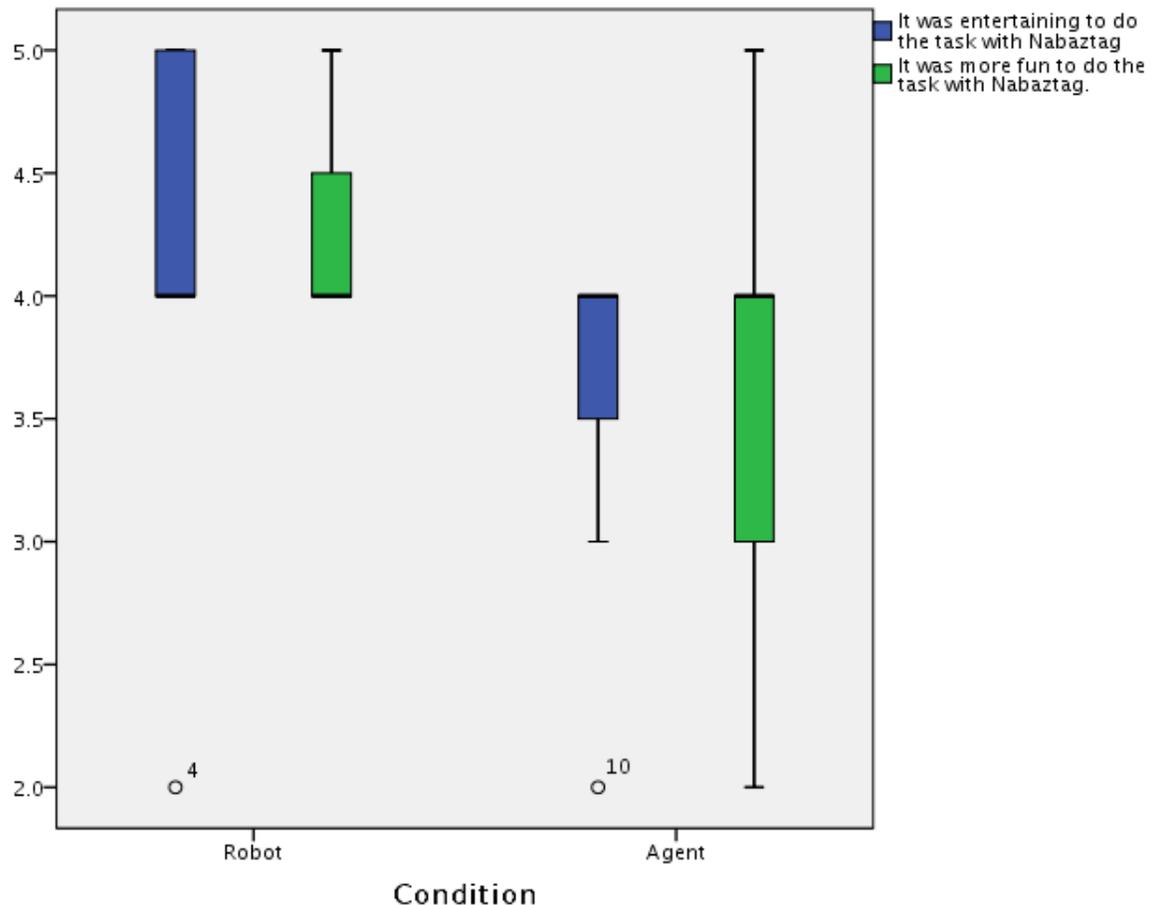


Figure 5. Rating of the robot's and the robot-like agent enjoyability.

Participants, who were asked whether they felt comfortable with Nabaztag performing the task with them, had indifferent opinion when it was the robot ($M = 3.25$, $SD = .89$) or the ECA ($M = 3.5$, $SD = 1.2$) – $U = 28$, $N1 = N2 = 8$, $p = .66$.

Moreover, there was no statistically significant difference in liking of Nabaztag in both conditions $U = 27$, $N1 = N2 = 8$, $p = .54$. Both Nabaztag robot ($M = 3.88$, $SD = .99$) and Nabaztag agent ($M = 3.75$, $SD = .46$) were relatively liked. However, predicted higher liking of the robot was not confirmed.

In addition, contrary to assumption that physical presence in the real world will increase a feeling of presence, participants were indifferent in the agent's ($M = 3.63$, $SD = 1.06$) and the robot's condition ($M = 3.38$, $SD = 1.19$) $U = 25$, $N1 = N2 = 8$, $p = .4$.

Furthermore, participants had indifferent opinion about Nabaztag's feedback being irritating. Respectively $M = 3.75$ ($SD = .89$) and $M = 2.25$ ($SD = 1.04$). There was no statistically significant difference $U = 28$, $N1 = N2 = 8$, $p = .64$; however there was a trend observed that the robot's repetitive feedback was seen as less irritating than the agent's.

4.3 Task performance

Time required by subjects to solve modular arithmetic problems was analyzed to see the robot's and the agent's impact on task performance. Participants interacting with the robot solved on average $M = 44.75$ tasks ($SD = 25.42$), while these interacting with the ECA $M = 57.38$ tasks ($SD = 14.27$). T-test for independent samples did not show any statistically significant differences between the conditions $t(14) = -1.23$, $p = .24$.

However, there was a significant difference in times required to solve one mathematical problem during interaction with Nabaztag $t(8.04) = 1.84$, $p = .1$. Participants interacting with the robot needed more time ($M = 15.69$ sec, $SD = 10.51$) to solve each problem than these interacting with the agent ($M = 8.6$ sec, $SD = 2.87$). In addition, the same analysis was done for the first 5 and the last 5 statements. There was a statistically significant difference between the robot's ($M = 20.58$, $SD = 11.86$) and the agent's ($M = 12.35$, $SD = 4.51$) conditions during the first 5 modular arithmetic problems; $t(8.98) =$

1.83, $p = .1$. However, during the last 5, no significant difference was observed, respectively $M = 14.58$ ($SD = 11.76$) and $M = 7.85$ ($SD = 3.35$); $t(8.13) = 1.56$, $p = .16$.

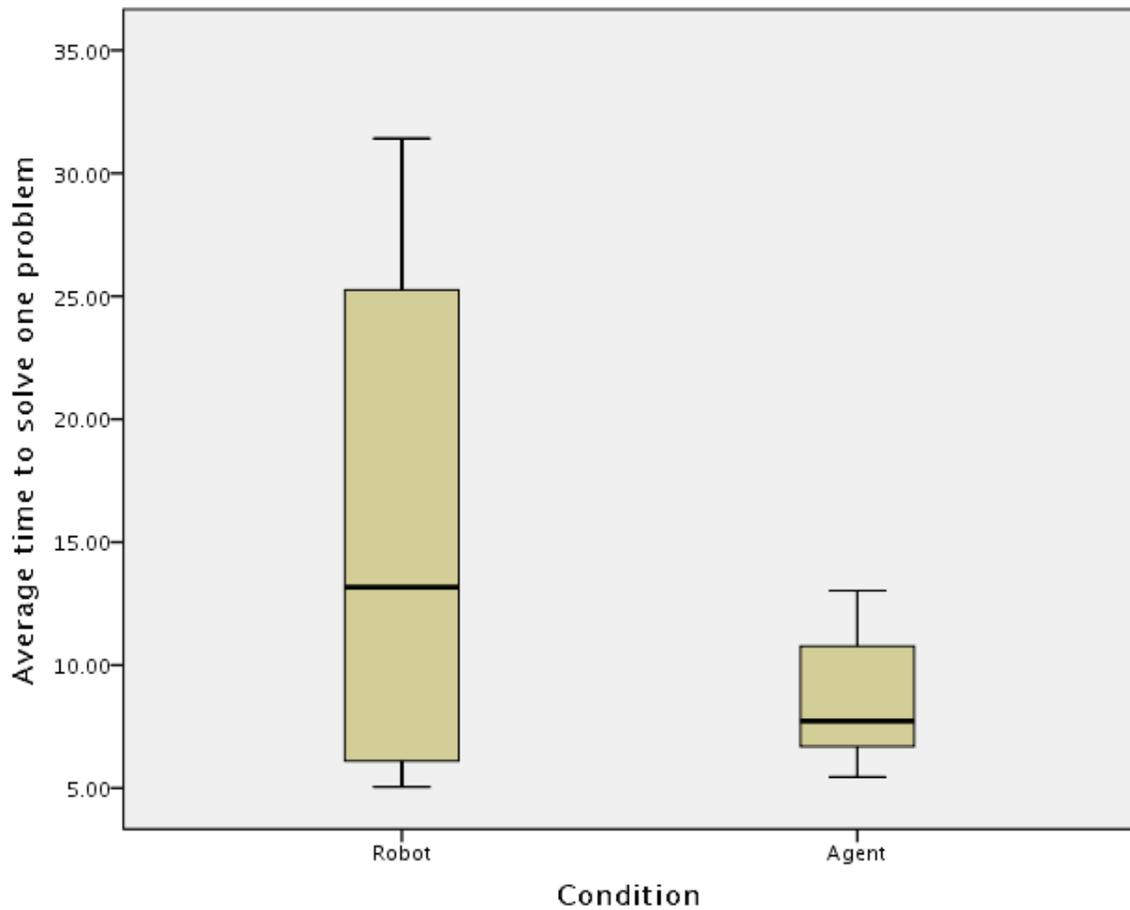


Figure 6. Average time (in seconds) spent on modular arithmetic task in the robot's and the agent's conditions.

T-test for matched pairs showed that in both conditions participants' response times decreased comparing the first and the last 5 problems solved. In the robot's condition $t(7) = 2.11$, $p = .07$ and the agent's $t(7) = 5.11$, $p = .001$.

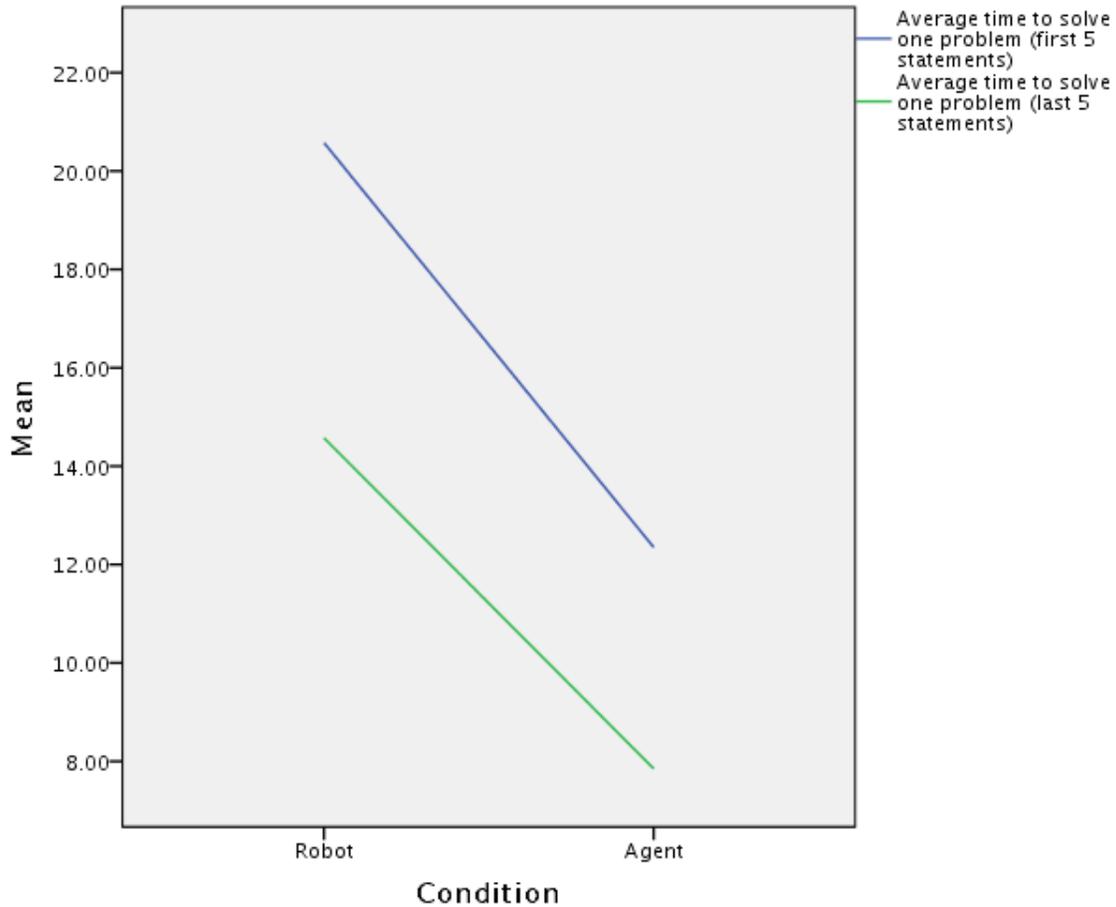


Figure 7. Comparison of times required to solve the first 5 and the last 5 mathematical problems grouped by the experimental condition.

Moreover, correctness of the answers provided by subjects was analyzed. Participants very rarely made mistakes: in the robot's condition $M = 2.5$ ($SD = 1.31$), in the agent's condition $M = 1.88$ ($SD = 1.73$). No significant differences were found between the conditions $t(14) = .82$, $p = .43$. As a result, in both conditions participants almost did not hear any negative feedback, respectively $M = .38$ ($SD = .52$) and $M = .13$ ($SD = .35$). Therefore, no further analysis to compare impact of negative and positive feedback was conducted.

Nabaztag's feedback had little impact on responses following it. No significant differences were found in users' response times for statements displayed just after Nabaztag's feedback $t(14) = 1.38$, $p = .19$. Participants interacting with the robot required

$M = 12.88$ sec ($SD = 7.67$) and with the agent $M = 8.74$ sec ($SD = 3.6$). Comparison of the first 5 feedback messages resulted in marginally significant difference, respectively $M = 12.85$ ($SD = 7.67$) and $M = 7.8$ ($SD = 2.98$); $t(14) = 1.74$, $p = .11$.

Since after each time Nabaztag spoke, participants had to press “Refresh” button to see a new statement, the time that elapsed after Nabaztag’s message was compared between the conditions. No significant differences were found between the ECA ($M = 2.92$, $SD = 1.18$) and the robot ($M = 3.03$, $SD = .77$); $t(14) = .22$, $p = .83$. In addition, same comparison for the first 5 feedback brought similar results, respectively $M = 3.15$ ($SD = .88$) and $M = 3.48$ ($SD = 1.33$); $t(14) = -.58$, $p = .57$.

Moreover, subjects’ perception of Nabaztag’s impact on the task performance was analyzed. Participants believed that Nabaztag’s presence helped them to focus on the task – the robot $M = 4$ ($SD = 1.07$) and the agent $M = 3.75$ ($SD = 1.04$). There was no statistically significant difference between conditions $U = 27$, $N1 = N2 = 8$, $p = .58$. In addition, Nabaztag’s feedback was perceived as having no effect on the task performance $U = 20$, $N1 = N2 = 8$, $p = .17$. Mean for people interacting with the robot $M = 2.75$ ($SD = .89$) and the agent $M = 3.38$ ($SD = .92$).

5. Discussion

The results presented in the previous chapter are discussed in this section. They are presented in a relation to the other papers which explored the same area. Since all the results obtained in this experiment were marginally significantly different or only non significant trends, it is important to keep that in mind when interpreting the results or drawing any conclusions.

5.1 Task performance

One of the assumptions of this experiment was that the task given to participants was easy and due to its repeatability also boring. While the post-test questionnaire confirmed that the task was not interesting, contrary to the experimenter's predictions participants rated the task as relatively difficult. Since it was assumed that a robot will induce stronger social facilitation effect than an ECA, perception of high task difficulty would also reverse the direction in H1 of Nabaztag's impact on the task performance - presence of a robot would impair participant's performance (in difficult tasks). The results of this experiment are in opposite direction than H1 predicted, with participants who were interacting with the agent solving the task faster than these who were receiving feedback from the robot and consistent with the above explanation. Participants spent almost twice as much time on single modular arithmetic problem in the robot's ($M = 15.69$ sec) than in the agent's condition ($M = 8.6$ sec). However, very high standard deviation ($SD = 10.51$) in the former condition shows high variance and the result should be taken with special caution.

These findings could support Powers et al. [2007] and Kiesler et al. [2008] conclusions that presence of a robot lures user's attention away from a main task more and results in worse performance than presence of an ECA. While these authors did not interpret their results as an example of the social facilitation effect, Bartneck [2003] suggested it as a potential explanation in his paper. However, it was the robot who improved people's task performance more than the agent. Unfortunately, perceived task difficulty was not measured in that experiment. To interpret, in the context of the social facilitation effect, Bartneck's [2003] finding and the result of the current experiment, both

would be consistent with the social facilitation effect if Bartneck used an easy task and the current experiment a difficult task. While the latter part of this assumption was reported in this experiment, future study should answer the question how robots and agents affect users' performance on easy tasks.

Further analysis showed that the difference in times required for solving modular arithmetic problems was bigger at the beginning of the interaction as it was statistically significant for the first 5 statements. When only times of solving the last 5 statements were analyzed, there was no statistically significant difference. Although we can see a trend in the same direction as for the first 5 statements with subjects solving the task faster in the agent's condition. It could be interpreted as a robot slowing down participants more at the beginning, as they were paying more attention to it. Perhaps, due to its presence in the real world, which potentially could be more dangerous for the users if a robot started behaving unexpectedly than if it was an ECA displayed on a computer screen.

Furthermore, novelty factor can be responsible for these differences. While both ECAs and robots are relatively new technologies, there is no doubt that despite participants coming from different countries, they had less chance to see a robot in their environment than a computer agent. Therefore, it is possible that they were observing the robot at the beginning of the experiment because they were curious what it can do. It would also explain why towards the end of the experiment, when Nabaztag started repeating its feedback, they focused more on the task and paid less attention to the robot/agent and there was no difference between the conditions.

This potential explanation for the results can be also supported by the finding that it took participants more time to solve a problem after Nabaztag's feedback in the robot's condition, only after the first 5 feedback messages. When subjects were accustomed to feedback messages, they possibly stopped listening to them and provided feedback had little impact on their performance.

At the same time, it is important to note that in both conditions the time required to solve a problem decreased during the course of the experiment. Unfortunately, due to the lack of a control group it is impossible to say how much of this improvement can be attributed to the robot's or the agent's feedback and what is simply a result of task

learnability. Nevertheless, it is a promising starting point for utilizing robots and computer agents in education as they both improved subjects' task performance.

In spite of participants rating the task as relatively difficult, the results suggest that they performed it very well. On average participants made less than 2.5 mistakes in either condition, which means that the majority of them did not even hear negative feedback messages from Nabaztag. Since they were asked before the experiment to solve problems as fast and as accurately as they can, it seems that they spent additional time to ensure the accuracy even if it resulted in longer task times. Moreover, no statistically significant difference was observed in the amount of solved problems in both conditions. It is very interesting and hard to interpret finding considering that the time required to solve a modular arithmetic task in the agent's condition was lower than in the robot's condition.

Since the task was perceived as boring it was assumed that participants would not perform it for the whole duration of the experiment, as they were informed that they can stop it at any time and go directly to the post-test questionnaire without any consequences. It was predicted in H1 that assumed higher entertainment value of a robot could convince them to perform the task for a longer period of time. Since all participants in both conditions decided to do the task for the maximum duration of time allowed (10 minutes), it seems that this measure was not valid in this experiment. This finding could also imply that both the robot and the agent were so engaging that participants continued the task for the maximum amount of time.

However, it seems highly improbable, as due to the limited amount of feedback messages provided by Nabaztag, users heard them many times and could not have expected that something new would happen. Alternative and more convincing explanation is that participants misunderstood instructions given before the experiment. They were informed that they can stop participation at any time. Moreover, later on, when the task was explained, experimenter said that if they want to stop the task and go directly to the questionnaire they can do so by pressing "End the task" button. It is possible that participants interpreted it as a form of opting out from the experiment and decided to continue the task longer than they would have otherwise wished.

5.2 Perceived impact of Nabaztag

When participants were asked in the post-test questionnaire whether Nabaztag's feedback made them more confident in doing the task, they had indifferent opinion in both conditions. Since the task lasted for 10 minutes and participants heard the same feedback many times, it is very probable that at some point they stopped listening to it. Moreover, since they were solving problems almost always correctly, the feedback provided was positive. It is possible that participants, despite feedback being 100% accurate, started having doubts if it is really correct or it is just a random positive message played without checking their answer. While there was no question in the post-test questionnaire devoted to this problem, many participants asked the experimenter about the accuracy of provided feedback after the experiment. However, as it was not scientifically measured, no conclusions can be drawn. Nevertheless, it could be an interesting question to answer in future research as it would impose that even accurate robots and ECAs might be unable to prevent users from doubting their feedback.

It is possible that this lack of impact of the robot's or the agent's feedback could be responsible for subjects' indifferent opinion about comfortability of Nabaztag performing the task with them. Since they did not see any influence of what it was saying, it should not be very surprising that they did not mind whether Nabaztag was present when solving the task.

As suggested by Bartneck [2003], participants showed high forgiveness for the robot's repetitive feedback. There was no statistically significant difference between conditions in perceiving the feedback as irritating, but we can notice a trend for rating the robot's feedback as less irritating than the agent's. If this trend in future research with bigger samples becomes statistically significant, that would be an indication that users are keener to forgive robots their imperfection rather than ECAs. Therefore, it would mean that a choice between a computer agent and a robot should take into consideration systems reliability.

On the other hand, subjects declared that Nabaztag helped them focus on the task. This finding is very important since it is a crucial factor for potential use of robots or ECAs in education where the system should not distract users from the main task. It is possible

that the motivational factor of the robot/agent observing users' task performance led them to focus more on the task as proposed by Lester et al. [1997].

5.3 Entertainment

While no statistically significant differences were found between conditions for the entertainment and fun values of performing the task with the robot or the ECA, we can observe a trend for rating the robot higher on both scales. This trend is in the same direction as reported by Powers et al. [2007] as subjects enjoyed more interaction with a robot. Nevertheless, this trend was not statistically significant and considering that Bartneck [2003] did not find any differences between these conditions, further research should be conducted to investigate entertainment value of both systems. Moreover, it is also worth noting that both, the agent and the robot, were highly entertaining, which is a very promising result for their potential use as attention gathering devices in various environments.

In addition, in spite of very limited interaction, both the agent and the robot managed to build a rapport with participants and were relatively liked by them. It would suggest that building positive relationship with users can be relatively easy even with very limited capabilities of technology. This result, together with previously mentioned high entertainment and fun of interaction with the robot and the ECA, further supports Koda and Maes [1996] conclusion that ECAs are well suited for the entertainment domain. Moreover, it implies that robots' entertaining values are equally high and they also could be successfully employed in that domain.

However, it is also possible that the "cute" appearance of the robot/agent used in this experiment was the main factor for high liking of Nabaztag and if another pair of a robot/agent, with different physical appearance, was used in this experiment, the results could have not been so encouraging. Since even small changes in a robot's (and probably an agent's) appearance can affect users' mental models [Powers and Kiesler, 2006] it is possible that the conclusions being discussed cannot be generalized for other robots or ECAs. There is a strong evidence that people might create wrong expectations regarding capabilities of technology due to its human-like appearance [Norman, 1994; Shneiderman

and Maes, 1997; Wilson, 1997]. Since the robot's/agent's "sweet" animal-like appearance used in this experiment could have led to lower expectations regarding its capabilities and its very limited interaction capabilities met subjects' mental models, participants built a positive relationship with the system. However, if a humanoid robot was used in this experiment, it is possible that people could have expected much more sophisticated communication capabilities from it, which not met by it, could lead to users' frustration and lower affection towards it. Nevertheless, the experiment presented in this paper gives a strong indication that by choosing a proper physical appearance for a robot or an ECA we can easily create a positive rapport between users and technology.

Possibly a bad wording in the question about feeling Nabaztag's presence is responsible for indifferent answers. It was not defined what "presence" means in this experiment as the word can refer to a mental or physical perception. Since participants did not know how to answer this question, they opted for neither agreeing nor disagreeing response.

6. Conclusions

There is a visible trend in technology for making user interfaces resemble living creatures. It is believed that by providing them natural human communication channels, such as speech or touch, we can improve quality of an interaction and make them more accessible for ordinary users. Moreover, human or animal appearance of technology could make it easier for users to relate to it compared with text-based interfaces of the past. Two greatest examples of this trend are Embodied Conversational Agents and robots. The aim of introducing these autonomous agents is not only to support people in different tasks, but also to make them capable of doing things without human supervision. One of domains where robots or ECAs would mean great improvement is education. Since they could be personal and do not have limits of current human teachers, who can be present in only one place at a time, they could monitor students' progress continuously.

The objective of this study was to compare special qualities of ECAs and robots and see their impact on users' task performance. An experiment was conducted in which participants were solving modular arithmetic problems and receiving feedback on their task performance from the rabbit-like robot or the ECA. Sixteen participants took part in this study. In addition to the task performance, answers of the post-test questionnaire were analyzed.

Statistical analysis was carried out using independent samples and matched pairs t-tests and Mann-Whitney U test. The most interesting results are summarized below. Due to small samples used in this experiment the results were marginally statistically significant or not significant. Nevertheless, few interesting trends were observed.

No difference was found in the amount of problems solved by participants in both conditions. However, participants interacting with the robot needed on average almost twice as much time to solve one modular arithmetic problem compared with the ECA. The difference was especially visible at the beginning and vanished towards the latter part of the experiment. Moreover, also only in the first part of the experiment Nabaztag's feedback had impact on the speed of solving the task. There are two potential explanations for these results. Powers et al. [2007] implied that people interacting with robots are focused more on them rather than a main task compared to ECAs. It is also possible that the social

facilitation effect is responsible for these differences [Bartneck, 2003]. It would also mean that robots invoke stronger social facilitation effect than ECAs.

While unexpectedly participants declared the task as relatively difficult, which meant that the direction of H1 was reversed, they performed it very well hardly making any mistakes. Moreover, what is very important for the use of robots and ECAs in the educational domain, in both conditions users significantly decreased required time to solve one modular arithmetic task during the course of the experiment. In other words, they have learnt to solve problems faster after receiving supporting feedback from Nabaztag. Unfortunately, due to the lack of a control group it is impossible to say how big impact Nabaztag's feedback had and what was a result of simply learning better techniques for solving the task, which would also occur in absence of Nabaztag. Nevertheless, participants declared that the robot/agent helped them to focus on the main task.

These findings have important practical consequences. Both ECAs and robots seem to be well suited for the educational domain since subjects showed improvements in the speed of solving modular arithmetic problems. However, a choice between these 2 technologies should be based on how well practiced material is processed by users. If robots induce stronger social facilitation effect, they should be used in situations when users are doing well trained tasks, as a robot's presence will result in better task performance than an ECA's. On the other hand, if a task is difficult or not well trained, a robot's presence would impair user's task performance. Similarly, a choice between utilizing a robot or an ECA at work should be based on the same criteria.

Moreover, in both conditions participants enjoyed interaction with Nabaztag. It is also very important as it confirms high value of these technologies in the entertainment domain. Furthermore, it is also positive information for a potential use of them in education, since pupils who enjoyed interaction with a robot or an ECA will also have more positive perception of their study experience and spend more time on it. By seeing it as a fun, they will probably also process material better than if they were just reading it.

Finally, the last important finding of this experiment with practical implications is that people tend to be more forgiving for robots than ECAs for their imperfections. In both conditions the same repetitive feedback was used. However, only in the agent's condition it

was perceived as irritating. While this difference was not statistically significant, there was a strong trend in favor of the robot. Considering that currently robots' technology is still in its infancy, it is possible that people will be more understanding in case system fails in some situations. This would be also very important as it could accelerate the social acceptance of robots and foster their development. Moreover, it is another aspect for consideration when making a choice between robots and ECAs in education, entertainment or any other sector as users require flawless and sophisticated capabilities from agents.

On the other hand, there were some methodological flaws in the design of presented experiment. Any future research should include bigger samples. Small samples used in this experiment had less statistical power and the results were only marginally statistically significant or non-significant trends. Moreover, it increased the probability of individual differences influencing the results. It is hard to draw any solid conclusions from such results and future research should answer the question whether real differences exist.

Furthermore, boredom effect could have affected subjects' answers to the post-test questionnaire. Since there was very limited amount of feedback provided by Nabaztag, it is possible that participants got tired of them and, in both conditions, rated the impact of the robot/agent lower than they would have if the question was asked earlier. A lack of negative feedback, due to participants' superior task performance, made it also impossible to see how they would have responded to negative messages. It would be interesting to see whether negative feedback brings different reactions between conditions.

Moreover, there was only one robot and one agent used in the experiment. There is strong evidence that even seemingly small differences in shape of a robot's head can affect participants' perception and attributions of a robot [Powers and Kiesler, 2006]. It is possible that if another robot was used in the experiment, the results would have been different. Further questions could be asked also about the use of different types of agents. A researcher must make a choice between 2D and 3D or animated and cartoon ECAs. It is currently not known how such differences would affect users. Both the robot and the ECA used in this experiment were static with light signaling that they are going to speak. However, the biggest advantage in general of robots and agents is that they are capable of moving themselves or moving parts of their bodies. Such a robot/ECA would be definitely more entertaining for users and could have influenced the results. In addition, the ECA used

was based on the robot's appearance, but lacked key functionality of computer agents – moving mouth when speaking. Therefore, while making it mimic the robot helped to ensure equal conditions, it made the comparison unfair for computer agents' technology.

Future research should also focus more specifically on few areas reported in this paper. There is some evidence in the previous [Bartneck, 2003] and the current study that robots induce stronger social facilitation effect than ECAs. However, systematic research that would compare their impact on users' performance on easy and difficult tasks is required. If these suggestions are confirmed, it could also add a new light on potentially higher anthropomorphization of robots than ECAs [Kiesler et al., 2008].

Finally, this research was unable to answer the question whether receiving feedback from a robot or an agent can improve user's task performance. It is necessary to include a control group in future research. Only then we will be able to know to what degree increase of a speed of solving problems is a result of learning new solutions and to what confidence boost after a robot's or an ECA's feedback.

References

- [Aiello and Douthitt, 2001] Aiello, J. and Douthitt, E. Social facilitation from Triplett to electronic performance monitoring. *Group Dynamics: Theory, Research, and Practice* **5**, 3 (2001), 163-180.
- [Barley, 1988] Barley, S. The social construction of a machine: Ritual, superstition, magical thinking and other pragmatic responses to running a CT scanner. In M.M. Lock and D.R. Gordon, *Biomedicine Examined*. Reidel, Hingham, MA, USA, 1988, 497-540.
- [Baron, 1986] Baron, R. Distraction-conflict theory: Progress and problems. In L. Berkowitz, *Advances in Experimental Social Psychology*. Academic Press, Orlando, FL, 1986, 1-40.
- [Bartneck, 2003] Bartneck, C. Interacting with an embodied emotional character. *Proceedings of the 2003 International Conference on Designing Pleasurable Products and Interfaces - DPPI '03*, ACM Press (2003), 55-60.
- [Beilock, et al., 2004] Beilock, S., Kulp, C., Holt, L., and Carr, T. More on the fragility of performance: Choking under pressure in mathematical problem solving. *Journal of Experimental Psychology: General* **133**, 4 (2004), 584-600.
- [Bickmore and Picard, 2005] Bickmore, T. and Picard, R. Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Computer-Human Interaction (TOCHI)* **12**, 2 (2005), 293-327.
- [Breazeal and Scassellati, 2000] Breazeal, C. and Scassellati, B. Infant-like social interactions between a robot and a human caregiver. *Adaptive Behavior* **8**, 1 (2000), 49-74.
- [Burgard et al., 1999] Burgard, W., Cremers, A.B., Fox, D., et al. Experiences with an interactive museum tour-guide robot. *Artificial Intelligence* **114**, 1-2 (1999), 3-55.
- [Catrambone et al., 2004] Catrambone, R., Stasko, J., and Xiao, J. ECA user interface paradigm: Experimental findings within a framework for research. In Z. Ruttkay and C. Pelachaud, *From brows to trust: Evaluating Embodied Conversational Agents*. Kluwer Academic Publishers, Dordrecht, the Netherlands, 2004, 239-267.

- [Dehn and van Mulken, 2000] Dehn, D. and van Mulken, S. The impact of animated interface agents: A review of empirical research. *International Journal of Human-Computer Studies* **52**, (2000), 1-22.
- [Dennett, 1987] Dennett, D.C. *The Intentional Stance*. MIT Press, Cambridge, MA, 1987.
- [Fogg and Nass, 1997] Fogg, B. and Nass, C. How users reciprocate to computers: An experiment that demonstrates behavior change. *Conference on Human Factors in Computing Systems*, ACM (1997), 331-332.
- [Friedman et al., 2003] Friedman, B., Kahn Jr, P., and Hagan, J. Hardware companions?: What online AIBO discussion forums reveal about the human-robotic relationship. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM (2003), 273-280.
- [Hall and Henningsen, 2008] Hall, B. and Henningsen, D. Social facilitation and human-computer interaction. *Computers in Human Behavior* **24**, 6 (2008), 2965-2971.
- [Huang et al., 2002] Huang, W., Olson, J.S., and Olson, G.M. Camera angle affects dominance in video-mediated communication. *CHI '02 extended abstracts on Human factors in computing systems - CHI '02*, ACM Press (2002), 716-717.
- [Judge and Cable, 2004] Judge, T.A. and Cable, D.M. The effect of physical height on workplace success and income: Preliminary test of a theoretical model. *The Journal of Applied Psychology* **89**, 3 (2004), 428-441.
- [Kawamichi et al., 2005] Kawamichi, H., Kikuchi, Y., and Ueno, S. Magnetoencephalographic measurement during two types of mental rotations of three-dimensional objects. *IEEE Transactions on Magnetics* **41**, 10 (2005), 4200-4202.
- [Khan, 1998] Khan, Z. Attitudes towards intelligent service robots. Royal Institute of Technology, IPLab, NADA, Report TRITA-NA-E98421 - IPLab-154, August 1998, 1-30. Also available as
http://scholar.google.fi/scholar?hl=en&q=Attitudes+towards+intelligent+service+robots&btnG=Search&as_sdt=2000&as_ylo=&as_vis=0#0.
- [Kiesler and Hinds, 2004] Kiesler, S. and Hinds, P. Introduction to this special issue on human-robot interaction. *Human-Computer Interaction* **19**, 1 (2004), 1-8.

- [Kiesler et al., 2008] Kiesler, S., Powers, A., Fussell, S.R., and Torrey, C. Anthropomorphic interactions with a robot and robot-like agent. *Social Cognition* **26**, 2 (2008), 169-181.
- [King and Ohya, 1996] King, W. and Ohya, J. The representation of agents: Anthropomorphism, agency, and intelligence. *Conference Companion on Human Factors in Computing Systems: Common Ground*, ACM (1996), 289-290.
- [Klein et al., 2002] Klein, J., Moon, Y., and Picard, R. This computer responds to user frustration: Theory, design, and results. *Interacting with Computers* **14**, 2 (2002), 119-140.
- [Koda and Maes, 1996] Koda, T. and Maes, P. Agents with faces: The effect of personification. *Proceedings 5th IEEE International Workshop on Robot and Human Communication. RO-MAN'96 TSUKUBA*, IEEE (1996), 189-194.
- [Laurel, 1997] Laurel, B. Interface agents: Metaphors with character. In B. Friedman, *Human Values and the Design of Computer Technology*. Center for the Study of Language and Information, Stanford, CA, USA, 1997, 207-219.
- [Lee et al., 2000] Lee, E., Nass, C., and Brave, S. Can computer-generated speech have gender?: An experimental test of gender stereotype. *CHI '00 Extended Abstracts on Human Factors in Computing Systems*, ACM (2000), 289-290.
- [Lee and Nass, 2003] Lee, K. and Nass, C. Designing social presence of social actors in human computer interaction. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM (2003), 289-296.
- [Lester et al., 1997] Lester, J., Converse, S., Kahler, S., Barlow, S., Stone, B., and Bhogal, R. The persona effect: Affective impact of animated pedagogical agents. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM (1997), 359-366.
- [Nass and Lee, 2000] Nass, C. and Lee, K. Does computer-generated speech manifest personality? An experimental test of similarity-attraction. *Conference on Human Factors in Computing Systems*, ACM (2000), 329-336.
- [Nass et al., 2000] Nass, C., Isbister, K., and Lee, E. Truth is beauty: Researching embodied conversational agents. In J. Cassell, J. Sullivan, S. Prevost and E.

- Churchill, *Embodied Conversational Agents*. MIT Press, Cambridge, MA, 2000, 374-402.
- [Nass et al., 1997] Nass, C., Moon, Y., Morkes, J., and Kim, E. Computers are social actors: A review of current research. In B. Friedman, *Human Values and the Design of Computer Technology*. Cambridge University Press, Stanford, CA, USA, 1997, 137-162.
- [Norman, 1994] Norman, D. How might people interact with agents. *Communications of the ACM* **37**, 7 (1994), 68-71.
- [Nowak and Biocca, 2003] Nowak, K.L. and Biocca, F. The Effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments. *Presence* **12**, 5 (2003), 481-494.
- [Park and Catrambone, 2007] Park, S. and Catrambone, R. Social facilitation effects of virtual humans. *Human Factors* **49**, (2007), 1054–1060.
- [Powers and Kiesler, 2006] Powers, A. and Kiesler, S. The advisor robot: Tracing people's mental model from a robot's physical attributes. *Proceeding of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction - HRI '06*, ACM Press (2006), 218-225.
- [Powers et al., 2007] Powers, A., Kiesler, S., Fussell, S., and Torrey, C. Comparing a computer agent with a humanoid robot. *Proceeding of the ACM/IEEE International Conference on Human-Robot Interaction - HRI '07*, ACM Press (2007), 145-152.
- [Rajaniemi, 2007] Rajaniemi, J. Java framework for WiFi-based Nabaztag device. University of Tampere, Dept. of Computer Science, Series of Publications D-2007-11, September 2007, 44-63. Also available as <http://www.cs.uta.fi/reports/dsarja/D-2007-11.pdf>.
- [Rickel and Johnson, 2000] Rickel, J. and Johnson, W. Task-oriented collaboration with embodied agents in virtual worlds. In J. Cassell, J. Sullivan, S. Prevost and E. Churchill, *Embodied Conversational Agents*. MIT Press, Cambridge, MA, 2000, 95-122.
- [Robins et al., 2005] Robins, B., Dautenhahn, K., Boekhorst, R., and Billard, A. Robotic assistants in therapy and education of children with autism: Can a small humanoid

- robot help encourage social interaction skills? *Universal Access in the Information Society* **4**, 2 (2005), 105-120.
- [Schmitt et al., 1986] Schmitt, B., Gilovich, T., Goore, N., and Joseph, L. Mere presence and social facilitation: One more time. *Journal of Experimental Social Psychology* **22**, 3 (1986), 242-248.
- [Scholl and Tremoulet, 2000] Scholl, B. and Tremoulet, P. Perceptual causality and animacy. *Trends in Cognitive Sciences* **4**, 8 (2000), 299-309.
- [Searle, 1980] Searle, J.R. Minds, brains, and programs. *The Behavioral and Brain Sciences* **3**, 3 (1980), 417-457.
- [Shechtman and Horowitz, 2003] Shechtman, N. and Horowitz, L. Media inequality in conversation: how people behave differently when interacting with computers and people. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM (2003), 281-288.
- [Shiomi et al., 2006] Shiomi, M., Kanda, T., Ishiguro, H., and Hagita, N. Interactive humanoid robots for a science museum. *Proceeding of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction - HRI '06*, ACM Press (2006), 305-312.
- [Shneiderman and Maes, 1997] Shneiderman, B. and Maes, P. Direct manipulation vs. interface agents. *Interactions* **4**, 6 (1997), 42-61.
- [Sproull et al., 1997] Sproull, L., Subramani, M., Kiesler, S., Walker, J., and Waters, K. When the interface is a face. In B. Friedman, *Human Values and the Design of Computer Technology*. Cambridge University Press, Stanford, CA, 1997, 162-190.
- [Takeuchi and Naito, 1995] Takeuchi, A. and Naito, T. Situated facial displays: Towards social interaction. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM Press/Addison-Wesley Publishing Co. (1995), 450-455.
- [Turkle, 1984] Turkle, S. *The second self: Computers and the human spirit*. Simon & Schuster, New York, NY, 1984.

- [van Mulken et al., 1998] van Mulken, S., André, E., and Muller, J. The persona effect: how substantial is it? *People and Computers: Proceedings of HCI'98*, Springer (1998), 53-66.
- [Walker et al., 1994] Walker, J., Sproull, L., and Subramani, R. Using a human face in an interface. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Celebrating Interdependence*, ACM Press (1994), 85-91.
- [Weiss et al., 2009] Weiss, A., Bernhaupt, R., Lankes, M., and Tscheligi, M. The USUS evaluation framework for human-robot interaction. *AISB2009: Proceedings of the Symposium on New Frontiers in Human-Robot Interaction*, (2009).
- [Wilson, 1997] Wilson, M. Metaphor to personality: the role of animation in intelligent interface agents. *Proceedings of the IJCAI-97 Workshop on Animated Interface Agents: Making them Intelligent*, Citeseer (1997), 1-10.
- [Winograd and Flores, 1987] Winograd, T. and Flores, F. *Understanding computers and cognition: A new foundation for design*. Addison-Wesley, Reading, MA, USA, 1987.
- [Wright et al., 1999] Wright, P., Milroy, R., and Lickorish, A. Static and animated graphics in learning from interactive texts. *European Journal of Psychology of Education* **14**, 2 (1999), 203-224.
- [Yamato et al., 2001] Yamato, J., Shinozawa, K., Naya, F., and Kogure, K. Evaluation of communication with robot and agent: Are robots better social actors than agents. *Proc. of the 8th IFIP TC 13*, (2001), 690-691.
- [Yerkes and Dodson, 1908] Yerkes, R. and Dodson, J. The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology* **18**, 5 (1908), 459-482.
- [Zajonc, 1965] Zajonc, R. Social Facilitation. *Science* **149**, (1965), 269-274.
- [Zlotowski, 2010] Zlotowski, J. Social human-robot interaction: Review of existing literature. University of Tampere, Dept. of Computer Sciences, Series of Publications D-2010-1, January 2010, 266-284. Also available as <http://www.cs.uta.fi/reports/dsarja/D-2010-1.pdf>.

[Zuboff, 1988] Zuboff, S. *In the age of the smart machine: The future of work and power*.
Basic Books, New York, NY, 1988.

Appendix A

Possible Nabaztag feedback for subjects' task performance

Negative Nabaztag statements:

- This is a wrong answer. Next time it will go better.
- Unfortunately this is an incorrect answer. Try next equation.
- Incorrect answer.

Positive Nabaztag statements:

- Well done! Keep on good work.
- Correct answer.
- Good! Please continue the task!

Appendix B

Post-test questionnaire statements (on a 5-point Likert scale)

1. Nabaztag encouraged me to focus on the task.
2. It was entertaining to do the task with Nabaztag.
3. I felt comfortable with Nabaztag performing the task with me.
4. Nabaztag's feedback made me more confident doing the task.
5. The task was easy.
6. I liked Nabaztag.
7. I felt Nabaztag's presence.
8. Nabaztag's feedback was irritating.
9. The task was boring.
10. It was more fun to do the task with Nabaztag.

Appendix C

Instructions given to subjects before the experiment

You are asked to perform series of mathematical tasks on a computer. Please do them as fast and as accurately as you can. During the process, Nabaztag (robot rabbit), placed above the computer screen, will regularly give you some feedback on your performance.

You will be shown series of modular arithmetic statements. Your goal is to judge whether problem statements are true or false. You will do it by pressing “True” or “False” button displayed under each statement. The task will last for 10 minutes after which you will be asked to fill the questionnaire. However, you can stop the mathematical task at any time by pressing “End the task” button, which will lead you straight to the questionnaire.

Example of modular arithmetic statement

$$50 \equiv 38 \pmod{4}$$

To solve the problem you need to subtract a middle number (i.e. 38 - between 50 and 4) from the first number (i.e. 50 – 38) and then the result of this (i.e. 12) is divided by the last number (i.e. 12/4). If the dividend is a whole number (as here 3) then the statement is true. On the contrary, if the dividend is a decimal number then the statement is false.