ALL EYES ON THE AGENT'S APPEARANCE?! Investigation of Target-group-related Social Effects of a Virtual Agent's Appearance in Longitudinal Human-Agent Interactions

Von der Fakultät für Ingenieurwissenschaften,
Abteilung Informatik und Angewandte Kognitionswissenschaft
der Universität Duisburg-Essen

zur Erlangung des akademischen Grades

Doktor der Philosophie (Dr. phil.)

genehmigte Dissertation

von

Carolin Straßmann
aus
Recklinghausen

Erstgutachterin: Prof. Dr. Nicole C. Krämer
Zweitgutachter: Prof. Dr. Stefan Kopp
Tag der mündlichen Prüfung: 17.12.2019

DOI: 10.17185/duepublico/70205
URN: urn:nbn:de:hbz:464-20190624-150844-3

ACKNOWLEDGEMENTS

Before the chapter Ph.D. will be closed, it is time to say thank you to all, who contributed in manifold ways to this dissertation.

First, I would like to thank my supervisor Prof. Dr. Nicole Krämer for encouraging me during all the years, giving me great opportunities and sparking my joy for the academic world. Thanks for trusting in me, for the space you gave me to unfold my abilities and the support and advice whenever I needed it. I would also like to thank Prof. Dr. Stefan Kopp for volunteering to be second supervisor of this thesis and valuable inputs especially for Study 2.

A lot of thanks belong to the whole SP family! I found in you probably the best colleagues and friends I could ever have (thank you Nicole for casting this great people together). Stephan Winter, German Neubaum, Laura Hoffmann, Leonie Rösner, Sabrina Sobieraj, Sabrina Eimler and Astrid Rosenthal-von der Pütten, thank you for being a "big sibling", on who I know I could always count on. Thank you, Elias Kyewski for spending the Ph.D. time together since our early student assistant days. It was always a great help to have someone to share experiences (and drinks) with. Also thanks to the Robot-Girls: Jessica Szczuka and Aike Horstmann for the Terrific Tuesdays, best Amsterdam-trip and the emotional and topic-wise exchange especially in the last year! Within our journey together, the joint conferences, drinks and enlightening discussions I was grateful to have all of you by my side. I hope for a lot of more fun moments to come!

Thanks to all KOMPASS-team members for valuable discussions and a lot of help with the implementation of the studies. Further I would like to acknowledge all students, who contributed to this dissertation, especially Tabea Holtbernd, Nurzada Reina Henkel and Sueyda Yüzer. Moreover, I would like to thank my current co-workers at the HRW for supporting me and giving me free up space to finish this thesis.

Thank you once again to all, who proofread and commented different parts of my dissertation. I owe you one!

Last but definitely not least, I would like to thank my family and friends for always believing in me, giving emotional support and being patient during the entire Ph.D. time. The biggest thank belongs to my parents Sabine and Thomas Straßmann, who encouraged me my whole life and without whom this thesis would not have been possible. Thank you, Katharina Schwarz not only for you excel skills, but the major encouragement in multiple ways.

Parts of this dissertation have been presented in or accepted for talks and posters at conferences, PhD workshops or have been published in peer-reviewed journals. Please refer to the following references for more information:

Overview

Straßmann, C. (2017). The influence of a virtual agent's appearance on social relations in a long-term interaction. *Research proposal presented at the PhD Workshop of the German Media Psychology Division at the 10th conference of the media psychology division of the German Psychological Society DGPs (September, 2017)*. Landau, Germany

Categorization & Study 1

Straßmann, C., (2016, June). *The influence of a virtual agent's appearance on social relations in a long-term interaction*. Poster presented at the 2nd international summer school "Living with Media" at the Nanyang Technology University, Singapore.

Straßmann, C. & Krämer, N.C. (2017). Categorizing a virtual agent's visual appearance and discovering age-related user preferences. *Presentation at the 10th conference of the media psychology division of the German Psychological Society DGPs (September 2017)*. Landau, Germany

Straßmann, C., & Krämer, N. C. (2017, August). A categorization of virtual agent appearances and a qualitative study on agerelated user preferences. In International Conference on Intelligent Virtual Agents (pp. 413-422). Springer International Publishing.

Study 2

Straßmann, C., and Krämer, N.C. (2018). A two-study-approach to explore the effect of user characteristics on users' perception and evaluation of a virtual assistant's appearance. *Multimodal Technologies and Interaction. 2 (66).* doi:10.3390/mti2040066

Study 3

Straßmann, C. & Krämer, N.C. (2018). Seniors like. It, but students don't. Age-related differences in the evaluation of appearance and embodiment in a human-agent interaction. *Presentation at the 51th conference of the German Psychological Society DGPs (September 2018)*. Frankfurt, Germany

Study 4

Straßmann, C. & Krämer, N.C. (2018). Effects of attractiveness and humanlikeness of an intelligent virtual agent in a long-term interaction on social relations between humans and agent. *Poster to be presented at the Intelligent Automation Symposium (November 2018)*. Münster, Germany

ABSTRACT

The present dissertation examines the effect of a virtual agent's appearance within longitudinal interactions between the agent and the user. Prior studies have already demonstrated that feelings like bonding toward the agent occur within a long-term usage. As from human-human interaction it can be derived that the appearance of interlocutors is important for the development interpersonal relationships; this dissertation investigated the role of appearance in terms of social relations. Moreover, the impact of the aimed target-group was focused, since applications in terms of health or assistance are promising application fields for specific target groups such as seniors. Therefore, an empirical approach consisting out of four different studies was used. Results from qualitative interviews (Study 1) indicate valuable insights about target-group-related differences. Students indicate that they perceive the agent more as a technical entity and want prevent from social processes that an agent might evoke. In contrast, seniors seem to value the familiarity of a humanoid appearance, since they like to address an interlocutor and assume a humanoid appearance to trigger trust and bonding. This was partly supported by Study 2, which demonstrated that seniors rely more on species and students on realism. In an actual interaction between human and agent (Study 3), the results for seniors have been substantiated. Seniors benefit from social outcomes of a humanoid appearance, as a humanoid agent was also found to trigger higher usage intention, trust and bonding for seniors, while this is not the case for students. Hence, seniors seem to be more sensitive for appearance cues than students, who might relay more on functions. However, also for students appearance was found to have promising effects. In Study 4 an attractive agent was more liked and enhanced users' intention to live healthy. Although no development of appearance effects over times have been found, still the evaluation of the agent and users' closeness towards the agent seem to change over time. This might be useful to foster the usage of virtual agents and their effectiveness. Altogether, this work makes valuable contribution to research regarding appearance effects: (1) appearance variables have been categorized and investigated in a more systematic manner, (2) the impact of user characteristics has been examined, (3) appearance effects have been investigated over time and with regard to an interpersonal relationship between the user and the agent. To conclude, the agent's appearance can especially enhance a longitudinal human-agent interaction for seniors, but also for students, whose evaluation is less affected by the agent's appearance, beneficial outcomes have been found (e.g., enhancement of health intention).

ZUSAMMENFASSUNG

Die vorliegende Dissertation untersucht die Auswirkung des Erscheinungsbildes eines virtuellen Agenten in längerfristigen Interaktionen zwischen dem Agenten und den Nutzenden. Vorherige Studien haben bereits gezeigt, dass soziale Reaktionen wie eine emotionale Bindung an den Agenten bei langfristiger Nutzung auftreten. Aus der Mensch-Mensch-Interaktion lässt sich weiter ableiten, dass das Auftreten von Gesprächspartnern für die Entwicklung zwischenmenschlicher Beziehungen wichtig ist. Daher untersucht diese Dissertation die Rolle des Erscheinungsbildes in Bezug auf soziale Beziehungen in langfristigen Interaktionen mit virtuellen Agenten. Darüber hinaus wurde die Wirkung der Zielgruppe fokussiert, da die Nutzung von virtuellen Agenten in Bezug auf Gesundheit oder Assistenz vielversprechende Anwendungsfelder für spezielle Zielgruppen wie beispielweise Senior*innen sind. Daher wurde ein empirischer Ansatz bestehend aus vier verschiedenen Studien verwendet. Ergebnisse aus qualitativen Interviews (Studie 1) weisen auf wertvolle Erkenntnisse über zielgruppenbedingte Unterschiede hin. Studierende geben an, dass sie den Agenten eher als technische Entität wahrnehmen und möchten, dass soziale Prozesse, die ein Agent hervorrufen kann, verhindert werden. Im Gegensatz dazu scheinen Senior*innen die Vertrautheit einer humanoiden Erscheinung zu schätzen, da sie während des Gespräches gerne eine Person adressieren wollen und erwarten, dass ein humanoides Erscheinungsbild Vertrauen und emotionale Bindung auszulösen. Diese Erkenntnisse wurden größtenteils von Studie 2 unterstützt, in der gezeigt wurde, dass Senioren stärker auf die Spezie des Agenten schauen und Studierende auf den Realismusgrad. In einer echten Interaktion zwischen Mensch und Agent (Studie 3) wurden die Ergebnisse für Senioren weiter gestützt. Senioren profitieren von den sozialen Einflüssen eines humanoiden Aussehens, da ein humanoider Agent eine höhere Nutzungsabsicht, Vertrauen und Bindung für Senioren auslöst, während dies für Studierende nicht der Fall ist. Infolgedessen scheinen Senioren sensibler für Erscheinungsbildvariablen zu sein als Studierende, welche sich eher auf Funktionen beziehen. Es zeigte sich jedoch auch, dass für Studierende das Erscheinungsbild eines virtuellen Agenten ebenfalls vielversprechende Auswirkungen haben kann. In Studie 4 wurde gezeigt, dass ein attraktiver Agent mehr gemocht wurde und die Intention gesund zu leben der Nutzenden gesteigert hat. Obwohl sich die Effekte des Erscheinungsbildes im Lauf der Zeit nicht verändert haben, scheint sich die allgemeine Bewertung des Agenten und die wahrgenommene Nähe zum Agenten im Laufe der Zeit zu ändern. Dies kann hilfreich sein, um die Nutzung von virtuellen Agenten und deren Effektivität zu fördern. Insgesamt leistet

diese Arbeit einen wertvollen Beitrag zur Erforschung von Erscheinungsbildeinflüssen: (1) Erscheinungsvariablen wurden systematisch kategorisiert und untersucht, (2) der Einfluss von Nutzendeneigenschaften wurde untersucht, (3) Effekte des Erscheinungsbildes wurden über mehrere Interaktionen und eine längere Zeit sowie in Hinblick auf eine zwischenmenschliche Beziehung getestet. Zusammenfassend kann gesagt werden, dass das Erscheinungsbild von virtuellen Agenten insbesondere die Mensch-Agenten-Interaktion für Senior*innen positiv beeinflussen kann. Aber auch für Studierende, deren Beurteilung durch das Aussehen des Agenten weniger beeinflusst wird, wurden positive Ergebnisse gefunden (z. B. Verbesserung der Gesundheitsabsicht).

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

I INTRODUCTION

The presented dissertation covers the effects of a virtual agent's visual appearance in long-term interactions. Virtual agents are autonomous, visual representations of computer interfaces that are able to interact humanlike (Cassell, 2000). Due to the very natural communication interface, which is mostly based on speech, the fields of applications for virtual agents are manifold. While some applications are based on one-time events (e.g. at a sales-counter or information desk), most of the possible future applications are not based on single but multiple social interactions (e.g. as a learning companion or daily-life assistant). Although there is ample evidence for social reactions towards computers or virtual entities (Reeves & Nass, 1996) in short-term interactions, long-term effects remain understudied.

During the past years, Bickmore and colleagues conducted long-term studies to explore the application of virtual agents as health advisors. The results from those studies indicate that interlocutors of a relational agent had the feeling to establish some bond between themselves and the agent, which fostered their motivation to be physically active (Bickmore, Caruso, Clough-Gorr, & Heeren, 2005). Moreover emotionally communication pattern were found to enhance this feelings of bonding towards a virtual agent (Bickmore, Gruber, & Picard, 2005). Such social effects and interpersonal relationships within a human-agent interaction might be helpful to foster the users' liking of the agent and their usage intention, which both are expected to result in a higher efficacy of the agent. Those processes can especially help with regard to the aimed target group. While long-term interactions are most likely to occur in health-related and assistive applications, which focus on people in need of support as a promising target group (Kopp et al., 2018; Yaghoubzadeh, Kramer, Pitsch, & Kopp, 2013), those people were also found to be rather skeptical towards such artificial entities (Scopelliti, Giuliani, & Fornara, 2005). Therefore, social processes might foster the initial attraction, as well as the continuous usage of virtual agents.

Overall, these studies give first hints that users feel some bond between themselves and the agent when they are interacting over a longer period. However, they mostly focused on the user's satisfaction with and the efficacy of the agent and did not concentrate on the social relation between the user and the agent. Hence, it is still unclear which cues will drive those effects.

In the development of social relationships between humans (like friendships or romantic relationships), qualities like physical attractiveness need to be taken into account

(Fehr, 2000). As mentioned before, humans transfer social reactions they would show toward other humans also to computers and virtual entities (Reeves & Nass, 1996). Consulting this assumption, physical attractiveness of a virtual agent might also impact the development of a social bonding. Thus, a virtual agents appearance might influence the user's social bonding toward the agent in longitudinal interactions.

Prior research already determined the importance of an agent's appearance in a human-agent interaction. There are multiple studies focusing on this aspect and different effects on motivation (Baylor, 2009), buying intention (Terada, Jing, & Seiji Yamada, 2015), learning outcomes (Sträfling, Fleischer, Polzer, Leutner, & Krämer, 2010), persuasive effects (Hanus & Fox, 2015) and the agent's overall evaluation (Ring, Utami, & Bickmore, 2014) were found. In sum, the virtual agent's appearance affects the human-agent interaction in many ways. However, most of this effects are not focused on social relations and these results are derived from one-time interactions and single evaluations. Hence, to understand the role of appearance with regard to social relations over a longer period, long-term interactions between users and agents are needed. Since effects of appearance in one-time interactions are supported, it is questionable how these effects develop within a longer period of human-agent interactions. Similar to human-human interaction, the first impression of a virtual agent is of great importance with regard to social relations. However, Bergmann et al. (2012) found that the first evaluation of a virtual agent can be revised afterwards. Thus, it is worth to investigate the development of the mere visual appearance evaluation over time and in which way the visual appearance drives social relations in long-term interactions.

In contrast to the look of humans, the visual appearance of a virtual agent is fictional and can be manipulated in endless ways. Therefore, it is hard to predict what kind of appearance variables will be preferred and which variables will be the most effective ones in the context of social relations. Most of the prior studies compared different agents whose appearance was not varied systematically, since different types of agents with various appearance features were used (e.g. different species with different degrees of realism have been compared, so that no implications are derivable which factor leads to the resulting outcomes). To examine the effects of appearance in long-term interactions a systematic categorization of various appearance variables is needed. Thus, the current work also aims to construct a categorization of appearance variables and further to study the effects of these variables more widely.

The overall goal of the current dissertation is to investigate the influence of appearance in long-term interactions. One typical application of a human-agent long-term

interaction might be the application as a virtual assistant (e.g. for health advices or daily-life support). Therefore, all constructed studies are framed in the context of a virtual assistant that might give support in daily-life and health. As this applications focus on special target groups such as people in need of support, who might have specific preferences and needs, in the current approach also target group related differences are investigated.

On the basis of the aforementioned research gap, this work aims to (1) systematically categorize and investigate effects of different appearance variables, (2) investigate target group related differences with regard to the evaluation of the appearance variables, (3) build an understanding of the development of appearance effects within a long-term interaction and (4) examine the effect of an agent's appearance on social relations in a long-term interaction.

To tackle this research goals, in the theoretical framework (Chapter I) the possible applications of virtual agents and the derivations made from it are discussed. Moreover, psychological effects in a human-agent interaction with special focus on long-term interactions are presented. Thereafter, based on prior research, a categorization of appearance variables is constructed and empirical evidences are discussed along this categorization. Accordingly, the derivations for the current work are summarized. In Chapter III the empirical part of this work is presented. Within the empirical framework of this work, four studies have been planned and conducted consecutively based on the constructed categorization. First, to explore this categorization and to extract the categories of most interest (especially with regard to different target groups), a qualitative interview study was conducted (Study 1). Building on the results of the interviews, the effect of species and realism was examined in a systematical manner, while also target group related differences were explored (Study 2). While the first two studies investigated the evaluation of a virtual agent by the mere perception of the appearance without any interaction, the latter two studies investigated the effect of different appearance variables in a human-agent interaction. Study 3 investigated age-related differences with regard to the effects of an agent's embodiment, species and realism with regard to the agent's evaluation and social outcomes such as trust and bonding. Finally, the effects of an agent's appearance (here with regard to attractiveness and humanlikeness) were examined in a long-term interaction over 4 weeks (Study 4). The outcomes from the empirical work are outlined in a global discussion, which also presents the theoretical as well as practical implications from this work (Chapter IV).

II THEORETICAL FRAMEWORK

This section presents an overview of the relevant concepts and aims to build a framework based on theoretical assumptions and empirical evidence. Firstly, we describe what a virtual agent is and wherefore it is useful and applicable to explain why it is important to look at the social effects in long-term interactions. Then, the theory of social effects in human–agent interactions is discussed. As this work broaches the issue of appearance effects, appearance variables are categorized and prior findings are sorted along this categorization. This gives an important overview of the recent state of research, on which the later empirical part can build on. Lastly, the theoretical assumptions are summarized and the general research model is presented.

1 VIRTUAL AGENTS AND THEIR APPLICATION

1.1 DEFINITION

Virtual agents are not a novel concept and have been developed for over 20 years; however, owing to digitalization and steady technological advancements, virtual agents are becoming more and more ubiquitous. A virtual agent is described as an autonomous computer interface associated with artificial intelligence. Wooldridge and Jennings describe the notion of artificial intelligence as: “A subfield of computer science which aims to construct agents that exhibit aspects of intelligent behavior” (Wooldridge & Jennings, 1995 p. 116). The authors further define an agent as a software-based computer system that is characterized by its autonomy, social ability, reactivity, and proactiveness (Wooldridge & Jennings, 1995). Justine Cassell (2000) goes a step further with her definition of conversational agents and highlights the specifically humanlike characteristics of an agent:

> Embodied conversational interfaces are not just computer interfaces represented by way of human or animal bodies. And they are not just interfaces where those human or animal bodies are lifelike or believable in their actions and their reactions to human users. Embodied conversational interfaces are specifically conversational in their behaviors, and specifically humanlike in the way they use their bodies in conversation. That is, embodied conversational agents may be defined as those that have the same properties as humans in face-to-face conversation. (p. 70)

Besides the term *virtual agent*, several other terms are frequently and often synonymously used, which is also seen in the aforementioned notions where the terms *intelligent agents* (Wooldridge & Jennings, 1995) and *embodied conversational agents* (Cassell, 2000) are used. Krämer (2008) presents several synonyms of virtual characters that are often used such as *avatars*, *anthropomorphic interfaces*, *autonomous agents*, *interface agents*, *embodied conversational agents*, *virtual assistants,* and *virtual friends*. In line with this, Soliman and Guetl (2010) list eight different terms (that mostly match the terms presented by Krämer (2008)) such as *agent*, *virtual character,* or *avatar* as example synonymous terms. Table 1 presents the different terms and their conceptual meanings adopted by Soliman and Guetl (2010).

Table 1.
Overview over used terms referring to virtual agents adopted from Soliman and Guelt (2010)

Term	**Meaning and Characteristics**
Agent	Autonomous entity with goals
Virtual character	Has character An animated form is named animated virtual character *HCI term*
Embodied agent	Has a physical body Stresses the visual appearance Embodied conversational agents have conversation abilities *Artificial intelligence term*
Pedagogical agent	Stresses pedagogical functions
Intelligent agent	Stresses intelligence abilities such as learning
Guidebots	Stresses guidance functions to simulate and encourage learning
Avatar	Incarnation of the user in the virtual environment Is a selected character by the user Mainly used in 3D virtual environments to emphasize personal preferences Guided by user
Intelligent pedagogical agent	Combines different abilities including intelligence and pedagogical orientation Autonomous (not directly guided by user)

The authors further ague that these different concepts have been used for the same goals. While this can be assumed for most of the listed concepts, avatars and agents have to

be distinguished with caution. Although Soliman and Guetl (2010) mention that an avatar is guided by the user and is an incarnation of the user in a virtual environment, they still see all terms and concepts to share the same goals. By its definition, an agent is an autonomously acting entity through artificial intelligence (Balakrishnan & Honavar, 2001) with no intervention of a human needed. By contrast, an avatar is a virtual representation of a human, which is controlled completely by the human (von der Pütten, Krämer, Gratch, & Kang, 2010). There is a body of research on the investigation of this concept called *agency* (for an overview, see Fox et al., 2015). Empirical work that varied agency (compared the influence of agents with that of avatars) and tested the threshold model of social influence (Blascovich et al., 2002), which claims that avatars always have greater social impact than agents, did not find differences between both agency levels. In line with this, a meta-analysis revealed that it depends on the measurement method (objective vs. subjective), the tasks, and the overall study design whether differences between agents and avatars are found (Fox et al., 2015). Thus, under specific circumstances both concepts seem to trigger similar social outcomes by the user.

Nevertheless, in this work only findings and implications for virtual agents are presented. Since this dissertation aims to investigate the influence of appearance on social outcomes in a human–agent interaction, it is important to distinguish between the two concepts. As described, avatars represent the human user in a certain way. It can therefore be assumed that users have different preferences and the design of an avatar looks different than the one of an agent. An avatar aims to represent the user and therefore other concepts like impression management (Leary & Kowalski, 1990) or identification (Klimmt, Hefner, Vorderer, Roth, & Blake, 2010; D. D. Li, Liau, & Khoo, 2013) need to be taken into account when it comes to the design of an avatar. Users might want to be represented by a big ugly monster during a game in order to impress and quench opponents (impression management), while they probably do not want to use a virtual assistant represented by the same virtual character. Unfortunately, to date, no study exists that separates these effects with regard to appearance. However, to avoid confounding effects, in this work only empirical evidence and derivations of virtual agents are presented and applied.

Referring to the aforementioned definitions and synonyms, the current work assumes a virtual agent to be a based on an artificial intelligence, autonomously acting computer interface that acts and communicates in a humanlike way. Mainly, embodied virtual agents (compare the notion of embodied conversational agents (Cassell, 2000)) will be discussed in this dissertation, since appearance factors only become critical when a physical body is

present (for a more detailed explanation of embodiment, see Chapter 3.1.1). However, to some degree agents without an embodiment are also taken into account. Therefore, the overarching term *virtual agents* is used. When virtual agents are applied in the educational context, it is often referred to as *pedagogical agent* (c.f. Krämer, 2008; Soliman & Guetl, 2010). In this work, the application of assistance and support is used, and therefore the term *virtual assistance* or *virtual assistant* might occur. Potential application fields will be discussed in the next Chapter.

1.2 APPLICATION FIELDS AND TARGET GROUPS

Owing to their natural interface and their ability to communicate in a humanlike way (Cassell, 2000), virtual agents are ubiquitously applicable to different tasks. Often-discussed application fields are: education, e-commerce, entertainment, recommendations, assistive living, and medical applications. In the following, these application fields are described and related target groups are mentioned to build a holistic understanding of the tasks and users of virtual agents.

1.2.1 Education

On the basis of prior research, Gulz (2005) summarized multiple benefits of virtual agents applied in an education context: Learners relate to the agent and this leads to collaboration and can further function as a role model (Ryokai, Vaucelle, & Cassell, 2003), an agent reduces anxieties during learning (Baylor, Shen, & Warren, 2004), frustration is reduced (Baylor, Warren, Park, Shen, & Perez, 2005), and agents can foster self-efficacy during the learning process (Baylor & Kim, 2005). A pedagogical agent can help especially in a learning system where motivational support is better when offered separately from information, or as a co-learner to enhance companionship (Baylor, 2009a). Embedded in teaching classes, virtual agents (or in this context specifically pedagogical virtual agents) can enhance motivation (Baylor, 2009a; Gulz, 2005) and learning outcomes (cf. Johnson & Lester, 2015).

Pedagogical agents have already been embedded in different educational systems, where they act as teacher, instructor, or coach (for an overview, see Agneta Gulz & Haake, 2006a). Thus, the agent can also adopt different instructional roles within the learning process such as expert, instructor, mentor, or learning companion (Baylor & Kim, 2004). While a motivator was seen as someone who supports the learners in performing a task and should foster their self-esteem, a mentor was assumed to guide the learners in the learning process

more than simply presenting information (Baylor & Kim, 2004). These roles were already found to affect psychological outcomes of the learners in the educational process (Baylor & Kim, 2003). Baylor and Kim (2004) embedded the pedagogical virtual agent in an e-learning tool in a computer literacy class, where it should help learners to develop instructional plans for e-learning. Similarly, Shiban et al. (2015) embedded a pedagogical agent in an e-learning lesson during a statistics course in a university. Thus, pedagogical agents are most likely applied in e-learning sessions. While the application during online classes can therefore be seen as the most common one, Trinh, Ring, and Bickmore (2015) presented a system where a virtual agent is used as a co-presenter. Using this technique, pedagogical agents could assist not only during online classes, but also in a real-world learning session.

The target group of pedagogical agents might be broad, as all kinds of people can learn with and from it. Nevertheless, in the aforementioned work mostly students and school children were studied. Since children and young adults have a higher chance of being in an educational context (e.g., school or university), they can be seen as the main target group, although the target group is not restricted to them.

1.2.2 E-Commerce and Recommendations

Virtual agents can be applied as company representatives in online stores, who are implemented via audiovisual media or text-to-speech techniques (Yang, Ou, & Zhou, 2017). One popular example for this application of a virtual agent is IKEA's Anna, who was embedded on the IKEA webpage for about 10 years (Wakefield, 2016). These agent are assumed to have a big impact on e-commerce processes: "Recommendation Agents (RAs) as an obvious application for predictive IT have now been widely considered as one of the major technological emphasis on e-commerce development" (Yang et al., 2017 p. 1). As presented with regard to an educational application, also in e-commerce applications can a virtual agent function in different roles. McGoldrick and colleagues (2008) extracted three main roles for agents in the application of e-commerce that were evaluated as applicable by customers:

(1) Helper that solves problems and saves labor

(2) Friendly, sociable, and welcoming host

(3) Personal shopper and recommending agent

Building on their empirical research, the authors further stated that customers want to keep the social interaction they are used from offline consumer practices and therefore virtual

agents can be embedded (McGoldrick et al., 2008). Moreover, Terada and colleagues (2015) demonstrated that an agent that gives product recommendations can, depending on its appearance, persuade people to buy a certain product. In line with this, Lunardo and colleagues (2016) summarized that agents in online shops can foster trust, positive attitude, positive emotions, and purchase intentions. To conclude, virtual agents can enhance commercial outcomes thanks to their humanlike interface that triggers inter alia social interactions and trust.

The target group of commercially applied agents is broad, as all customers of online shops and commercial websites are potential users of these agents.

1.2.3 Entertainment

Virtual agents can also be used for entertainment reasons. While in games the players are mostly represented via avatars (compare definition in Chapter 1.1), virtual agents are also embedded, for example, as opponents or extras (Emmerich & Masuch, 2016). Koda and Maes (1996) emphasized that personified interfaces fit well for entertaining applications and can enhance people's engagement. In line with this, social entities can decrease the perceived loneliness of players and can therefore enhance experiences during the game, when the agents are carefully designed (Liszio, Emmerich, & Masuch, 2017). Due to these advantages, virtual agents are broadly included in current game frameworks (De Kort & Ijsselsteijn, 2008; Söbke & Streicher, 2016). When an architecture framework for serious games is developed, it is required to "provide serious games with believable, emotional agents to help players learn skills and evaluate their performance" (Söbke & Streicher, 2016 p. 160). The authors further claim that these agents are linked to the dialogues and narratives of the game (Söbke & Streicher, 2016). Thus, virtual agents are beneficial in entertainment, mainly in games, to foster engagement and player experience.

As entertainment applications are mostly computer games, the target group here is properly characterized by higher computer literacy and greater experiences with technologies in general. Nevertheless, there are developments of games for specific target groups. Serious games in particular are often developed with a specified user group in mind such as children (Fuchslocher, Niesenhaus, & Krämer, 2011; Gerling, Fuchslocher, Schmidt, Krämer, & Masuch, 2011) or seniors (Nguyen et al., 2017; Phan Tran, Robert, & Bremond, 2016).

1.2.4 Medical and Health Application

Like in the aforementioned area of serious games, virtual agents can also function in the realm of health. They can be used as health advisor or medical assistant. The overarching goal of this application task it to beneficially enhance health-related outcomes of humans. Since there is a lack of health caregivers (SpiegelOnline, 2018), these applications are becoming more and more beneficial.

A health advisor can, for example, be used in terms of self-management support, where they help people manage the handling of their (e.g., chronic) illnesses (van Wissen, Vinkers, & van Halteren, 2016). In this application, the agent is expected to advise and guide the user with health promotion and to initiate and maintain changes in the user's self-management routines (van Wissen et al., 2016). Virtual health advisors might in some circumstances be even more effective than human health advisors, since they are accessible for the user at all times and not restricted to appointments only (van Wissen et al., 2016).

In line with this, Bickmore and colleagues (2010) also see big advantages in regard to self-care and health behavior change. They especially see benefits for people who have low health literacy, as the usage of and communication with a virtual agent is natural. This could also break up health disparities, since independent of their literacy, people can use the health-care agent and improve their health behavior (Bickmore, Pfeifer, et al., 2010).

Even more than giving advice on health-related self-management behavior, virtual agents are assumed to help in making decisions about specific health tracks such as treatments. Robertson et al. (2015) presents three advantages to applying virtual agents in health decisions:

> 1) Patients can learn essential health information without requiring time from their provider. 2) The interactive, conversational modes of communication used by ECAs can overcome passivity limitations of traditional health decision aids and can promote active learning and decision-making. 3) ECAs can allow patients to take adequate time to understand important information, repeating content or explaining content in simpler terms as necessary, all critical features for patients with low health literacy. (p. 430)

Thus, in sum, virtual agents are not restricted to time slots and have no limit in duration of use, which in combination with the natural communication style might lead to a more intense analysis of the information and in the end hopefully to a better understanding and more profound decision-making process. These assumptions are in line with the findings of Bickmore, Pfeifer, et al. (2010), who found virtual health-care systems to be easy to use and accepted by people regardless of their literacy. The authors build on these advantages and

aim to develop an agent that can help during the decision-making process of prostate cancer patients (Robertson et al., 2015).

Virtual agents can also be used in a therapeutic sense. Bickmore, Mitchell, et al. (2010) investigated a virtual human as a therapeutic tool for depressive patients and found it to be useful, as people with a high depression score reported a high therapeutic alliance after the intervention.

Additionally, virtual agents might also be used as a diagnostic tool, either as the first step before people visit a human medical professional (Shaked, 2017) or within a clinical context (Rizzo et al., 2014). Vault and colleagues (2014) developed in the SimSensei Kiosks project a tool with which distress indicators can be assessed, such as nonverbal behavior that is correlated with depression. Thus, SimSensei Kiosk is a virtual agent that interviews patients and aims to be engaging and make the patients feel comfortable to disclose any psychological distress problems that are highly sensitive (De Vault et al., 2014).

Moreover, virtual agents can be used for training purposes in the health context, where difficult and dangerous procedures cannot be reproduced in the real world (Volante et al., 2016). With regard to this application, Volante et al. (2016) list several tasks where virtual agents are used in health training applications such as routine patient surveillance, interview emulation, or uncomfortable situations. It is further argued that for a high realism of the training process and similar emotional responses as shown toward real humans, the applied agents are required to imitate human behavior as accurately as possible (Volante et al., 2016). This application overlaps with the educational contexts explained in Chapter 1.2.1. It can be seen as a specific case of educational use. However, the specific target group of a medical training process also has specific requirements, since human patients are imitated and partly emotional responses are needed during the training process.

To conclude, virtual agents in the health-care realm can be used as health advisors, therapists, diagnostic tools, decision-making assistants, and training tools for health professionals. Based on these functions, the target group of these applications is mostly participants with certain health problems such as people with chronic illnesses, risk groups (e.g., seniors), or people who have to make health decisions. As findings pointed out (Bickmore, Pfeifer, et al., 2010), people with all degrees of literacy can use these systems, which might be especially beneficial to break up boundaries. Besides patients and people with health problems, also health professionals are potential users of such technologies, especially of the training tool but also of, for example, the diagnostic tool.

1.2.5 Assistive Living

Several authors described virtual agents as a promising application in terms of assistive elderly care (Kopp et al., 2018; Tsiourti et al., 2016; Yaghoubzadeh et al., 2013). Like in the health sector, there is a lack of care persons and at the same time we have an aging population, so that assistive technologies are of great interest (Tsiourti et al., 2016).

Virtual agents applied in assistive technologies should respond to the users' needs and help users with their daily-life routines and their care self-management (Tsiourti et al., 2016). There are several groups of people with need of support, for whom cognitive limitations raise problems for mastering daily-life activities. Within this type of application, health-related topics (as discussed in Chapter 1.2.4) might also occur, but in a more preventive sense than as an actual treatment method. However, the context of assistive living is even broader than only health-related issues. These people might have problems with regular activities such as having meals, taking medications, or meeting social contacts (Yaghoubzadeh et al., 2013). More precisely, assistive agents can help the elderly to locate objects, can give reminders, and guide the user with household activities (Tsiourti et al., 2016). Additionally, they can serve in emergency detection (Tsiourti et al., 2016). Spiekman et al. (2011) summarized the possible tasks of assistive agents as follows: memory domains, daytime activities, personal information, social contact, health monitoring, and security. By helping with all these things, virtual assistants can be used to ensure a longer autonomous living for this target group.

This application was found to be accepted and well-usable for these kinds of target groups (Yaghoubzadeh et al., 2013). Within this application, the agent not only needs to process the tasks correctly, but must also demonstrate and use social skills too (Tsiourti, Joly, Wings, Ben Moussa, & Wac, 2014), since it is integrated in daily-life and is often used for vulnerable target groups.

All in all, there are several tasks in which a virtual agent can assist in the user's daily-life. Therefore, daily face-to-face interactions between the user and the agent are aimed for, where the agents give assistive advice and helps users manage their schedules and routines.

While assistive technologies are especially beneficial for people in need of support (Yaghoubzadeh et al., 2013), such technologies can also be helpful for all kinds of users. This is demonstrated through the enormous success of virtual speech-based assistance such as Siri (Apple), Alexa (Amazon), Watson (IBM), or OK Google (Google). Nevertheless, the benefit for people in need of support such as seniors or people with cognitive impairments is even

greater: "Assistive agents are expected to interact with users of all ages; however, the development of assistive intelligent technology has the promise of increasing the quality of life for older adults in particular" (Beer, Smarr, Fisk, & Rogers, 2015 p. 2).

1.2.6 Deviation from the Application Fields

All kinds of user groups can be regarded as potential target-groups of virtual agents. Nevertheless, as described above, the aimed target-group is highly depending on the chosen application. While educational applications mostly address younger users, seniors and other people in need of support are also often considered as target-groups with low literacy. This is mostly the case for health-related and assistive applications. It is decisive that the target-group has to be taken into account, in order to match the special needs and make these applications even more promising (Kopp et al., 2018).

What most applications have in common is that they require multiple interactions between the user and the agent. The exception to this is e-commerce and recommendations, where users mostly interact with the agent one at a time to buy a certain product or receive information about something in a specific situation and within a restricted time slot. Besides these application forms, the others call for longitudinal and steady interactions. Thus, the agent will either be ubiquitous in the user's daily-life (e.g., assistive living, where the user can rely on the agent's help whenever he or she wants to) or the user and the agent will interact on a regular basis (e.g., teaching sessions, where the agent is periodically used in specific classes). Owing to these longitudinal interactions, social processes become more crucial. Agents should engage users in a beneficial long-term relationship (Tsiourti et al., 2016). However, to date, there are only a few studies available that investigate long-term human–agent interaction and the social processes that occur within it.

2 HUMAN-AGENT INTERACTION

As described in the former chapter, virtual agents are ubiquitous applied in our daily-life within certain application fields. Therefore, it is important to examine humans' reactions towards virtual agents from a psychological perspective. This chapter covers the underlying processes that play a role during human-agent interactions. As a basic assumption the media equation theory and the computers are social actors paradigm are considered in this work. The underlying theory, empirical findings and its impact for the present work are discussed in this chapter. Moreover, as long-term effects with regard to the relationship between human and agents are investigated, the current state-of-art with regard to long-term studies is also discussed.

2.1 HUMANS SHOW SOCIAL REACTIONS TOWARDS VIRTUAL ENTITIES

Two decades of research support the assumption of the media equation theory (Reeves & Nass, 1996) or computer are social actors paradigm (CASA; Nass, Steuer, & Tauber, 1994), i.e., humans treat computers and virtual entities similar to other humans. Reeves and Nass (1996) postulate that media equals real life and that people unconscious react toward interactive behavioral cues presented by the media or artificial entities.

Empirical studies demonstrated that people perceive and evaluate virtual entities such as computers and virtual agents based on same person perception rules that exist in human-human interaction. Moreover, humans perceive computers as a teammate (Nass, Fogg, & Moon, 1996), apply gender stereotypes (Nass et al., 1994), are polite (Nass, Moon, & Carney, 1999) and apply rules of reciprocal behaviors (Nass & Moon, 2000). Assumptions like social facilitation can also be transferred to virtual entities (Rickenberg & Reeves, 2000). Those behaviors were not only observable towards computers, but also towards virtual agents (Hoffmann, Krämer, Lam-chi, & Kopp, 2009) and they are also seen to be transferable to other media such as smartphones or cars (Xu & Lombard, 2016).

According to the human tendency to treat computers and virtual entities as social actors (Nass et al., 1994), psychological theories found in human-human interaction can be transferred to interactions with virtual agents. To test these assumption, Nass et al. (1994) postulates that one picks a finding from human-human interaction and replaces at least one human in the theory and study method with a computer (or virtual entity) that features humanlike characteristics. When the same outputs occur that were assumed from human-human interaction, the media equation is regarded as supported (see Nass et al., 1994).

The social reactions human show towards virtual entities are seen to be mindless (in accordance to the theory by Langer, 1989) and unconsciously produced by the presence of social cues such as a human voice (Nass & Moon, 2000). A voice can elicit those social responses, even when it is obvious that the voice is pre-recorded, while a rock without any social cues evokes no such reactions. When enough social cues are presented, people are assumed to ignore other cues that hint to the asocial nature of the entity and apply social scripts (Nass & Moon, 2000). Those reactions are seen to be a product of the highly social nature of humans, as humans are seen to be ultra-social animals (Tomasello, 2014). Based on an evolutionary approach it is assumed that those reactions occur because people have not adapted those mechanisms to current technologies (Reeves & Nass, 1996). People are assumed to believe that what they see is social and respond accordingly (Reeves & Nass, 1996). Therefore, this reactions are assumed to be neither irrational, nor shown only by specific target-groups, but apply to everybody in high frequency and consistency (Reeves & Nass, 1996). It was even shown that people who have higher experiences with technical devices show stronger social reactions (D. M. Johnson & Gardner, 2007). The authors assume that more experienced people are more relaxed during the usage of technical devices and therefore show more unconscious/mindless social reactions (D. M. Johnson & Gardner, 2007). Thus, those reactions are no irrational behavior that is only present by people, who are not familiar with technical entities. The social cues that trigger those social reactions, might be a voice or speech-based outputs, an anthropomorphic appearance, interactive behavior or a fulfilled role that equals human roles (c.f. Fogg, 2000; Nass, Steuer, Tauber, & Reeder, 1993)

With regard to these cues, an early study emphasized that an anthropomorphic appearance elicits a higher perception of competence, social attractiveness and trustworthiness by humans than less anthropomorphic appearances (E. Lee & Nass, 2002). A perceived similarity between the human and the artificial entity is seen as a prerequisite for (Sundar & Nass, 2000). Although cues like a voice already elicit social reactions, it was found that an embodied character, which offers even more social cues, triggers more social outcomes (e.g. attracted more attention and higher sense of mutual awareness, (Appel, von der Pütten, Krämer, & Gratch, 2012). Thus, while social reactions are also observable towards entities with lower cues, an appearance that incorporates more social cues enhances these effects. This is in line with multiple studies that reported effects of an agent's appearance (including varying degrees of cues) on various variables (please consult Chapter 3 for an overview).

Besides the mindlessness explanation (see above), there are several possible explanations for the obtainable social reactions towards media, which are nicely summarized by Krämer, Rosenthal-von der Pütten and Hoffmann (2015). It has been discussed whether those reactions are just demand characteristics (Kiesler & Sproull, 1997) or real social behavior. Although the reactions can be criticized not to be truly social, they happen unconsciously and are denied by humans. In addition, the social reactions are depending on the entity's agency (Fox et al., 2015).

While partly contradicting evidences for the different explanations exist (c.f. Krämer et al., 2015), there is a call for long-term studies to investigate the development of such social responses toward virtual entities and find answers to the nature of social reactions towards virtual entities. It is questionable, whether users get used to the interaction with virtual social entities or whether the aforementioned effects are stable even over a longer period of time. Thus, long-term and field studies are seen to be highly promising:

> (…) future research should extend field and long-term studies in order to identify patterns and factors that determine the conditions under which people show social reactions and whether they ultimately even develop relationships with artificial entities. (Krämer et al., 2015 p. 154).

This examination is especially needed with regard to the aforementioned application fields, which expect the users to interact with virtual agents over a longer period of time.

Against the background of media equation research, the present dissertation relies on the methodological approach developed within CASA research. As derived above, the media equation claims that humans treat computers and virtual entities as social actors (Nass et al., 1994). This assumption makes it possible that psychological theories found in human-human interaction can be transferred to interactions with virtual agents. Nass et al. (1994) postulates that to test these assumption, one picks a finding from human-human interaction and replaces at least one human in the theory and study method with the computer (or virtual entity) that presents humanlike characteristics. When the same outputs occur that were assumed from human-human interaction, the media equation or CASA paradigm was proven (see Nass et al., 1994). Hence, theoretical assumptions and findings form psychological research from human-human interaction are used and applied to the relation between humans and agents, since similarities in the reactions are expected.

2.2 SOCIAL RELATIONS BETWEEN HUMANS AND AGENTS IN LONG-TERM INTERACTIONS

Although long-term studies are highly promising, only few of those with regard to virtual agents exist. The majority of long-term studies have been conducted by the research group around Timothy Bickmore. Those studies are discussed in the following in detail, because they built the basis for the long-term approach pursuit in the present dissertation.

To investigate the acceptance and efficacy of a virtual health advisor, Bickmore, Caruso et al. (2005) compared a virtual (relational) agent with a control condition with regard to their effects on the evaluation of the health advisor system and participants physical activity. Elderly participants interacted either with a virtual health advisor (embodied as a female agent with cartoon-stylized shades), which asked the participants to enter their steps walked and helped to self-monitor their physical activities, or used printed materials where they could note their physical activities over a period of two month. Results indicate that participants' steps walked increased in the virtual agent condition, while this was not the case for the control condition. Thus, interacting with a virtual agent over a substantial period enhanced the physical activity of elderly people. This enhanced activity did not result in higher life satisfaction, since no differences between the groups occurred. However, the study's results from qualitative statements further highlight that in the virtual agent condition, participants established a feeling of bonding with the agent and that this bonding motivated the participants to be physically active (Bickmore, Caruso, Clough-Gorr, et al., 2005).

In a second study, Bickmore, Gruber et al. (2005) examined the effect of emotional and relational communication strategies utilized by a virtual agent in a long-term study. Therefore, a relational agent, a non-relational agent and a control group (again only print material) have been compared after 6 weeks of use. The relational agent used strategies like a social dialogue, empathic feedback, meta-relational communication and humor. In contrast, the non-relational agent did not show these strategies. In both agent conditions the virtual agent assessed peoples physical activity and monitored their behavior. The main research goal was to investigate whether the relational agent enhances the feeling of bonding, resulting in higher motivation and physical outcomes. Their results indicate that indeed the relational agent evoked higher ratings of bonding compared to the non-relational agent. Moreover, participants liked the relational agent more and state a higher intention to keep on working with it. However, contrary to the initial assumptions (and findings presented in the aforementioned study by Bickmore, Caruso et al. (2005)), no differences between both agent conditions were found with regard to the physical activity of the participants. However, again,

an agent applied as health advisor enhanced the physical activity in general, but is not affected by emotional communication strategies (Bickmore, Gruber, & Picard, 2005). Overall, the results of this study emphasized that emotional communication strategies used by a virtual agent can enhance the feeling of bonding, the liking of the agent and participants' usage intention.

Doering, Veletsianos and Yerasimou (2008) conducted a long-term study to investigate how students evaluate the usefulness of virtual agents in a their learning process. Students of an online course had the chance to interact with a virtual agent. During the course the agent answered questions about the course content and interact with the students during the classes. Results of two focus groups indicated that participants were disappointed about the support of the agent during their learning process, since they perceived only little topic-wise support to enhance their learning outcomes. But in contrast, students did also expressed that the agent gave them humanized social support (Doering et al., 2008). Thus, one implication from this study might be that virtual agents can rather lend social than content-related support. Additionally, participants explained that they wished to have control over individual characteristics of the agent, i.e. they wanted to customize the virtual agent's features (Doering et al., 2008). The effect of customization (with regard to the agent's appearances) is discussed in Chapter 3.3.3.

Furthermore, it was shown that variety in a virtual agents dialogue is beneficial with regard to long-term interactions (Bickmore & Schulman, 2009). Bickmore and Schulman (2009) tested the perception of repetitiveness in a long-term study where participants interacted with the agent at least for 40 days. They found that an agent who varied its dialogue structure was perceived as less repetitive what further lead to higher physical activity by the participants.

Additionally, it was tested how participants perceive a virtual agent in a long-term interaction (of about 28 days) that declares itself even more human by means of telling personal backstories (Bickmore, Schulman, & Yin, 2010). Results indicate that while people perceive the agent that tells stories about itself more enjoyable, they showed enhanced physical activity when the interacted with an agent who tells stories about a third person. However, the initial positive effect of the stories seems to vanish over time, since participants overall showed a decrease of completing a session with the agent, as well as lower levels of engagement and enjoyment (Bickmore, Schulman, et al., 2010).

Furthermore, the effect of an embodied character compared to a solely text-based interaction was tested, in order to check human responses to social cues over time (Pfeifer &

Bickmore, 2011). Results of this study demonstrate that users, who interacted over six weeks with an embodied character instead of a text based interaction, were even more likely to state their walked steps the more steps they walked (Pfeifer & Bickmore, 2011). When people interacted with a solely text-based agent, the number of steps walked at the prior day influenced the likelihood to state the walked steps, but this effect decreases over time. In contrast, for people who interacted with an embodied character, the mentioned effect increases over time (Pfeifer & Bickmore, 2011). Thus, overall this study demonstrates that people are more ashamed to state that they have not been physically active when an embodied character is presented. Therefore, an embodied agent might trigger more social processes and this effect was found to be even stronger over time.

Overall, those studies give relevant implications for the design of a virtual agent applied in a long-term interaction. A virtual agent should be designed with a relational dialogue, since it enhances the feeling of bonding, which lead to higher motivation (Bickmore, Caruso, Clough-Gorr, et al., 2005). In addition, as feelings of repetitiveness can especially occur within long-term and multiple interactions, the dialogue of a virtual agent has to be variable, to prevent from feelings of repetitiveness, which also enhance the effectiveness of the agent (Bickmore & Schulman, 2009). Personal backstories of virtual agents might first be perceived as enjoyable, but this effect vanishes over time (Bickmore, Schulman, et al., 2010). Additionally, a virtual character was found to produce stronger social effects over time (Pfeifer & Bickmore, 2011). However, this study did not test the interpersonal effects with regard to a relationship between the agent and the user as prior studies partly did. To conclude, a virtual agent should be design relational and with high variability in its dialogue and an embodied virtual character might enhance the social effects.

In most studies described above, users were only able to interact with the agent via pre-scripted answer options. So far no solely-speech based interaction, where the agent and the user interacted via speech has been conducted. However, the speech-based interaction is one major advantage of virtual agents, especially with regard to specific target-groups. Likewise, a reciprocal speech-based interaction might trigger even more social cues, since the communication pattern is more humanlike. Therefore, a long-term study is needed in which human and agent communicate solely via speech.

Moreover, previous studies did not investigate the processes with regard to social relations and their development over time. Most studies demonstrated that users feel connected to the agent within a long-term interaction (Bickmore, Caruso, Clough-Gorr, et al., 2005; Bickmore, Gruber, & Picard, 2005; Bickmore, Pfeifer, et al., 2010; Ring, Barry,

Totzke, & Bickmore, 2013) and that social support of a virtual agent is important (Doering et al., 2008). However, within those studies mainly the acceptance and effectiveness of different agent characteristics were investigated over time, while there was no focus on social processes, although results indicate the importance of this. In the mentioned studies mostly a pre-post comparison of user's bonding was used. Therefore no assumption over the development over multiple interactions can be made. Moreover, did the number of interactions and time span in which the participants of the mentioned studies interacted with the agent vary a lot within the studies, since users were free in their usage of the agent (owing to the used field studies). Therefore, no controlled examination of the interpersonal relationship can be made, as interaction time is correlated with affection and closeness (McAdams & Constantian, 1983). Hence, the studies highlight that people build interpersonal relationships with the virtual agent, but they did not explain under which circumstances these relationships occur and how they develop over time.

Building on findings from human-human interaction, the perceived similarity (Byrne, 1997; Montoya, Horton, & Kirchner, 2008; Urber, Degirmencioglu, & Tolson, 1998) and physical attractiveness (Berscheid & Walster, 1974; Poulsen, Holman, Busby, & Carroll, 2013; Walster, Aronson, & Abrahams, 1966) were found to be important factors for relationship development. Both factors can be manipulated in an agent's appearance. However, the effect of appearance in long-term interactions on social outcomes like the perceived bonding, which has been shown to be a crucial factor in long-term interactions, has not been investigated yet. The current dissertation thus aims to fill this research gap by means of considering the impact of virtual agent's appearance on social evaluation and bonding in long-term interactions. Therefore, the following Chapter summarizes the state-of-the-art on appearance research.

3 VIRTUAL AGENT'S APPEARANCE

The effects of a person's outer appearance have been demonstrated in human–human interaction. Especially the attractiveness stereotype supports the impact of appearance variables (Dion, Berscheid, & Walster, 1972). Although the same principles are also applicable for virtual agents (Khan & De Angeli, 2009; Sobieraj & Krämer, 2014), the appearance of a virtual character can be more multifarious than that of a human. Since virtual agents are fictional and computer-generated, the design space of a virtual agent's appearance can take up endless forms that can have different effects on the user's perception of the agent and the interpersonal relationship between the user and the agent. Therefore, the effects of these manifold appearance variables in interpersonal interactions need to be investigated. From human-human interaction appearances is seen to transport many information about the interlocutors perception: "our appearance telegraphs more information about us than we would care to reveal on a battery of personality inventories" (Berscheid & Walster, 1974 p. 158). In line with this, regarding virtual agents there are already many studies that present the effect of a virtual agent's appearance (see Table 2 for an overview). However, most studies explore these effects in an unsystematic way by comparing virtual agents that differ in multiple appearance variables instead of manipulating it systematically. In order to explore the effect of different appearance variables more systematically, these have to be categorized first based on previous research. Thus, a categorization of different appearance variables is needed, on which later research can be built. In the following Chapter, first the constructed categorization that builds on prior research and then the empirical evidence regarding appearance effects are discussed. Moreover, explanations that were assumed with regard to appearance effects got summarized and in the end shortcomings from prior research are presented and implications for the present dissertation are derived.

3.1 CATEGORIZATION OF APPEARANCE VARIABLES

Since the design space of a virtual agent's appearance contains endless forms of multiple variables, an extensive literature review was performed to gather the main variables of interest. Different terms like appearance, look or visual design have been used to find literature that consult appearance factors of a virtual agent. As claimed above, also for virtual agents different synonyms are common, therefore all this terms (see Chapter 1.1) have been included in the literature retrieval.

Table 2.
Overview over literature used for the construction of the categorization of appearance variables

Author	Primary research objective	Method	Sample	Appearance Variables	Variables according categorization	Main Dependent Variable	Application context
Embodiment							
(Terada, Jing, & Seiji Yamada, 2015)	Which appearance of an product recommendation virtual agent is effective in making recommendations in an online shopping environment ?	Experiment (Within-Subject Design)	N = 41 (graduate students, Japan)	Text, Virtual Human, Robot-like agent, Dog-like Agent, Real Person (Video), Buddha-statue	Embodiment, Species, Realism	Buying intention, Impression of the agent	Online Shopping / Product Recommendation
Shiban, Schelhorn, Jobst, Hörnlein, Puppe, Pauli & Mühlberger (2015)	Does a virtual tutor influences the student's exam performance, interest in the course material and enjoyment during learning and unter which circumstances?	Experiment (Between Subjects Design)	N = 108 (students, Germany)	feminine & attractive, male & older, no agent	Embodiment, Feature Specifications (Socio-demographics)	exam performance, impression of the agent	Learning enviorment
(Sträfling et al., 2010)	Does solely the presence of an agent has an effect on motivation and learning even if the agent does not directly teach the learning material?	Experiment (Between Subjects Design)	N = 45 (university students, Germany)	no agent vs. humanoid agent vs. cartoon-like rabbit	Embodiment, Species, Realims	Learning outcomes, mood of the learners, impression of the learning program and the pedagogical agent	Learning environment

(Koda & Maes, 1996)	How do people perceive different appearances in an interactive entertainment enviroment? Does peoples impression of the faces differ by their sex or opinion about personification?	3 Pilot Studies & Online Experiment	N = 157 (mixed sample)	Face (Woman vs. white square), Sex (Woman vs. man), Humanity (Man vs. Dog), Realism (Real man vs. caricature man vs. smiley)	Embodiment, Species, Realism, Feautre Specification (Socio-demographics)	Impression of the agent, engagingness, level of comfort	poker game/ entertainment
Species							
(James et al., 2015)	To which extent will the effects of humanness and realism on people's responses to neutral faces in mediated messages be determined by patterns of brain activation in the whole brain and specifically in regions of the face network?	fMRI study	N = 21 (healthy adults)	Human vs. Animal, Realistic vs. Cartoon	Species, Realism	brain activity	non
(Terada, Jing, & Seiji Yamada, 2015)*	Which appearance of an product recommendation virtual agent is effective in making recommendations in an online shopping environment?	Experiment (Within-Subject Design)	N = 41 (graduate students, Japan)	Text, Virtual Human, Robot-like agent, Dog-like Agent, Real Person (Video), Buddha-statue	Embodiment, Species, Realism	Buying intention, Impression of the agent	Online Shopping / Product Recommendation
Krumhuber, Kappas, Hume, Hall & Aylett (2015)	How do adults and children infer human qualities of virtual characters that vary in appearance from non-human to very human?	not specified	N = 75 (adults and children, not defined)	Humanlikeness (Blob, cat, cartoon and human)	Species, Realism	impression of the agent, inference of human qualities	non/ not defined
Li, Kizilcec, Bailenson & Ju (2015)	Do students retain more information when instructed with a video or an animation of a human instructor? Do students retain more information when instructed with a video of a	Experiment (Between Subjects Design)	N = 40 (undergraduates, USA)	Morphology (human vs. robot), Realism (non-virtual vs. virtual)	Species, Realism	Knowledge recall, social presence, interpersonal attraction	Learning environment

	human instructor or a video of a robot?						
Tsiourti, Joly, Wings, Moussa & Wac (2014)	Are ECAs accepted as assistive companions in the living environment of older adults? What are users' expectations towards a "useful" virtual assistive companion? What are users' expectations towards a "socially intelligent" virtual assistive companion in terms of "social skills" communication behavior and character traits? What are users' expectations towards the appearance of a virtual assistive companion?	Focus Groups & individual interviews	N = 20 (elderly, Netherlands and Switzerland), N= 12 (care professionals, Netherlands), N=2 (neuropsychology experts, Switzerland)	no manipulation (different paper mock-ups: Smiley, Males, Children, Females)	Species, Realism, Feature Specifications	impression of the agent	assistive living
(Bergmann et al., 2012)	How do warmth and competence ratings change from a first impression after a few seconds to a second impression after a longer period of human-agent interaction depending on manipulations of agents' visual appearance and nonverbal behavior ?	Experiment (Between subjects Design with repeated measures)	N = 20 (university students, Germany)	Robot vs. Human	Species	Warmth & Competence	not defined
(Sträfling et al., 2010)	Does solely the presence of an agent has an effect on motivation and learning even if the agent does not directly teach the learning material?	Experiment (Between Subjects Design)	N = 45 (university students, Germany)	no agent vs. humanoid agent vs. cartoon-like rabbit	Embodiment, Species, Realism	Learning outcomes, mood of the learners, impression of the learning program and the pedagogical agent	Learning environment

(Koda & Maes, 1996)	How do people perceive different appearances in an interactive entertainment environment? Does peoples impression of the faces differ by their sex or opinion about personification?	3 Pilot Studies & Online Experiment	N = 157 (mixed sample)	Face (Woman vs. white square), Sex (Woman vs. man), Humanity (Man vs. Dog), Realism (Real man vs. caricature man vs. smiley)	Embodiment, Species, Realism, Feature Specification (Socio-demographics)	Impression of the agent, engagingness, level of comfort	poker game/ entertainment
Realism							
Matsui & Yamada (2017)	Which dimensions have a high emotional contagion ability?	Online Experiment (not defined)	N = 200 (mid aged, Japan)	Sex (male vs. female), Ethnic group (Caucasian vs. Asian), Abstraction (high vs. low)	Realism, Feature Specification (Socio-demographics)	Impression of the agent, emotional contagion	non/ not defined
Schwind, Leicht, Jäger, Wolf & Hence (2017)	Is there an uncanny valley of virtual animals? Do very realistic virtual pets tend to cause a similar aversion as humanlike characters?	2 Online Survey Experiments	Study 1: N = 339 (university students, Germany), Study 2: N = 214 (university students, Germany)	Realism(Reference Photo, HQ 3d Model with realistic fur, HQ 3D model with simple fur, HQ 3D model without fur, Low Poly 3D model, Ultra low Poly 3D Model, Toon Model with painted texture, vector Style 2D model), stylization of proportions	Realism (Rendering, Stylization)	impression of the agent (naturalness & aesthetic)	non/ not defined
Volonte,Babu, Chaturvedi, Newsome, Ebrahimi, Roy, Daily & Fasolino (2016)	How does the appearance of a virtual human affect the emotional response of participants in a medical virtual reality system that was designed to educate users in recognizing the signal and symptoms of	Experiment (Mixed Design)	N = 62 (students, USA)	Rendering (Realistic rendering, Cartoon rendering & Sketch rendering)	Realism (stylization)	Physiological arousal, valance of emotion, social presence	medical application

	patient deterioration?						
van Wissen, Vinkers & van Halteren (2016)	How can a virtual coach for chronic patients, that is able to support and stimulate behavior change be developed? Which visual characteristics of the agent yield the highest user acceptance among the target population?	Experiment (Mixed Design)	N = 65 (seniors, Caucasian)	Age (Younger vs. Older) + Familiar(actual nurse's photo), Realism (photorealistic vs. stylized)	Feature Specifications (Socio-demographics), Realism (Stylization)	User acceptance, attitudes towards use	medical application
(James et al., 2015)	To which extent will the effects of humanness and realism on people's responses to neutral faces in mediated messages be determined by patterns of brain activation in the whole brain and specifically in regions of the face network?	fMRI study	N = 21 (healthy adults)	Human vs. Animal, Realistic vs. Cartoon	Species, Realism	brain activity	non
Schwind, Wolf, Henze & Korn (2015)	How do people design faces to be liked or rejected? Do they use stylized cartoon faces and how do they control the visibility of details if they have the choice?	Online Study (Design Process Testing)	N = 431 (adults, mainly Germany)	arbitrarily face, uncanny, repulsive face, attractive heroine, attractive hero, female villain, male villain	Realism	Usage Parameter of the design software, self-assessment of the success of the design process	non/ not defined
(Terada, Jing, & Seiji Yamada, 2015)	Which appearance of an product recommendation virtual agent is effective in making recommendations in an online shopping environment ?	Experiment (Within-Subject Design)	N = 41 (graduate students, Japan)	Text, Virtual Human, Robot-like agent, Dog-like Agent, Real Person (Video), Buddha-statue	Embodiment, Species, Realism	Buying intention, Impression of the agent	Online Shopping / Product Recommendation
Krumhuber, Kappas, Hume, Hall & Aylett (2015)	How do adults and children infer human qualities of virtual characters that vary in appearance from non-human to very human?	not specified	N = 75 (adults and children, not defined)	Humanlikeness (Blob, cat, cartoon and human)	Species, Realism	impression of the agent, inference of human qualities	non/ not defined

Li, Kizilcec, Bailenson & Ju (2015)	Do students retain more information when instructed with a video or an animation of a human instructor? Do students retain more information when instructed with a video of a human instructor or a video of a robot?	Experiment (Between Subjects Design)	N = 40 (undergraduates, USA)	Morphology (human vs. robot), Realism (non-virtual vs. virtual)	Species, Realism	Knowledge recall, social presence, interpersonal attraction	Learning environment
Zell, Aliagra, Jarabo, Zibrek, Gutierrez, McDonnell & Botsch (2015)	How do the combination of different levels of stylization in shape and material affect the perception of a virtual character across different expressions?	not specified (within-subjects design)	Study 1: N = 22 (not defined), Study 2: N = 20 (not defined), Study 3: N = 21 (not defined)	Study 1: material, shape style, expression; Study 2: Shading & Lighting, Study 3: Texture	Realism	impression of the agent	non/ not defined
Robertson, Solomon, Riedl, Wicklin Gillespie, Chociemski, Mater & Mohan (2015)	How does different level of socio-economic status and health literacy influence the creation process of a embodied conversational agent as an e-Coach?	Focus Groups	sample size not defined (mainly male Caucasian diagnosed with prostate cancer)	Rendering Style (Photograph, Vectorized Photo, Realistic Illustration, Cartoon)	Realism	impression of the agent	medical application
(Ring et al., 2014)	How does shading Style and Character Proporties affect the user's attitude toward the virtual agent? Are there any correlations with the task domain?	2 Online Experiments (Within-Subject Design)	Study 1: N = 67 (not defined), Study 2: N = 47 (not defined)	Shading Style (Realistic vs. Toon), Character Properties (Realistic vs. Toon)	Realism (Stylization)	Attitude toward agent	Medical Application/ Social Application
Zibrek & McDonnell (2014)	Can we build a reliable dialogue resource to convey a range of complex personalities grounded in psychology research? Are certain personalities considered more appealing than others? Could changing the render style alone alter the perception of personality?	Experiment (Mixed Design)	N = 36 (students, Ireland)	Rendering Style	Realism	impression of the agent	non/ not defined

Tsiourti, Joly, Wings, Moussa & Wac (2014)	Are ECAs accepted as assistive companions in the living enviroment of older adults? What are users' expectations towards a "useful" virtual assistive companion? What are users' expectations towards a "socially intelligent" virtual assistive companion in terms of "social skills" communication behavior and character traits? What are users' expectations towards the appearance of a virtual assistive companion?	Focus Groups & individual interviews	N = 20 (elderly, Netherlands and Switzerland), N= 12 (care professionals, Netherlands), N=2 (neuropsychology experts, Switzerland)	no manipulation (different paper mock-ups: Smiley, Males, Children, Females)	Species, Realism, Feature Specifications	impression of the agent	assistive living
McDonnell, Breidt & Bülthoff (2012)	Are there differences in how truthful we perceive real and virtual humans? Does rendering style influence trust? If we choose a stylized rendering, will this affect how appealing, friendly, or reassuring the character appears?	Experiments (Mixed Design)	Study 1: N = 34 (not defined) Study 2: N = 12 (not defined)	Rendering Style (Toon Pencil, Toon Flat, Toon Shaded, Toon Bare, Toon CG, Human III, Human Basic, Huamn SSS, Human HQ1, Human HQ2)	Realism	impression of the agent	not defined
(Sträfling et al., 2010)	Does solely the presence of an agent has an effect on motivation and learning even if the agent does not directly teach the learning material?	Experiment (Between Subjects Design)	N = 45 (university students, Germany)	no agent vs. humanoid agent vs. cartoon-like rabbit	Embodiment, Species, Realism	Learning outcomes, mood of the learners, impression of the learning program and the pedagogical agent	Learning environment
Saidatul Maizura, Farah, Nabila, Noorizdayantie, Zuraidah, Omar, Hanafi, Fong & Wong (2010)	How does realism of a pedagogical agent influence student's achievement during online learning in terms of sex ?	Quasi-Experiment (Between-Subjects Design)	N = 130 (secondary school students)	Realism (Unrealistic, Moderately realistic & highly realistic)	Realism	Learning Achievement	Learning environment

Baylor & Kim (2004)	Which impact does sex, ethnicity, realism and instructional role of a virtual pedagogical agent have in learning situations?	Experiment (Between Subjects Design)	Study 1: N = 312 (students, USA), Study 2: N = 229 (undergraduates, USA)	Realism (realistic vs. cartoon-like), ethnicity (black, white), sex (male, female), instructional roles (expert, motivator, mentor)	Realism, Feature Specification (Socio-demographics)	self-regulation, self-efficacy, learning outcomes	Learning environment
(Koda & Maes, 1996)	How do people perceive different appearances in an interactive entertainment environment? Does peoples impression of the faces differ by their sex or opinion about personification?	3 Pilot Studies & Online Experiment	N = 157 (mixed sample)	Face (Woman vs. white square), Sex (Woman vs. man), Humanity (Man vs. Dog), Realism (Real man vs. caricature man vs. smiley)	Embodiment, Species, Realism, Feature Specification (Socio-demographics)	Impression of the agent, engagingness, level of comfort	poker game/ entertainment
Feature Specifications							
Matsui & Yamada (2017)	Which dimensions have a high emotional contagion ability?	Online Experiment (not defined)	N = 200 (mid aged, Japan)	Sex (male vs. female), Ethnic group (Caucasian vs. Asian), Abstraction (high vs. low)	Realism, Feature Specification (Socio-demographics)	Impression of the agent, emotional contagion	non/ not defined
Lunardo, Bessolles & Durrieu (2016)	Does the presence of corporate clothing on an OVA moderates the effects of OVA sex on attractiveness? How does such OVA attractiveness affect consumer online behavior?	Experiment (between-subjects design)	N = 252 (half students, France)	Sex (male vs. female), Corporate Clothing (presence vs. absence)	Feature Specifications (Socio-demographics, Styling)	Attractiveness, social presence, behavioral and purchase intention	online shopping

van Wissen, Vinkers & van Halteren (2016)	How can a virtual coach for chronic patients, that is able to support and stimulate behavior change be developed? Which visual characteristics of the agent yield the highest user acceptance among the target population?	Experiment (Mixed Design)	N = 65 (seniors, Caucasian)	Age (Younger vs. Older) + Familiar(actual nurse's photo), Realism (photorealistic vs. stylized)	Feature Specifications (Socio-demographics), Realism (Stylization)	User acceptance, attitudes towards use	medical application
Shiban, Schelhorn, Jobst, Hörnlein, Puppe, Pauli & Mühlberger (2015)	Does a virtual tutor influences the student's exam performance, interest in the course material and enjoyment during learning and under which circumstances?	Experiment (Between Subjects Design)	N = 108 (students, Germany)	feminine & attractive, male & older, no agent	Embodiment, Feature Specifications (Socio-demographics)	exam performance, impression of the agent	Learning environment
Khan & Sutcliffe (2014)	Which effect has the agent's appearance on user's perception and reaction and it's persuasiveness levels?	Experiment (Between Subjects Design)	not defined	Attractiveness	Feature Specifications	impression of the agent, social competence, social adjustment, anthropomorphism, performance	persuasion (Desert Survival Game)
Tsiourti, Joly, Wings, Moussa & Wac (2014)	Are ECAs accepted as assistive companions in the living enviroment of older adults? What are users' expectations towards a "useful" virtual assistive companion? What are users' expectations towards a "socially intelligent" virtual assistive companion in terms of "social skills" communication behavior and character traits? What are users' expectations towards the appearance of a virtual assistive companion?	Focus Groups & individual interviews	N = 20 (elderly, Netherlands and Switzerland), N= 12 (care professionals, Netherlands), N=2 (neuropsychology experts, Switzerland)	no manipulation (different paper mock-ups: Smiley, Males, Children, Females)	Species, Realism, Feature Specifications	impression of the agent	assistive living

Qui & Benbasat (2010)	How do demographic embodiments influence the users' social experiences of interacting with an anthropomorphic product recommendation agents?	Experiment (between-subjects design)	N = 188 (Caucasian & Asian)	Sex (male vs. female), Ethnic group (Caucasian vs. Asian)	Feature Specifications (Socio-demographics)	Social Presence, Perceived Enjoyment, Perceived Usefulness	Online Shopping / Product Recommendation
Kahn & DeAngeli (2009)	Does the attractiveness stereotype apply to the interaction with virtual agents?	Experiments (Within-Subjects & Between-Subjects Design)	Study 1: N =30 (students, England), Study 2: N = 48 (students, England)	Attractiveness	Feature Specifications	impression of the agent	non/ not defined
Baylor & Kim (2004)	Which impact does sex, ethnicity, realism and instructional role of a virtual pedagogical agent have in learning situations?	Experiment (Between Subjects Design)	Study 1: N = 312 (students, USA), Study 2: N = 229 (undergraduates, USA)	Realism (realistic vs. cartoon-like), ethnicity (black, white), sex (male, female), instructional roles (expert, motivator, mentor)	Realism, Feature Specification (Socio-demographics)	self-regulation, self-efficacy, learning outcomes	Learning environment
(Koda & Maes, 1996)	How do people perceive different appearances in an interactive entertainment environment? Does peoples impression of the faces differ by their sex or opinion about personification?	3 Pilot Studies & Online Experiment	N = 157 (mixed sample)	Face (Woman vs. white square), Sex (Woman vs. man), Humanity (Man vs. Dog), Realism (Real man vs. caricature man vs. smiley)	Embodiment, Species, Realism, Feature Specification (Socio-demographics)	Impression of the agent, engagingness, level of comfort	poker game/ entertainment
not defined							
Chattaraman, Kwo, Gilbert & Shim (2011)	Which representational characteristic of virtual agents do existing agent software providers employ in designing agents for various commercial and non-commercial web applications?	Content Analysis & Focus group	Focus groups: N = 25 (elderly, southeastern)	Slide show with example virtual agents (not defined)	–	preferences for usage in web-applications	commercial web-applications

Heidig & Clarbout (2011)	Do pedagogical agents facilitate learner motivation and learning? Do they make a difference? Under what conditions do pedagogical agents facilitate learner motivation and learning outcomes? When are they effective? How should pedagogical agents be designed to foster learner motivation and learning outcomes?	review	–	–	–	Motivation, Learning Outcomes	Learning environment
Baylor (2009)	What drives motivational and affective changes from observing or interacting with anthropomorphic agents?	Review	–	–	–	motivational outcomes	–
Gulz & Haake (2006)	How should a animated pedagogical agent be designed?	Review	–	–	–	motivational outcomes	Learning environment

When no virtual agent was used but an avatar that was controlled by a human, these papers have been excluded. From the resulting literature that is presented in Table 2 all used appearance variables have been extracted. When the research papers presented a definition of the appearance variable, this has also been added to the extraction of appearance variables. In the end, all extracted variables were sorted that have either been mentioned in the papers or were experimentally investigated in empirical studies. Based on descriptions and the presented stimulus materials synonym variables have been combined. Taking all used stimulus material and manipulated variables into account, a categorization based on that research and the authors' definitions of different variables (such as realism) was created. See the column "Appearance Variables" for a list of all used terms for the manipulated variables that have been used. Table 2 lists all the studies used for the categorization. Since many studies focused not only on one appearance variable but on multiple, and the table was sorted along the constructed categories, studies were listed several times (once for each category).

From the literature presented in Table 2, five main categories can be identified: (0) embodiment, (1) species, (2) realism, (3) 2D vs. 3D, and (4) feature specifications. Figure 1 gives an overview of the constructed categorization with its five main categories and all subcategories.

These categories are not distinct and they influence each other. For example, the type of species also has an impact on the perceived realism of the agent and affects the possibilities of the agent's feature specifications. When a humanoid character is chosen, features such as sex, clothing, etc. can be designed, while the available dispositions for zoomorphic or mechanical characters might be slightly different. In the following, the categories will be derived from prior research and the final category descriptions are presented in more depth.

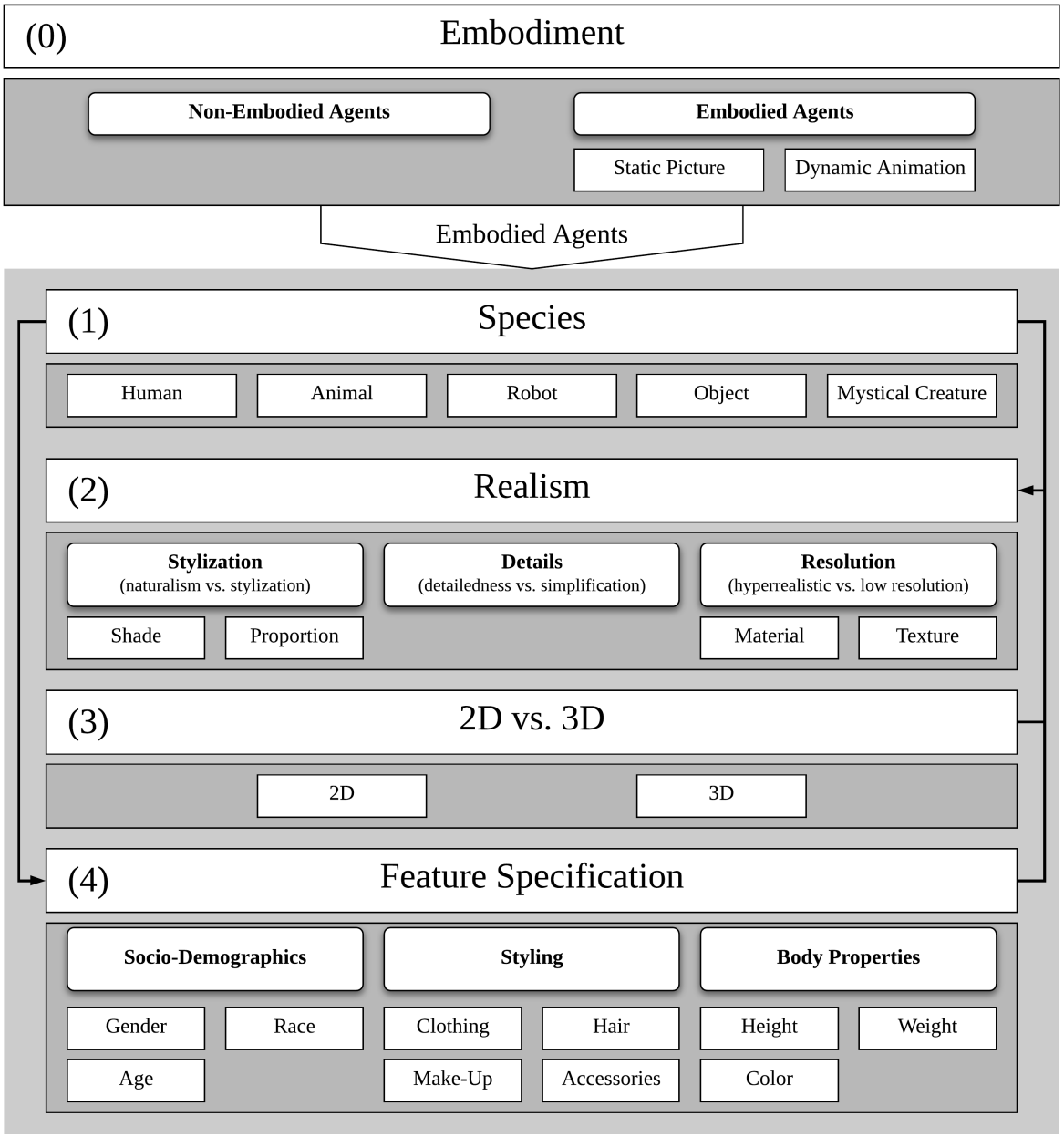


Figure 1. Overview over all categories that have been extracted from prior research.

3.1.1 (0) Embodiment

When it comes to the design space of a virtual agent, the first decision needs to be whether the virtual agent should have an embodied appearance or not. In the literature, the notion of embodiment can have several meanings (K. M. Lee, Jung, Kim, & Kim, 2006). Because embodied conversional agents are defined as computer interfaces with a virtual representation that are able to communicate humanlike (Cassell, 2000), the notion of embodiment is connected to a virtual representation. In a narrow sense, embodiment can be described as having a physical body (Ziemke, 2001), but this core definition excludes software agents, since they have no physical connection to the environment. However, an embodied virtual agent is characterized by a virtual physical body (Cassell, 2000) and, thus, having any kind of virtual body can be seen as embodiment. In this sense, either a static picture as an interface representation or a dynamic animated body is a form of embodied agents. The form of the represented body is usually not defined precisely, and although some authors mention that this body might have a humanoid or zoomorphic look (e.g. Cassell, 2000), this characterization is not exclusive. Thus, the appearance of an embodied agent can take up different forms and it can be represented by a human just as well as by an abstract ball.

By contrast, non-embodied agents do not own a virtual physical body and they communicate without any virtual representation. In this sense, recent speech-based systems like Apple's Siri or Microsoft's Cortana mostly do not have an embodied character and these agents are solely represented by a voice. Beside speech-based systems, also chat-bots, who only communicate via text, can be classified as non-embodied agents.

Since this work deals with the effects of appearance variables and their effects on social interaction, mainly embodied agents are of interest. When an embodied character is used, multiple factors can be manipulated and these factors are described within the following categories.

3.1.2 (1) Species

Most of the embodied agents used are humanoid (c.f. Chattaraman, Kwon, Gilbert, & Shim [2011], who conducted a content analysis of 64 different e-commerce websites that present virtual agents and found that 91% of them were humans or humanoid), while virtual agents could easily take various other forms. Koda and Maes (1996), for instance, compared in their study a human face, a dog, and a smiley. In line with these different forms, Gulz and Haake (2006b) called this differentiation *humanness* and stated that agents could be designed

as human, animal, or another creature and nonliving object. Furthermore, the authors do not see these forms as distinct and mention combinations of them(Gulz & Haake, 2006b). Examining different degrees of realism, Sträfling et al. (2010) distinguished between zoomorphic and anthropomorphic agents and therefore compared a rabbit with a female human in their experimental study. Beside humans, animals, and objects, machinelike agents are also considered (e.g., Bergmann et al., 2012; Terada et al., 2015). Bergmann et al. (2012) compared a human agent with a robot agent, while Terada et al. (2015) investigated the following agent representations: real and virtual human, dog, buddha, text, and robot. In most computer games (e.g., "World of Warcraft") avatars and agents may also be represented as mystical creatures. Therefore, mystical creatures (e.g., dragons, monsters, or fairies) are also possible forms of virtual agents. Ring et al. (2014) sum up these differentiations using the term *species* and list humanoid, animal, and robot as examples for this category. Sticking to this term, the described category of species is defined as different classes of individuals that have common attributes and are identified by a common name such as humans, animals, or objects. There are at least five different types of species: humans, animals, robots, objects, and mystical creatures. As Gulz and Haake (2006b) already claimed, combinations of the species do exist, since most animals and objects are mainly designed in a humanoid way (e.g., Microsoft's Clippy that is a humanoid paper clip with eyes and mouth).

3.1.3 (2) Realism

Numerous studies examined the realism of the agents' appearance (e.g., Koda & Maes, 1996, James et al., 2015, Ring et al., 2014, Sträfling et al., 2010 & Terada et al., 2015). According to the Oxford Dictionary, realism is defined as, "the quality or fact of representing a person or thing in a way that is accurate and true to life" (Oxford Web Dictionary, 2018). In line with this lexical definition, realism was described as, "in a simple binary manner; things are more real, the more they look exactly like a real object" (James et al., 2015, pp. 109–110). James and colleagues argue further that photographs of humans or animals score higher on realism, while cartoon stylized pictures are less real (James et al., 2015). However this definition is quite vague, and since there is no universal characterization of realism, various studies focus on different dimensions of it.

While many studies compare a cartoon-stylized agent with a realistic agent, they mainly do not describe what kind of real appearance is used. It is obvious that there are multiple subdimensions that affect the perceived realism on a spectrum from unrealistic to highly realistic.

James and colleagues (2015) describe in their definition the opposite of a stylized character as a photograph of a real object (human or animal). Therefore, they combine the stylization of a character and its photorealism or also named *resolution* (e.g. Sträfling et al., 2010). Koda and Maes (1996) investigated the effect of realism by comparing a realistic and caricature male and a smiley face. However, again the authors' understanding of realism is not defined. The authors distinguished their used appearances into realism and humanity (Koda & Maes, 1996). Therefore, in some way the stylization of the character (realistic vs. caricature) is seen as part of realism, while the species or humanlikeness is also taken into account (human vs. smiley).

In line with these gradations, Sträfling et al. (2010) defined three realism dimensions. As already mentioned, the species of the agents is seen as one part of realism by comparing zoomorphic characters with anthropomorphic ones (Sträfling et al., 2010). Since in this categorization species is seen as a separate category, it is excluded from realism (although it can also affect realism).

Another dimension by Sträfling et al. (2010) is resolution, which is described by the poles low resolution and hyperrealistic. Further, the authors defined naturalness versus cartoon-likeness (Sträfling et al., 2010), which fits the comparison of realistic and caricature humans by Koda and Maes (1996). However, it should be mentioned that these dimension are not described in more detail and therefore it is unknown which features of appearance exactly cause a hyperrealistic look or a less cartoonlike look.

In contrast to many other authors (Baylor & Kim, 2004; James et al., 2015; Koda & Maes, 1996; Saidatul Maizura et al., 2010; van Wissen et al., 2016), Sträfling and colleagues (2010) named the lower endpoint of cartoon-likeness *naturalness* instead of *realism*. Similarly, Gulz and Haake (2006b) describe this dimension of realism as a subdimension of graphical style; they called it *naturalism* versus *stylization* and characterized it through artistic quality. The authors further describe this dimension as complex since there are various expressions of this appearance dimension and no simple linear relation in this design space (Gulz & Haake, 2006b). As a second subdimension of graphical style, the degree of detailedness versus simplification was listed (Gulz & Haake, 2006 b). In this sense, a photo is detailed and can be transferred into a semidetailed contour drawing by reducing the details. The degree of detail might also influence the stylization, while a naturalistic design could also have more or fewer details (Gulz & Haake, 2006b).

Ring et al. (2014) mention new characteristics of realism, which describe stylization in a more detailed manner: shading style and proportions. They explore the effect of a realistic

versus stylized shading and realistic versus stylized proportions (Ring et al., 2014). In accordance with that, Schwind, Wolf, Henze, and Korn (2015) describe enlarged or unrealistic properties as a common way of stylization (Schwind et al., 2015). In the same vein, many authors (e.g. James et al., 2015; McDonnell, Breidt, & Bülthoff, 2012; Volante et al., 2016) describe stylization as a type of shading or drawing. Volante and colleagues (2016) distingue shading further into cartoon and sketch, where cartoon is characterized by a cel-shaded character and a sketch stylization appears hand drawn. However, both types (cartoon and sketch) are characterized by a distinct type of shading or level of details. Therefore, the degree of stylization (naturalness vs. stylization) may be characterized by the two subcategories: shade and proportions.

Referring to Sträfling and colleagues' (2010) dimension of resolution (low resolution vs. hyperrealistic), other authors characterize a realistic appearance through photorealistic textures (McDonnell et al., 2012; Schwind et al., 2015). In line with this, Zell et al. (2015) investigated shape, shade, material, and texture as parameters of a virtual character. While shape and shade have already been characterized as subdimensions of stylization, according to McDonnell et al. (2012) and Schwind et al. (2015), material and texture define the resolution of a virtual character. Thus, the resolution of a virtual agent's appearance is characterized by two subcategories: material and texture. These range from an unrealistic/low-resolution material or texture to a photorealistic material or texture.

In sum, there is no universal definition of realism in the existing literature, but it can be seen as a spectrum where a virtual character is more realistic the closer it is to a real object (James et al., 2015) and is characterized by three subcategories: stylization, resolution, and details. While in the above-presented *Figure 1*, only the subcategories are listed, it is important to see the degree of realism as a spectrum with specific start and end points. Therefore*, Figure 2* presents the spectrum of realism with all the subcategories described earlier. A virtual agent's appearance can score differently on the listed subcategories and all those factors affect the perceived realism in a certain way. For example, a character can be designed with high resolution (realistic material and texture), while its proportions can be enlarged (e.g., unrealistic large eyes) and, thus, can be seen as highly stylized.

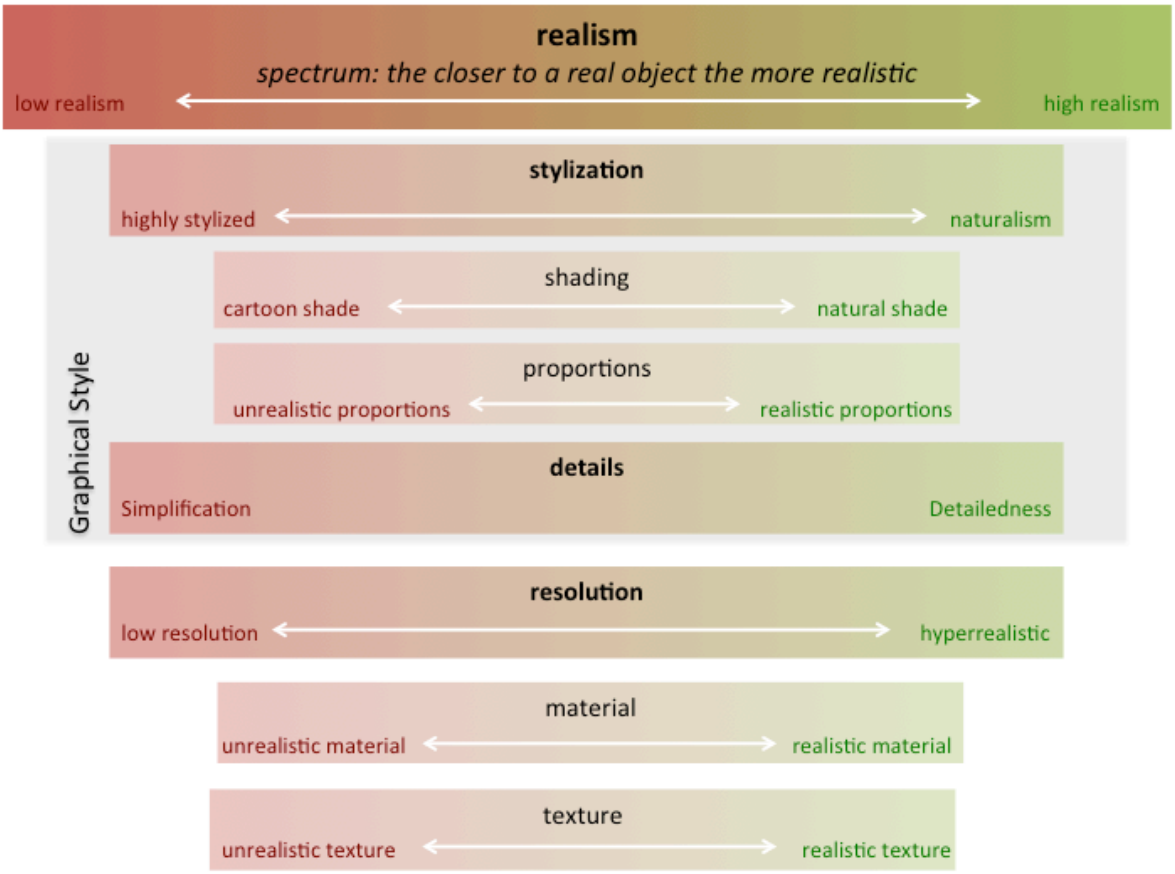


Figure 2. Definition of realism as a spectrum and the forms of all subcategories

Since the listed categories are not distinct ones, other main categories such as the species or the visual dimension (2D vs. 3D) also have an effect on the perceived realism.

In addition, it should be mentioned that all the subcategories presented are characteristic for appearance realism, while the behavior of a virtual agent can also affect its perceived realism (behavioral realism; e.g., Bailenson et al., 2005). However, since this work aims to categorize and explore appearance variables and their effects, behavioral realism will not be investigated in more depth and is therefore excluded from this categorization.

3.1.4 (3) 2D Versus 3D

While most of the modern embodied agents are 3D models (e.g., Ring et al., 2014 & Zell et al., 2015), it would also be possible to design a 2D character. Besides the presentation of 2D or 3D models within a 2D simulation (such as a computer screen), thanks to the rise of modern technologies (e.g., Oculus Rift) it is also possible to present these characters in a 3D environment. Since the current approach focuses on visual appearance only, the two former dimensions (2D vs. 3D presented in a 2D simulation) are of interest.

3.1.5 (4) Feature Specification

Especially for humanoid characters, there are various features that can be specialized in more detail. For instance, Ring et al. (2014) mention demographic parameters, selection of clothing and accessories, as well as hairstyle as further design decisions. Gulz and Haake (2006a), on the basis of the theory of physical personality (Branham, 2001), stated that features such as sex, race, hair, clothing, make-up etc. create an impression of personality. Therefore, even these detail decisions might have a big influence on the overall perception of the virtual agent. Within these specific features, there are different subcategories: First there are sociodemographic parameters (e.g., age, sex, or race), while the choice of clothing, hair, or make-up, for instance, can be summarized with the term *styling.* Additionally, feature properties like height, weight, or eye color can be seen as body specifications that characterize the perception of an agent.

Many studies (c.f. Baylor & Kim, 2004; Koda & Maes, 1996; Lunardo et al., 2016; Matsui & Yamada, 2017; Qiu & Benbasat, 2010; Shiban et al., 2015; van Wissen et al., 2016) focus on the sociodemographic parameters of the virtual agent. Most of these studies investigated the influence of sex (Baylor & Kim, 2004; Lunardo et al., 2016; Matsui & Yamada, 2017; Qiu & Benbasat, 2010; Shiban et al., 2015; Tsiourti et al., 2016) by comparing male with female agents. Ethnicity (Baylor & Kim, 2004; Matsui & Yamada, 2017; Qiu & Benbasat, 2010) and age (Shiban et al., 2015; Tsiourti et al., 2016; van Wissen et al., 2016) are also key factors of an agent's appearance that have been studied frequently.

Beside sociodemographic variables, the effect of styling (e.g., Lunardo et al., 2016, who investigated the effect of corporate clothing) has been examined and is one factor of an agent's appearance regarding feature specifications.

With regard to body properties, the attractiveness of a virtual agent has often been discussed (Kahn et al., 2012; Shiban et al., 2015; Sobieraj & Krämer, 2014).

Although most of these features belong to humanoid characters, they are also of interest for other species (e.g., the specific race of an animal as well as fur patterns are possible design decisions). Thus, the choice of species has an impact on the potential feature specifications.

3.2 EFFECTS OF APPEARANCE VARIABLES

Based on the categorization described earlier, state-of-the-art research findings are presented for the main categories. Since there is no empirical evidence to date for the comparison of 2D and 3D appearances, no subchapter exists for this category ("2D vs. 3D").

3.2.1 Embodiment

Prior research demonstrates that an embodied character is very beneficial. For instance, playing an online poker game against an embodied face was perceived as more engaging and comfortable (Koda & Maes, 1996).

Furthermore, Baylor and Ryu (2003) showed that a human voice presented together with an anthropomorphic character (static or animated) could enhance credibility within a learning environment compared with presenting the voice only. In line with this research, many studies examined the persona effect, which assumes an animated embodied agent's effect on learners' motivation and performance (Lester et al., 1997). However, this effect has been controversial (Shiban et al., 2015). Shiban and colleagues did not find any significant difference in performance between a learning environment with regard to embodiment (Shiban et al., 2015). Sträfling et al. (2010) also found no difference between an embodied agent (either human or rabbit) and a voice-only condition in users' learning outcomes, but a difference was found with regard to the time spent with the learning material (learning program). The zoomorphic embodied agent attracted people to spend more time in the learning program compared with a solely speech-based system (Sträfling et al., 2010). Still, controversial findings with regard to learning outcomes were found, but overall an embodied agent seems to be perceived more positively.

A study in the context of product recommendations showed that presenting only a text evokes less buying intention than the presentation of textual recommendations together with a picture of a virtual dog or human, but no difference was found between the non-embodied recommendation agent and an object (Buddha) or robot agent (Terada et al., 2015). In addition, an overall meta-analysis by Yee, Bailenson, and Rickertsen (2007) demonstrated that an embodied character evokes a more positive social interaction than a non-embodied one.

In sum, although there are partly controversial findings regarding learning outcomes in the literature, having an embodied agent seems to be beneficial for the human–agent interaction, as it enhanced the social interaction (Yee, Bailenson, & Rickertsen, 2007), was found to attract longer time spent with the material presented by the agent (Sträfling et al., 2010), and evoked higher buying intentions (Terada et al., 2015).

3.2.2 Species

Although most real-world applications use humanoid agents (Chattaraman, Kwon, Gilbert, & Shim, 2011) and people often aim to reproduce a human being on the screen, a

virtual agent could also be a different species such as robot, animal, or object. To investigate the effects of different appearances, prior studies compared different species (mostly in an unsystematic way and without naming the category *species*) with each other. Those studies either use a robot, an animal, or an object to investigate the effects and compare it with a human appearance. Thus, in all studies, a humanoid agent is present and is compared with other species. In the following, the comparisons of human versus animal, human versus Robot, and human versus object are presented separately.

Human Versus Animal

Animals are a big part of many people's lives. Since many people are used to sharing their social life with their pets, zoomorphic agents might be a promising option and has already been investigated.

Koda and Maes (1996) showed that when only pictures of an embodied agents are evaluated, differences between a humanoid and zoomorphic agent are obtainable, since a dog character was evaluated as being less intelligent than a human, but more likable and engaging. Nevertheless, these findings could not be replicated when the agents were used in an actual poker game, since there were no differences in the users' evaluation.

Terada and colleagues (2015) investigated appearance effects of a product recommendation agent and compared six different appearances in their effects on the perception of the agent and users' buying intention. All agents were static pictures combined with textual product recommendation information. In the comparison of humans and animals, no differences were found by the authors. Users who interacted with a doglike agent did not differ from users of both humanoid agents (virtual human and real human) in their buying intention (Terada et al., 2015).

While Sträfling et al. (2010) also found no significant difference between a virtual human and a virtual rabbit on users' learning outcomes, they found a significant difference in the evaluation of the agent and the learning program. The zoomorphic agent was evaluated most positively and as more friendly.

In sum, zoomorphic agents were partly found to be more positive than humanoid agents. They might be perceived as more friendly and engaging than humanoid agents, while this effect is not necessarily applicable to behavioral outcomes such as learning outcomes (Sträfling et al., 2010), buying intention (Terada et al., 2015), or experiences during a game (Koda & Maes, 1996).

Human Versus Robot

Robots are also becoming increasingly integrated in our lives. Children are used to toy robots (Pleo, Aibo, or Dash and Dot), robots are used at exhibitions (e.g., Nao or Pepper) and are also integrated in households (vacuum robots such as Roomba). Since robots and virtual agents are both fictional and artificial entities, a robot appearance might be an appropriate appearance for representing a virtual agent. Although one could argue that a machinelike appearance fits better with the artificial nature of a virtual agent, there is evidence that humanlike agents evoke more positive outcomes.

Terada et al. (2015) found differences in people's buying intention when people interacted with a robot agent and when they received a product recommendation from a virtual human. A virtual human achieved higher buying intention values than the virtual robot did. Similar results were found in a learning application. By investigating the use of a virtual agent as an instructor, which is giving a lecture, it was found that the experience during the lecture and the perceived presence, liking of the teacher and its presentation skills, were evaluated better for a humanoid agent than for a robot agent (J. Li, Kizilcec, Bailenson, & Ju, 2015).

While both of these studies found higher outcomes for humanoid agents, there is also evidence that a machinelike agent is evaluated better. Bergmann et al. (2012) investigated whether the impression of a virtual agent can change after the first impression and compared a humanoid agent with a robot agent. Participants evaluated the agent once after a few seconds and a second time after its presentation. Results indicate an interaction effect of the evaluation point and the presented appearance. While in the first impression, the robot was evaluated warmer than the human, no differences were noted at the second measuring point (Bergmann et al., 2012). Results indicate that people can revise their impression and that this might be dependent on the presented appearance. Nevertheless, after a while no differences between humanoid and machinelike agents were obvious.

In sum, these results indicate a tendency that humanoid agents will evoke more positive outcomes than a machinelike agent. Nevertheless, the first impression of a machinelike agent was found to be warmer than the impression of a humanoid one.

Human Versus Object

Objects might be more diverse than the other species and this makes it even harder to compare them with other appearances. In contrast to interactions with humans, animals, and robots, people are usually not used to communicating and interacting with objects and might

perceive an object more as a functional thing. To overcome this gap, when objects in prior research were used, they were not totally abstract but were humanized. The most famous agent that was represented by an object is Microsoft's Clippy, a humanized paper clip used as a helping assistant in Microsoft's Office applications. From this, one could derive that objects are nicely adaptable to the given context (like the paper clip that was used in an office application). However, the objects that were used in prior research were not adapted to the context and only little research on the comparison of objects and humans exists with contradicting findings.

In the study of Terada et al. (2015), the Buddha statue agent that can be seen as an object (although the authors did not call it this) evoked lower buying intention than both of the humanoid agents used (virtual human and real person). While here a humanoid agent was found to be more effective with regard to users' buying intention, depending on the target group, objects have also been found to be perceived as less stiff than a humanoid agent and more positive.

Krumhuber et al. (Krumhuber, Kappas, Hume, Hall, & Aylett, 2015) investigated the effect of a virtual agent's appearance on the person perception and mental capabilities of the agent by comparing an abstract egg-shaped character, a humanoid cat, a cartoon-stylized human, and a realistic human. They further examined age-related differences, since children and adults participated in their study. Children were found to evaluate the agents more positively in general and the results demonstrated an interaction effect of the target group and appearance. Especially the abstract object character was evaluated better by children than by adults for most variables such as trustworthiness and perceived intelligence. Children were found to evaluate the cognitive capabilities of the agent not based on its humanlikeness. When only adults' ratings are obtained, they seem to evaluate the cartoon-stylized humanoid agent less positively with regard to likability and trustworthiness, while the abstract object was perceived as less intelligent.

Further preferences of a specific age group were investigated in qualitative interviews with seniors, where contradicting findings were reported (Tsiourti et al., 2014). While one part of the sample stated they preferred a realistic humanoid agent as their health-care assistant, another part preferred an abstract object (Smiley), since they stated that a humanoid assistant would be to stiff and serious: "The humanlike avatars are too impersonal, too serious, maybe depressing after a while" (Tsiourti et al., 2014, p. 62). Although Tsiourti et al. (2014) did not experimentally investigate target-group-related differences, their results in combination with the results of Krumhuber et al. (2015) seem to indicate that children and

seniors evaluate objects more positively. However, as the results from Tsiourti et al. (2014) were also contradictory, no straightforward assumption can be made and further investigations are needed. Moreover, the results indicate that adults perceive humanoid agents as more competent (Krumhuber et al., 2015), serious (Tsiourti et al., 2014), and trust them more in the context of product recommendations (Terada et al., 2015). Thus, in applications where competence and seriousness are important, a humanoid agent seems to be more appropriate than an object.

3.2.3 Realism

There is a large body of research investigating the effect of an agent's degrees of realism. In this research area it is often discussed whether a realistic agent is favorable or not. Referring to the uncanny valley theory (Mori, MacDorman, & Kageki, 2012), it has to be assumed that increasing realism leads to increasing positive evaluation until a certain gap is reached where the effect is inverted and the appearances evoke negative feelings. However, although the uncanny valley is a very popular theory and is often assumed to be proven, there is a lack of systematical experimental evidences and also other cues like personality or behavior have to be taken into account (Rosenthal-von der Pütten & Weiss, 2015). Moreover, there are conceptual issues with the uncanny valley and its outcome variables, as neither affinity nor humanlikeness haven been determined conceptually (Rosenthal-Von Der Pütten, 2014). While there are evidences that there is a relation between affinity and humanlikeness, the exact form was not always proven to be the assumed valley, but also linear trends have been found (Rosenthal-Von Der Pütten, 2014). Therefore, this theory has to be kept in mind for the current dissertation, but is not in focus.

In reference to the uncanny valley, McDonnell et al. (2012) found highly realistic and highly abstract appearances to be perceived as most appealing, while the degrees of realism in between were evaluated as unappealing. This tendency was even more obvious when the agent showed motions. Moreover, the perceived unfamiliarity of the appearance let participants evaluate the presented appearance as more unappealing. Nevertheless, no differences in behavioral trust toward different degrees of realism were noted.

The uncanny valley was not only found for human appearances, but also for animals. Schwind et al. (2017) investigated the effect of realism for animal appearances. They found significant difference between different degrees of realism with regard to familiarity, realism, and aesthetics. The authors highlight that they found an uncanny valley for animal agents too

and that animals should either be designed naturally or with a cartoon-stylization to avoid the uncanny valley.

By contrast, with regard to the uncanny valley, Zell et al. (2015) found realism to be a bad predictor of eeriness and appeal. Zell and colleagues (2015) delved into the subcategories of realism deeper and examined the impact of stylized shape and material on the perception of a virtual agent. While the authors found shape to be the most critical factor for perceived realism, the stylization of material was found to affect the perception of eeriness, appeal, and attractiveness more.

Most studies present a-priori-designed agents to the participants for their evaluation, while Schwind et al. (2015) investigated how people will build appearances that are either liked or rejected. Their results demonstrated that with regard to cartoon stylization most people stuck to realistic features, since for more than 70% of the designed faces no stylization was used, only about 10% added no skin details, and when a stylization was used, only lower degrees were added to the design (Schwind et al., 2015). When a likable appearance is created, people rely on realism and naturalism. Thus, people seem not only to evaluate realistic appearances more positively, but they also design a likable agent with a realistic appearance.

The aforementioned evidence investigated the effect of appearance on the perception with regard to variables such as appeal, eeriness, or likability; yet it is also important to examine the effect of appearance on the user's feelings and behavior.

Based on pictures only, it was found that a realistic humanoid face is more intelligent, likable, and engaging compared with a cartoon-stylized human (Koda & Maes, 1996). These results are also transferable to a poker game between an agent and a user, where realistic agents were also perceived as more intelligent and likable. Moreover, users stated that it is more comfortable to play against a realistic agent than a cartoon-stylized agent (Koda & Maes, 1996). Hence, a realistic appearance can engage a player, since it is perceived as more intelligent and also likable.

With regard to the influence of realism on users' feelings, studies investigated how realism affects emotional contagion. Matsui and Yamada (2017) investigated which appearance dimensions evoke the highest emotional contagion by the users. Besides ethnic group and sex (Feature Specification described in Chapter 3.2.4), they compared different levels of abstraction. For this, a cartoon-stylized appearance (agents had cartoon-stylized shades and proportions) was compared with an abstract line drawing (which can be seen as highly stylized and with simplified details). Results indicated that agents with an abstract line-

drawing appearance evoked less emotional contagion (Matsui & Yamada, 2017). Emotional contagion was measured by asking people how happy they felt when they saw the agent. A simple line drawing evoked less joy than a cartoon-stylized appearance. In this sense, higher realism might lead to a more positive affect.

In line with this, the emotional contagion of a virtual patient's appearance was investigated by comparing a realistic appearance with a cartoon-stylized shade appearance and a line-drawing appearance (Volante et al., 2016). The authors found that the negative affect is lowest for a realistic appearance and differs from both of the other conditions. Thus, people felt less negative after interacting with a virtual patient who had a realistic appearance than having a stylized appearance or a line-drawn appearance. Moreover, a realistic appearance was perceived as more sentient, conscious, and alive than a cartoon-stylized appearance (Volante et al., 2016). Overall, people seem to feel better after an interaction with a realistic appearance than with a stylized appearance.

As described in Chapter 1.2.4, medical applications are a common usage of virtual agents and therefore many studies investigated the role of realism in a medical or health-related context.

Van Wissen et al. (2016) examined the appearance of a virtual health coach with regard to age, familiarity, and realism. They compared a photorealistic female coach with a stylized female appearance with one that had cartoon shades. Results of the study indicate that participants evaluated the realistic appearance more positively (friendlier, more competent, more attractive, liked it more, showed more trust, followed advice more, were more likely to continue interaction and were more satisfied; van Wissen et al., 2016). It should be mentioned that here photorealistic pictures of real humans were used as realistic stimuli, while most other studies compared computer-generated appearances only. In line with this, elderly participants were found to appreciate a realistic agent more as their health coach. By investigating the preferences of elderly users, Tsiouri et al. (Tsiourti et al., 2014) found that the elderly prefer a realistic looking agent, since they want to be able to look into its eyes.

Similar to the study goal of van Wissen et al. (2016), Robertson et al. (2015) investigated how a decision-making agent should look for health decisions. They gathered opinions of prostate patients regarding a photograph, a vectorized photo, a more realistic illustration (the authors describe this agent to be realistic, although it has fewer details and cartoon-stylized shades, but nevertheless it is more realistic than the cartoon-stylized agent), and a cartoon-stylized agent (with both stylized shades and proportions) in a qualitative approach. Results indicate that the patients mostly disliked the stylized versions of the agent

and would prefer to have a realistic appearance for their decision-making assistant. To some point, participants even stated anger about the unrealistic version ("Prostate cancer is not like a cartoon," or "This is a serious matter and having a cartoon seems to minimize it." (Robertson et al., 2015 p. 433)). On the other hand, the authors also received positive opinions about the cartoon stylization, since for some patients the cartoon version was seen to attract more attention and be less serious (Robertson et al., 2015). As these results mainly highlight that participants rejected a cartoon-stylized appearance in the context of serious illnesses, the context might be important for the evaluation of the agent's appearance.

To investigate the impact of an agent's application context on the evaluation of different degrees of realism, Ring and colleagues (2014) compared a social context with a medical application and investigated the evaluation of stylized shading and stylized proportion in these contexts. Results for the shading indicate that an appearance with stylized shade was perceived as more friendly than the appearance with realistic shades. Moreover, an interaction effect between shades and application occurred, where stylized shaded agents were perceived as more friendly than realistic shaded agents, but only in the social interaction. By contrast, the stylized shaded agent was perceived as more caring in the social interaction, while in the medical interaction the agent with realistic shades was evaluated as more caring (Ring et al., 2014). Similar results were found for stylized proportions, since an agent with cartoon-stylized proportions was perceived as more friendly than an agent with realistic proportions. Again, an interaction effect was found. Here an agent with stylized proportions seems to be more appropriate for a social interaction, while for medical interactions the agent with realistic proportions is more appreciate (Ring et al., 2014).

Contradictory findings were reported with regard to learning outcomes. It was found that students had higher performances when they worked in a learning environment with an agent that has a realistic appearance (Baylor & Kim, 2004). Moreover, Baylor and Kim (2004) found an interaction effect with participants' sex and the realism of the pedagogical agent. Men had higher outcomes with a realistic agent than with a cartoon-stylized agent, while no such effect was found for female students. By contrast, no such differences were found when a highly realistic, moderate realistic, and low realistic pedagogical agent was compared (Saidatul Maizura et al., 2010). Here the students' achievements did not differ between the different appearances and here, too, an effect of the participants' sex was found. While in these two studies human appearances were used, another study demonstrated the relevance of the interaction of realism and species during learning sessions. Li et al. (J. Li et al., 2015) found an interaction effect between species and realism in the perception of a

virtual lecturer. People who listened to a human lecturer were able to recall more of the lecture content when a real person held the lecture. By contrast, when a robot instructor was used, the recall was higher when a virtual robot gave the lecture. Thus, for humanoid appearances, high realism leads to higher learning outcomes, but when a machinelike agent is used, it should have lower realism.

To conclude, a realistic agent was found to evoke a more positive affect after the interaction (Matsui & Yamada, 2017; Volante et al., 2016). A realistic agent was also more accepted for medical tasks (Ring et al., 2014) and evoked more positive outcomes during this application (Robertson et al., 2015; Tsiourti et al., 2014; van Wissen et al., 2016). The positive effect of realistic appearances has also partly been found with regard to learning outcomes (Baylor & Kim, 2004); however, the evidence available is contradictory (Saidatul Maizura et al., 2010). However, when the realism of an agent is examined, the uncanny valley has to be kept in mind, since there is evidence that it is applicable to virtual agents (McDonnell et al., 2012) and not only humans but also animals (Schwind et al., 2017). Thus, virtual agents should either be kept realistic or a stronger stylization should be used, to avoid the uncanny valley.

3.2.4 Feature Specifications

With regard to feature specifications, mostly humanoid characteristics have been explored. The sex of the agent is thereby the most widely investigated feature and overall female agents were perceived more positively, while contradictory derivations were made for a learning environment.

By investigating the role of sex and corporate clothing of a product recommendation agent, Lunardo et al. (2016) found female agents to be perceived as more attractive than male agents. This effect was found to be even stronger when the agent wore corporate clothing. However, no difference between website familiarity and visit intention was found between the male and female agent was found. But the attractiveness rating leads to higher trust in the website; people trust the website more when the presented agent is perceived as attractive (Lunardo et al., 2016). Female agents were found to evoke more emotional contagion by the users compared with male agents (Matsui & Yamada, 2017). People felt happier after interaction with a female agent. In line with this, older users stated in qualitative interviews that they preferred a female agent to a male agent as their virtual assistant (Tsiourti et al., 2014).

In a learning environment, however, a male agent was found to be more credible, engaging, and humanlike than a female agent (Shiban et al., 2015). This perception also led to better performance after a learning session with the male agent (Shiban et al., 2015). By contrast, the female agent was found to evoke more interest in the learning topic than the male agent did (Shiban et al., 2015). However, the performances were still found to be better after interaction with the male agent. These results support the findings from Baylor and Kim (2004), who reported that a male agent increased self-efficacy and was perceived as more interesting, more intelligent, and more useful and students were more satisfied after an interaction with a male agent compared with a female agent.

With regard to sex, it is often assumed that a match of the user's sex with the agent's sex is the best option. However, the results of an experimental study did not support this assumption. No differences in perceived enjoyment and usefulness of a product recommendation agent were found between users where their sex matched with the sex of the agent and those where a mismatch was present (Qiu & Benbasat, 2010).

By contrast, the authors found that ethnic group had an effect on the perception of the agent. When people interacted with an agent that had the same ethnic group as themselves (Asian vs. Caucasian), the agent was perceived as more enjoyable and useful (Qiu & Benbasat, 2010). Moreover, an interaction between the ethnicity match and participants sex was found. For female users, the match of ethnic group had even stronger effects than for male users (Qiu & Benbasat, 2010). Results from Matsui and Yamada (2017) support the assumption that a match of the agent's ethnic group with that of the user is more effective, since they found Asian agents to have a stronger impact on emotional contagion: The user felt mostly happier after interacting with an Asian agent compared with a Caucasian agent, when Asian people participated in the experiment. However, unlike Qiu and Benbasat (2010), the authors did not experimentally vary the match of the ethnic group of the user and agent, but only Asian people evaluated the agent and results indicated stronger effects for Asian agents.

While Baylor and Kim (2004) did not take the match of ethnic group of the user and agent into account, they also found that the ethnic group of the agent (Black vs. White) had an effect on users' perception and outcomes. Black agents used in a learning environment were found to lead to higher self-regulation by the students than White agents did. Moreover, there was an interaction between the role of the agent and its ethnic group. A Black agent was found to be more effective as an expert and students were able to focus better on the material with a Black expert than with a White expert (Baylor & Kim, 2004).

Besides sex and ethnic group, the age of a virtual assistant was investigated. van Wissen and colleagues (2016) did not found a difference between older and younger agents. With regard to this finding, it should be noted that here again no match with the age of the participants was included and the age difference between both agents was rather small.

Overall, when a human agent is used, mainly its sex and ethnic group were found to affect the perception and interaction outcomes of a virtual agent. Results can be sorted along the warmth and competence stereotype content model of Fiske and coworkers (Fiske, Cuddy, Glick, & Xu, 2002). As Eckes (2002) highlighted, stereotypic females are associated more with warmth, while males are associated with competence. The findings summarized here agree with this assumption, since female agents were perceived more positively in social situations (where warmth is more relevant). Participants felt happier after an interaction with a female agent (Matsui & Yamada, 2017), they were evaluated as more attractive (Lunardo et al., 2016) and preferred as a personal assistant (Tsiourti et al., 2014). By contrast, regarding competence, male agents were found to be perceived as more intelligent and led to better outcomes in learning situations (Baylor & Kim, 2004; Shiban et al., 2015). To date, age has not been found to have an influence on the human–agent interaction.

Overall, only feature specifications of human agents have been researched and no assumptions about feature specifications of other species such as animals or robots can be made.

3.3 POSSIBLE EXPLANATIONS FOR APPEARANCE EFFECTS

There are plenty of explanations, why people respond in a certain way to a virtual agent's appearance. The uncanny valley has already been discussed in Chapter 3.2.3. Moreover, the appearance seems to trigger specific perception that matches the application context. Gulz and Haake (2006b) offer for example the assumption, that "visual naturalism is required to indicate that something is serious, whereas stylization signals leisure and especially entertainment". Therefore, some appearances might be preferred, because they are more serious or in contrast entertaining and this perception has to match the demands of the application context. This is in line with the findings by Ring et al. (2014) who demonstrated topic-related differences in the evaluation. Moreover, it is assumed that expectations that are triggered by a certain (especially very realistic and humanlike) appearance might not match the actual capability of a virtual human (Gulz & Haake, 2006b). For example, a highly realistic humanlike appearance might trigger the expectations to have the same intelligence as a real human and users are then disappointed when their expectations are not fulfilled.

Besides these assumptions, concepts found in human-human interaction are also applicable such as similarity and attractiveness.

In the following especially those explanatory assumptions are discussed that are specifically relevant for interpersonal relationships and therefore of importance for the present dissertation.

3.3.1 Attractiveness Stereotype

Physical attractiveness has been found to be very important in the construction of interpersonal relations (Walster et al., 1966). The authors found, that people are more likely to like an other person, more likely to have an intention to meet the other person again and to actually asks an other person for a date, when this person was physically attractive (Walster et al., 1966).

Referring to Dion et al. (1972) a "what-is-beautiful-is-good"-stereotype exist, in the sense that attractive people are also associated with other positive outcomes. The authors emphasized that attractive people are not only related to more positive personalities, but were also expected to have a happier and more successful life than unattractive persons. Hence, people judge other peoples' traits and characteristics based on their appearance and the attractiveness of it. This assumption is supported from multiple evidences with regard to different outcome variables over the past years.

Overall, attractive people are assumed to have more socially desirable personality traits (Brigham, 1980). A meta-analysis demonstrated that attractive people will be evaluated "as more sociable, dominant, sexually warm, mentally healthy, intelligent and sociably skilled than physically unattractive people" (Feingold, 1992 p. 304). The perception of attractiveness does not only trigger a more positive perception, but also more positive responses toward attractive people (Langlois et al., 2000). Depending on the attractiveness social judgments such as the perception of crimes those people might be guilty of or the tasks that people are useful for are made (Deska, Lloyd, & Hugenberg, 2018). Moreover, the attractiveness of people was found to influence the shape of their social networks (O'Connor & Gladstone, 2018). Thus, the perception of people and their behavior is also seen to be affected by people's attractiveness.

This stereotypy was not only found for real humans, but can with regard to the media equation (Reeves & Nass, 1996) also be transferred to virtual agents. Khan & DeAngeli (Khan & De Angeli, 2009) showed that attractive virtual agents are evaluated as more socially competent, more integer and more intellectual competent than unattractive agents.

Additionally, attractive agents were also found have stronger persuasive effects than unattractive agents (Khan & Sutcliffe, 2014). It was shown that same principles to assess the agent's attractiveness are used than for human attractiveness ratings, since people perceive overall similar physiognomic facial features as attractive for virtual and real humans (Sobieraj & Krämer, 2014).

To conclude, prior research demonstrated that perceived attractiveness affects the perception of interlocutors and our behavior towards them. Especially, in the construction of personal relations this is important, as it was shown that the liking and intention to see the person again was also affected by attractiveness. These, findings are not restricted to real humans, but are also applicable for virtual agents. Thus, it can be assumed, that the perceived attractiveness of a virtual agent's appearance has big influences on the overall evaluation and interpersonal relationship between the user and the agent.

3.3.2 Similarity

Another factor that affects the development of interpersonal relationship and perception of other people is the perceived similarity (see McPherson, Smith-lovin, & Cook, 2001 for an overview). Stronks and colleagues summarized the influence of similarity in a simple manner: "to put it simple, we don't dislike people that are just like us" (Stronks, Nijholt, van der Vet, & Heylen, 2002 p. 94).

Similarity was found to influence the attraction especially in early relationships formations (Byrne, 1997; Montoya et al., 2008) and people likes other people more, when they are perceived as more similar (Hampton, Fisher Boyd, & Sprecher, 2018). Thus, the influence of similarity with regard to appearance is seen to be mainly important in the beginning: "Superficial similarity (e.g. demographic characteristics) may be used as a criterion for selecting (or rejecting) friends, but as a friendship becomes close, the partners must be similar in deeper and more meaningful ways" (Fehr, 2000 p. 74). Therefore, an appearance that evokes feelings of similarity might be helpful to foster attraction, liking and first intentions to interact with the person, which can than lead to an interpersonal relationships and feelings of closeness.

The effect of similarity was also found for a virtual agents appearance. Agents that look like the user and thus have facial similarity evoke higher involvement by the user (when the agent was helpful) (van Vugt, Bailenson, Hoorn, & Konijn, 2010). Agents with similar body types have also been found to be more effective with regard to health outcomes (van Vugt, Konijn, Hoorn, & Veldhuis, 2006). Moreover, agents with a similar appearance can

enhance the self-efficacy of female students in learning application (Rosenberg-Kima, Baylor, Plant, & Doerr, 2008). There are also evidences that similarity with regard to ethnic group affecting the perception of a virtual agent (Qiu & Benbasat, 2010).

A virtual agent's appearance can evoke the perception of similarity via multiple cues. The most obvious cues are feature specifications such as same hair color, ethnic group, age or sex. Nevertheless, also categories like species or realism might affect the perception of similarity, especially in regard to perceived humanlikeness. Appearances that are closer to a real human (such as a human appearance and higher realism without stylization) might be perceived as more similar to the user. As similarity is expected to enhance the liking of an interlocutor and the formation of interpersonal relationships, it can also be assumed that an appearance, which is perceived as more similar, can produce these effects.

3.3.3 Customization

While in human-human interactions there is only an option to chose our interaction partners based on certain characteristics (for example as aforementioned based on attractiveness and similarity), but not to design our interaction partners with regard to our own preferences, this option is possible with regard to virtual agents. Since virtual agents are fictional characters design by humans, the design (in this case the appearance) can be adopted to the users needs and preferences. It might be the most efficient way to adjust the virtual agent's appearance to the users' needs and preferences, when the users have the opportunity to customize it. Designing or assembling something on your own can enhance the value of the customized item (e.g. product (Norton, Mochon & Ariely, 2012) or robot (Groom, Takayama, Ochi & Nass, 2009)). Accordingly, customizing the agent might enhance the willingness to use it and interact with it. In line with this assumption, Hurst and Tobias (2011) state that designing and creating their own assistive technology can enhance the users' adoption rate. Although it seems to be beneficial, there is a lack of research about customizing agent's appearance, while the customization of avatars (which are virtual representations of users) has been highly investigated. Since many games offer the users the opportunity to customize their gaming avatar, a lot of research is done regarding gaming. Customization enhances gamers' satisfaction and therefore their loyalty (Teng, 2010), players' identification with their avatars (Turkay & Kinzer, 2014) and is beneficial for engagement and presence (Ng & Lindgren, 2013). However, it has not been investigated whether those effects are transferable to the interaction with virtual agents. While the customization of agents might be beneficial in the way to tailor the appearance towards the users' preferences, in order to enhance identification,

immersion and perceived value, even the design process itself can have a positive impact. Marathe and Sundar (2011) stress that customization can enhance users' sense of identity and sense of control and therefore has a profound psychological value. Thus, customizing the own virtual assistant in regard to their preferences as well as the design process itself might be very beneficial for the human-agent interaction. Nevertheless, not a whole design process of the virtual agent might be required to evoke these positive effects. Diehl et al. (2017) demonstrated that an increasing degree of choice options lead to higher satisfaction with the agent, but that this is not true for a self-customization. Therefore, not only a whole customization process is required, but a choice out of multiple options can also have the promised effect described above.

3.4 SHORTCOMINGS OF PRIOR APPEARANCE RESEARCH AND OPEN RESEARCH QUESTIONS

The findings presented in the Chapters above, highlight the relevance of a virtual agent's appearance. Nevertheless, those studies have some shortcomings that have to be discussed and are used as a starting point for the current work.

Main problem of prior appearance research is the unsystematic empirical approaches that were used. As virtual agent's appearance has multiple facets, the controlled investigation and manipulation is difficult. While many studies aimed to investigate a certain factor of appearance, they did not control for other variables in which the used appearances differed. Sträfling et al. (2010) for example mixed up different species with different degrees of realism, since they compared a cartoon-stylized animal (stylized proportions and shade) to an unstylized human. As realism is the most complex appearance variable, many studies lack of the described problem. However, not only realism interferes with other appearance variables in former studies. Shiban et al. (2015) compared a male with a female agent. While the realism degree of both agents can be seen as equal, both agents differ not only in sex, but also in perceived age. When the appearance differs in multiple variables, it is hard to derive whether the variable of interest caused the studies findings or if the results are produced as a by-product of the confounding variables. In the aforementioned cases it is unclear whether species or realism cause the presented effects on engagement (Sträfling et al., 2010) and if sex is decisive for the effects with regard to learning outcomes or if those findings are produced by the perceived age (Shiban et al., 2015). Thus, when the appearances differ in multiple variables that are not examine in a systematical and controlled manner, it is hard to make derivations from those studies. The described categorization (see Chapter 3.1) was

constructed to overcome this problem, as one can design and describe the appearance of a virtual agent along this categorization. This will help to guarantee a similar level for each appearance variable that is not in focus of the research agenda and applies as an overview over possible variables that might interfere with the investigation.

In prior studies not only the problem of different variables that are confounding the actual investigated appearance variable occurred, but also the design of experimentally stimulus material is challenging in multiple other ways.

In some studies the quality of the used stimulus material is to some degrees improvable. Van Wissen et al. (2016) for example investigated the effect of cartoon stylization in a health application. While they did it in a rather controlled way, where the same stylization was applied to all realistic appearances, the resulting material does not match the current standards of cartoon stylization known from media. Thus, the poor done stylization might confound the achieved results.

Moreover, in some cases the appearance variables, which were in focus of the research, were not clearly manipulated. See for example Matsui and Yamada (2017) who wanted to manipulate ethnic group by comparing a Caucasian with an Asian agent. From the used stimulus material it is not clearly derivable that ethnic group was manipulated, but the appearances rather differed in realism. While the Asian agents were highly stylized in shade and had low details, the Caucasian agents might score higher in realism, since a more realistic material was used. Furthermore, the Caucasian female agent was designed with unrealistic feature specifications such as blue hair, while all other agents had natural feature specifications. In sum, again the appearances differed in many other variables than the intended and in addition, the pivotal variable (ethnic group) was not obtainable. Unfortunately, the authors did not check whether the manipulation was successful in the sense whether people perceive the agent's ethnic group in the intended way. Therefore, no clear implications from the findings can be made.

Due to the manifold appearance variables, it is hard to design material in a maximum controlled way. Often not only the manifold variables, but also technical restrictions can limit the used stimulus material. Therefore, it is of great importance to use pre-tests and manipulation checks. Using these methodological techniques, it can be clarified right from the beginning, in which variables the used appearances differ and whether the intended manipulation was successful or not. When other variables than realism are of research interest (such as species or feature specification variables like sex), than at least the perceived degree of realism has to be controlled as this was found to be a crucial variable with multiple

sub-categories. Although it is still hard to design good stimulus material, due to the enhancing technical developments, it seems to become easier to design controlled material.

Besides the problems during the design and selection of systematically varying stimulus material, the settings, in which the effects of appearance have been tested, have to be marked. So far nearly none of the mentioned studies above used a reciprocal interaction between the user and the agent that is speech-based.

Many studies just tested the perception and evaluation of appearance variables by presenting static pictures, which have been evaluated by the user with out any context (e.g. Krumhuber et al., 2015; Matsui & Yamada, 2017; Schwind et al., 2017; Tsiourti et al., 2014; Zell et al., 2015). This method is beneficial, since so the mere effect of the appearance can be derived, as no interaction variables (such as the topic or the agent's responses) can interfere with the appearance effects. Nevertheless, research already showed that the behavior can effect the perception of appearance (e.g. Bailenson et al., 2005). Thus, it is hard to derive implications for a human-agent interaction, when only static pictures were tested. As a result, the effects found by the evaluation of mere static pictures should always be investigated and replicated in a human-agent interaction. Moreover, the outcome variables that can be tested when static pictures with out any interaction are used are limited, since only perception variables (such as person perception) and no behavioral outcomes (such as learning outcomes) can be measured.

When interactions between users and the agent were used, the user interacted mostly not via speech, but either via multiple-choice buttons or via keyboard. In some studies, pictures of a virtual agent were used and combined with a chat-bot where the user and the agent interacted via text through the computer (e.g. Baylor & Kim, 2004; Koda & Maes, 1996; Lunardo et al., 2016; van Wissen et al., 2016). Other studies did already use a moving agent that communicated via speech, but still the user was not able to respond via speech (e.g. Ring et al., 2014; Volante et al., 2016). Thus, appearance effects have so far not been tested in a reciprocal speech-based interaction, where the user is able to interact and respond via speech to the agent. However, the speech-based interaction and the resultant natural and humanlike communication with the agent were seen as the most promising ability of virtual agents. Hence, speech-based interaction studies that examine the effect of an agent's appearance are missed and this lack is tackled in the current dissertation.

Additionally, there are some evidences that user characteristics can also effect the perception of appearance variables. For some appearance variables (mostly feature specification variables) male and female user were found to evaluate the appearance

differently (Qiu & Benbasat, 2010; Saidatul Maizura et al., 2010; Zibrek & McDonnell, 2014). Moreover, age was found to be a critical user characteristic that affects the perception of the agent in regard to its appearance. Krumhuber et al. (2015) found children to evaluate the presented appearances different than adults have done. In addition, specific preferences of elderly got accessed (Tsiourti et al., 2014). However, no experimental studies tested age-related differences in appearance preferences and the perception of appearance variables with regard to elderly users. This is important, since seniors are one main target-group especially for health and assistive applications of virtual agents, since this applications are most promising for seniors and people in need of support. Furthermore, the users attitude and personality have also so far not been taken into account.

In Chapter 1.2.6 it has been derived, that many application fields aim for a long-term interaction between the user and the agent. However, so far (to the authors knowledge) no long-term studies exist, in which the effect of appearance has been investigated. Bergmann et al. (2012) presented first evidences, that the impression of an agent can change over time and that this is connected to the agent's appearance. This is somehow in contrast to findings of the human-human interaction, where it is assumed that the first impression lasts and will even be deepened (Mathes, 1975). In addition, the study of Bergmann et al. (2012) is only a first hint, but since in this study still only a single interaction between the agent and the user was used, it has to be investigated how the effect of appearance develop over a longer period and within multiple interactions.

In sum, there are some issues with systematical and controlled investigations of appearance variables. It has to be marked, that the aforementioned problems does not apply to all studies that investigate effects of appearance. But due to the technical restrictions, those shortcomings are especially found in studies where behavioral outcomes and interactions between the agent and the user have been tested. Talking about interactions between humans and agents, so far no appearance study exist where the user and the agent interacted reciprocally speech-based. Additionally, there is a lack of examinations of the moderating effect of user characteristics such as age. And most important with regard to the application field of virtual assistants, no long-term study on appearance effects has been conducted. Thus, no conclusions about the development of the found appearance effects over time are possible.

4 SYNOPSIS OF THE THEORETICAL BACKGROUND AND GENERAL RESEARCH MODEL

To present the overall motivation of the present dissertation, first the general research questions are derived from the aforesaid theoretical and empirical background. The overall goal of this dissertation is to investigate whether peoples' intention to use or actual usage of a virtual agent can be enhanced with the agent's appearance and its effects in terms of social relations. Therefore, as a second step a theoretical framework that summarizes the assumed influences of a virtual agent's appearance with regard to social outcomes and usage behavior on prior research is proposed.

4.1 SUMMARY OF THE GENERAL RESEARCH QUESTIONS

The previous Chapters outlined the role of a virtual agent's appearance in a human-agent-interaction. At the same time, some methodological problems with prior appearance research have been extracted. While the body of research highlights the importance of an agent's appearance (e.g. with regard to entertainment (Koda & Maes, 1996), buying intention (Terada et al., 2015) and the overall evaluation of the agent (McDonnell et al., 2012; Ring et al., 2014; van Wissen et al., 2016)), the research was often done in a rather unsystematically way. Studies compared difference appearances, which varied in multiple variables and not only in the pivotal ones (e.g. while realism was aimed to be investigated different species have been used or sex was the pivotal variable and the agents differed also in perceived age) or the focused stimulus variation has not been designed and tested proper (e.g. while ethnic group was the pivotal variation, feature specifications lead to a different degree of realism). In contrast to these former unsystematic approaches, the perception and evaluation of a virtual agent's appearance have to be investigated in a more systematic manner. It is important to assess the overall perception and evaluation of appearance variables first, before the effects of this variables in an actual human-agent interaction and over a longer period are tested. Thus, the following general research questions are posed:

General Research Question 1 (GRQ 1): Which appearance variables are of pivotal interest for the evaluation of virtual agents in an assistive application?

General Research Question 2 (GRQ2): How should a virtual assistant's appearance be designed?

While virtual agents can be used by all kinds of users, people in need of support were highlighted as target group, for which the application of virtual agents will be especially promising. At the same time this target group was found to be more skeptical about the use of virtual entities (Scopelliti et al., 2005), therefore virtual agents have to be designed in a way that triggers usage intention. A first beneficial way with regard to findings from human-human interactions might be a favorable design of the appearance along the preferences of the users. While there are first hints that user characteristics influence the evaluation of appearance variables (Krumhuber et al., 2015; Qiu & Benbasat, 2010; Tsiourti et al., 2014), so far no experimental research has been conducted to investigate those target-group related differences. Accordingly, the following general research questions will be examined in the present dissertation:

General Research Question 3 (GRQ3): How do user characteristics and target-groups influence the evaluation and perception of a virtual agent's appearance?

General Research Question 4 (GRQ4): What is the most effective appearance for people in need of support?

The present work aims to investigate whether the agent's appearance can evoke a certain perception and evaluation that triggers social outcomes like trust or bonding. As most applications, for which virtual agent are designed, desire multiple and steady interactions (for an overview see Chapter 1.2), it is specifically important to investigate these processes in a long-term interaction. While prior long-term studies highlight that users build social relations such as bonding towards a virtual agent (Bickmore, Caruso, Clough-Gorr, et al., 2005; Bickmore, Gruber, & Picard, 2005), no study investigated the role of an agent's appearance over a longer period of interactions. But from human-human interaction it can be derived that appearance mainly with regard to attractiveness (Dion et al., 1972; Poulsen et al., 2013; Walster et al., 1966) and similarity (Byrne, 1997; Montoya et al., 2008; Urber et al., 1998) is a critical factor in the formation of interpersonal relationships. Additionally, it is important to investigate the development of these effects, as first hints exist that people can revise their impression of a virtual agent (Bergmann et al., 2012).

Nevertheless, this has not been investigated over multiple interactions. Consequently, the following general research questions are posed:

General Research Question 5 (GRQ5): Which effect has a virtual agent's appearance on the interpersonal relationship in a human-agent interaction?

General Research Question 6 (GRQ6): How do the effects of a virtual agent's appearance develop over time in a long-term human-agent interaction?

4.2 DERIVATION OF THE ASSUMED UNDERLYING PROCESSES

As explained above, especially for specific target-group an agent's design that fosters the user's usage intention is of importance. When an agent attracts the users to interact with it through the interpersonal relationship between human and agents, a steady usage can be guaranteed, which enhances the effectiveness of the agent. Bickmore (2005) highlights the benefits of an interpersonal relationship for several tasks such as persuasion, education, business and helping. The agents appearance can be seen as a pivotal variable in the formation of these interpersonal relationships (see findings regarding attractiveness (Dion et al., 1972; Poulsen et al., 2013; Walster et al., 1966) and similarity (Byrne, 1997; Montoya et al., 2008; Urber et al., 1998)) especially in the beginning.

Based on the media equation theory (Reeves & Nass, 1996), it can be assumed, that these social processes are also applicable to human-agent-interactions (e.g. attractiveness stereotypes are also applied towards virtual agents (Khan & De Angeli, 2009; Sobieraj & Krämer, 2014)). Thus, the following social processes were assumed for the interaction between humans and agents with regard to appearance effects.

The agent's appearance does influence the way in which humans will perceive the agent. Referring to the attractiveness-stereotype (Dion et al., 1972; Khan & De Angeli, 2009) inferences about the agent's personality and abilities are made based on its appearance. In line with this, there is a body of research presenting such evidences: Appearance was shown to affect the person perception (e.g. how likable or competent an agent will be perceived (e.g. McDonnell et al., 2012; Ring et al., 2014; van Wissen et al., 2016)) and the overall liking or attraction of the agent (e.g. J. Li et al., 2015; Zibrek & McDonnell, 2014).

In addition as an evaluative response of this perception, users' liking of the agent and their trust towards the agent are influenced by the agent's appearance (Stirrat & Perrett, 2010). Thus, users evaluate based on the agent's appearance and the therefrom-resulting person perception whether they likes the agent or they are attracted by it and if they show trust towards the agent (Stirrat & Perrett, 2010). Trust and attraction were both found to affect the interpersonal relationship in terms of bonding and closeness. More over people's

interpersonal relationship towards another person is highly correlated with spending time together (McAdams & Constantian, 1983). Hence, when people feel closer or more bonded to the agent, they will more likely show a higher intention to interact with the agent and their actual usage behavior. This relation might help to ensure a steady usage of the agent and to raise its effectiveness.

The prior described relations are social outcomes which are according to the media equation theory (Reeves & Nass, 1996) evoked by the social cues of virtual agents. Besides these social processes, virtual agents are still technologies. When the agent is more seen as a technical device, technology acceptance models can be consulted to explain the usage intention. One of the most frequent used models is TAM (Technology Acceptance Model (Davis, Bagozzi, & Warshaw, 1989)). This model describes the perceived usefulness and ease of use as two main components to explain user's intention to use a certain technology. This model has also been adapted to the acceptance of assistive social agents (Heerink, Kröse, Evers, & Wielinga, 2010). The authors included more components like trust and perceived enjoyment in the model. Accordingly, the perceived usefulness, ease of use (Davis et al., 1989) and perceived enjoyment (Heerink et al., 2010) can be assumed to influence participants usage intention.

From the mere-exposure theory (Zajonc, 1968) it can be derived that a higher exposure rate or in this context a higher number of interactions, does also enhance the perception and evaluation of the agent. As mentioned above, the interpersonal relationship is correlated with time spend together (McAdams & Constantian, 1983). Thus, the interpersonal relationship is not only assumed to influence the intention to interact, but the actual interaction frequency can also affect the interpersonal relationship positively. Nevertheless, it has to be mentioned that repetitiveness is a big influence factor in the context of human-agent-interaction (Bickmore & Schulman, 2009) that can decrease the interaction quality. When people got bored or perceive the boundaries of the agent's interaction capacities due to a repetitive communication structure, they evaluate the agent and the interaction therewith less positive. Moreover, with regard to technical devices a novelty effect is often proposed (Wells, Campbell, Valacich, & Featherman, 2010). Referring to this effect, especially technologies with a lower degree of innovation are assumed to attract people in the beginning, but than the evaluation will become more negative. This is in conflict with the aforementioned mere-exposure effect (Zajonc, 1968) and the assumption, that more interactions are related to a stronger interpersonal relationship. Therefore, the influence of the number of interactions on the assumed processes needs to be investigated thoroughly.

Since prior research indicates that user characteristics might affect the effects of an agent's appearance (e.g. Qiu & Benbasat, 2010; Zibrek & McDonnell, 2014) and for virtual assistance a very specific target-group has been extracted, those characteristics, preferences and need have to be investigated. So far there is a lack of experimental research that covers these effects. Hence, user characteristics are assumed to influence the described processes and the specific effects have to be examined in the present research.

Figure 3 presents the assumed theoretical framework. Within the framework two steps of the effect of appearance variables are assumed: first the mere perception of appearance variables can effect the impression of the agent and subsequent the appearance can be evaluated in combination with the interaction between the user and the agent. The later step is especially important with regard to long-term effects. In the presented framework only the relationships between the variables that are in focus of the present research are pictured. There are more relationships that are not specifically investigated in the empirical framework and those are therefore excluded to keep the theoretical framework simple. The same is true for the effects of user characteristics and number of interactions. Since so far no concrete assumptions based on prior evidence are possible, their influence is displayed on the overall process and has to be investigated with this work's empirical framework.

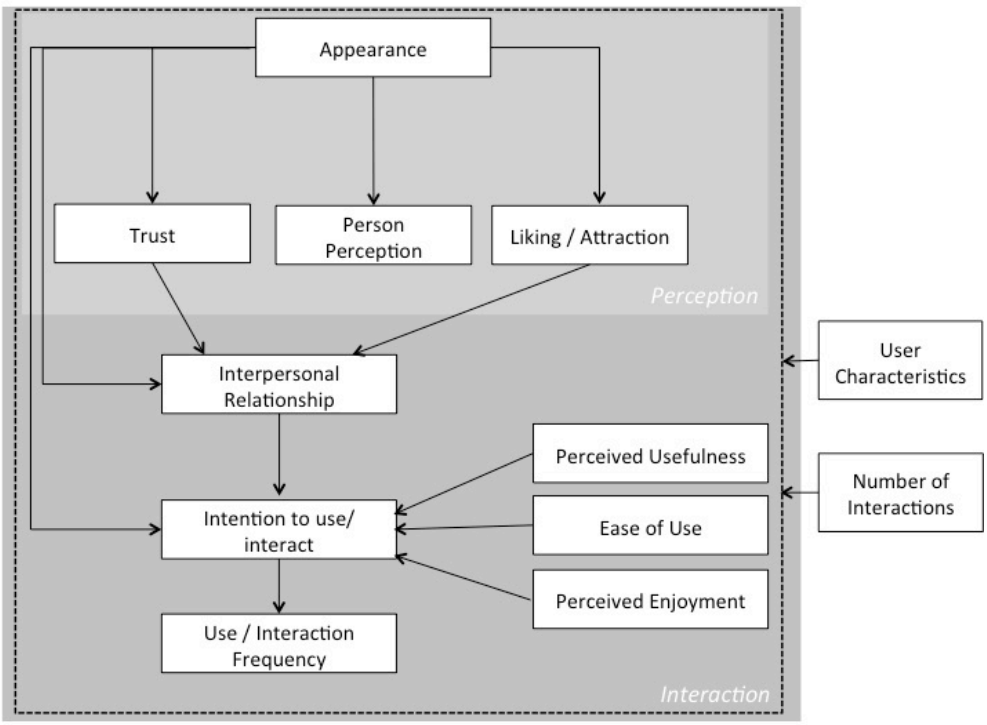


Figure 3. Theoretical Framework of appearance effects in terms of social relations and the enhancement of usage behavior.

III EMPIRICAL FRAMEWORK

This Section covers the empirical work that was used to answer the general research questions. Therefore, first the overall framework and its structure were described. Subsequently, all four studies are explained in detail.

5 INTRODUCTION TO THE EMPIRICAL FRAMEWORK

To gain a holistic view over humans' perception and evaluation of a virtual agent's appearance, a mixed method approach was used and qualitative as well as quantitative methods have been applied. Furthermore, the present work presents perception and interaction studies. To receive insights in the mere effects of different appearance variables without any confounding effects of the interaction with the agent (e.g. produced through the agents' speech and nonverbal behavior or by the interaction topic), studies based on pictures only (perception studies) were conducted. However, as the effect of these appearance variables in a social interaction and long-term interactions between humans and agents are of special interest, obviously interaction studies need to be done to answer the general research questions. Avoiding an unsystematically research agenda (as further explained in II3.4), all studies were designed based on the constructed categorization of appearance variables. *Figure 4* presents the consecutive study order and its applied methodical approaches.

The first study intended to gain deeper insight in age-related preferences regarding a virtual agents appearance. As qualitative research is recommended to be applied for research questions that aim to understand social phenomena and explore issues in more depth (Ritchie, Jane & Lewis, 2003), a qualitative interview studies was used. Due to this method the constructed categorization could be explored for the first time. As the categorization contains multiple variables and endless combinations are possible, not all variables can be investigated in experimental studies. Thus, Study 1 further extracted categories of most interest. Up until now (to the author's knowledge), nothing is known about senior's preferences regarding a virtual agents appearance in the context of virtual assistance. Accordingly, Study 1 furthermore was used to understand age-related differences in appearance preferences.

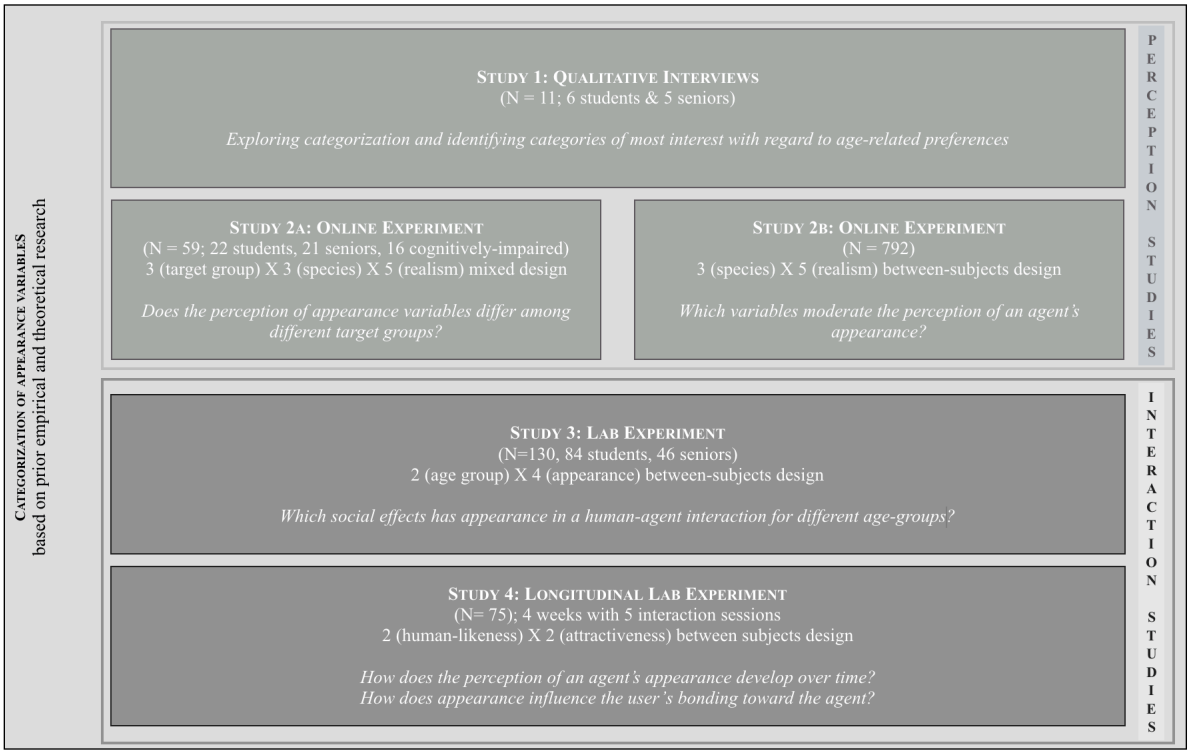


Figure 4. Empirical Framework with all conducted studies of this dissertation

Based on the results of Study 1 those categories with users' most interest have been selected and were investigated more intense in Study 2. In this two-parted study, appearance variables have been varied and manipulated thoroughly along the systematic categorization. Goal of Study 2 was to quantify the results of Study 1 and to gain in-depth insights in the effects of specific appearance variables. Moreover, differences between target-groups and effects of other user characteristics (e.g. attitudes and personality) were taken into account, to understand the mechanism of users' perception and evaluation of an agents' appearance holistically. As mentioned before, those perception studies were used to exclude confounding variables and observe the mere effect of appearance.

To complement the first two studies with contribution regarding the effects of appearance variables in an actual human-agent interaction, interaction studies were needed. Study 3 looks further into the differences between age groups and their perception and evaluation of appearance variables. Participants therefore interacted with an agent whose appearance was varied. As in described before, in an interaction between human and virtual entities, social cues are of great importance and social outcomes are highly likely. Thus, Study 3 further investigated the influence of an agent's appearance on social variables like trust or bonding. Those outcomes might be beneficial to enhance usage behavior and the interaction overall. In sum, Study 3 was used to gather comprehension about social

mechanisms in a human-agent interaction and was consequently a precondition of Study 4, where long-term effects have been addressed.

Study 4 is the core study of this work and investigated the influence of appearance variables in a long-term interaction. In a 4-weeks lab experiment the effects of appearance and interaction time on agent as well as interaction evaluation and user's social reactions towards the agent have been investigated.

All studies are consecutive explained extensively in the following Chapters.

6 STUDY 1: QUALITATIVE EXAMINATION OF AGE-RELATED USER PREFERENCES WITH REGARD TO A VIRTUAL AGENTS APPEARANCE

As described in Chapter II3.4, prior research investigated the effects of a virtual agent's appearance in a rather unsystematic manner. Therefore, in this dissertation a categorization of appearance variables was constructed (see Chapter II3.1). Based on this categorization, the preferences of users with regard to appearance can be explored. While it is impossible to investigate all appearance categories in a controlled experimental approach, the categories of highest interest have to be extracted.

Moreover, in order to create virtual agents that are motivating the user to interact with it, it is necessary to take the users' needs and preferences into account, thus it is important to know the preferences of the agent's target group. Since current developments aim to maintain the elderly's autonomy by using virtual agents (e.g. Yaghoubzadeh, Kramer, Pitsch & Kopp, 2013), it is of particular interest to explore the preferences of this target group. Until now, little is known about age-related differences with regard to appearance preferences (compare the detailed derivation in Chapter II3.4), but since assistive technologies like virtual agents can be extremely beneficial for elderly people, knowledge about their preferences is called for.

In sum, the current approach has two main contributions: (1) to explore the constructed categorization of different appearance variables and find the categories of highest interest, and (2) to explore age-related differences in users' preferences regarding those categories. Therefore, a qualitative interview study was constructed in the context of virtual assistance with six students and five elderly people. Interviewees stated their preferences and feelings regarding the different appearance categories. Furthermore, they had the chance to design their own virtual agent to explore the effect of customization and their preferences more explicitly.

6.1 BACKGROUND & HYPOTHESES

The current approach aims to give a more holistic and systematic overview of a virtual agent's appearance. Regarding the design of a virtual agent, first, it must be decided whether it should be embodied or not. Although, users are familiar with disembodied agents (e.g. Siri or Cortana), embodied virtual agents were found to enhance social interactions in a positive way (Yee et al., 2007) and to affect for instance game engagement (Koda & Maes, 1996) as

well as learning outcomes (Baylor and Ruy, 2003). Therefore, the following hypothesis is assumed:

Hypothesis 1 (H1): Users will prefer an embodied character as their personal assistant over a speech-based system only.

As explained in Chapter X, the appearance of an embodied agent influences the agent's overall evaluation and the interaction therewith in multiple ways. However, visual appearance contains many different factors (e.g. perceived realism or species of the character) which may be manipulated and, in turn, affect the outcome. The categorization is needed to explore the effect of the mentioned factors more systematically, e.g. in experimental studies. Since it is not possible to investigate every variable in quantitative studies, the current approach explores the categorization for the first time and can give initial hints to (a) the users' preferences regarding the categories and (b) what categories seem to be more important and should, therefore, be investigated more. Consequently, the following research questions are investigated:

Research Question 1 (RQ1): What are the users' preferences with regard to the main categories of appearance variables?

Research Question 2 (RQ2): Which categories of appearance variables are of specific interest for the design of a virtual assistant?

A growing amount of research focuses on the implementation of virtual agents in assistive technologies for elderly people, aiming to maintain their autonomy. For instance, Vardoulakis, Ring, Barry, Sidner and Bickmore (2012) demonstrated that elderly people do accept a conversational virtual agent in their homes. In a long-term study, Bickmore, Carisp, Clough-Gorr and Heeren (2005) investigated the usage of a virtual assistant in order to improve the physical activity of elderly people. Participants stated high likability and acceptance toward the agent and they were more physically active when they interacted with the virtual agent. Thus, virtual embodied agents seem to be beneficial for elderly people with regard to mental and physical health and are generally accepted by this target group as a virtual assistance in their daily lives (Yaghoubzadeh et al., 2013). What remains open is the question how these virtual agents should look like and what kind of embodied design the special target group prefers. There are some insights in regard to e-commerce. The results of focus groups with a total of 25 elderly people indicated that abstract agents were preferred

because they are less distracting than humans or humanoid agents (Chattaraman et al., 2011). These results relate, however, to the application of virtual agents within e-commerce. It is still unknown how a virtual assistant for elderly people in the context of daily life support should look like and whether there are any age-related differences regarding its preferred visual appearance. Therefore the qualitative interview study addresses the following research question:

Research Question 3 (RQ3): Does age influence the users' preferences concerning the agents' appearance in the context of daily life support?

It has been explained that people might like things more when they create it themselves (see Chapter II3.3.3). Since also for a virtual agent's appearance customization might be the best way to adapt the users' preferences, this opportunity should be investigated in more detail. Customizing is seen to enhance the adoption rate of technologies (Hurst & Tobias, 2011) and the perceived value of the customized item (Norton et al., 2012). Within games, customization increases the loyalty of the players (Teng, 2010), identification processes (Turkay & Kinzer, 2014) as well as engagement and presence during the game (Ng & Lindgren, 2013). Since the process of customizing is psychological very beneficial (Marthe & Sundar, 2011), the following hypothesis can be derived:

Hypothesis 2 (H2): Users of a personal virtual assistant prefer to customize the appearance of the own virtual agent.

Nevertheless, again age-related differences with regard to customization have to be investigated. As seniors are often not used to technologies, they might not be capable and willing to create their own agent, although it might be beneficial to do so. Thus, the following research question is posed:

Research Question 4 (RQ4): Are there age-related differences with regard to the perception of the customization's benefits?

6.2 METHOD

In order to explore the meaning of virtual agents' appearance in more depth, qualitative interviews with eleven interviewees were conducted. This method captures the thoughts and attitudes of individuals more deeply and is less strict than standardized questionnaires. Therefore, interviewees' inner thoughts about a virtual agents appearance can be explored. In these semi-structured interviews the conducted categorization of the appearance variables is explored for the first time. Therefore, the interviewees' preferences of different appearance variables can be derived from the conducted interviews. Since quantitative methods cannot concentrate on all variable categories, it is essential to investigate the importance of the different categories beforehand. Furthermore, this interview study focuses on the overall importance of the agent's appearance and the effects of customization of its appearance. Another main benefit is the examination of age-related differences with regard to the aforementioned investigations of appearance variables.

6.2.1 Procedure

The interviews were framed in the context of virtual assistance and interviewees were told to imagine that they had a virtual agent as personal assistant in their daily lives. In order to build the interviews on a shared understanding of the term virtual agent, the interviewer gave a definition to the interviewees and two example videos were presented to give an overview over the capabilities of these entities. When a shared comprehension of the concept of virtual agent was received, the main interview began (see below for more details). During the interviews, different stimuli (pictures of different virtual agents) were presented in order to clarify the discussed categories (e.g. species or realism). As a second step, interviewees had the opportunity to design their personal virtual assistant with an online tool by Autodesk (Character Generator, see Figure 2). The possibilities of this tool were limited so that only humanoid characters with a similar degree of realism (same shade, texture etc.) could be designed. However, various other features could be manipulated (e.g. skin tone, hairstyle, body and facial features, clothing).

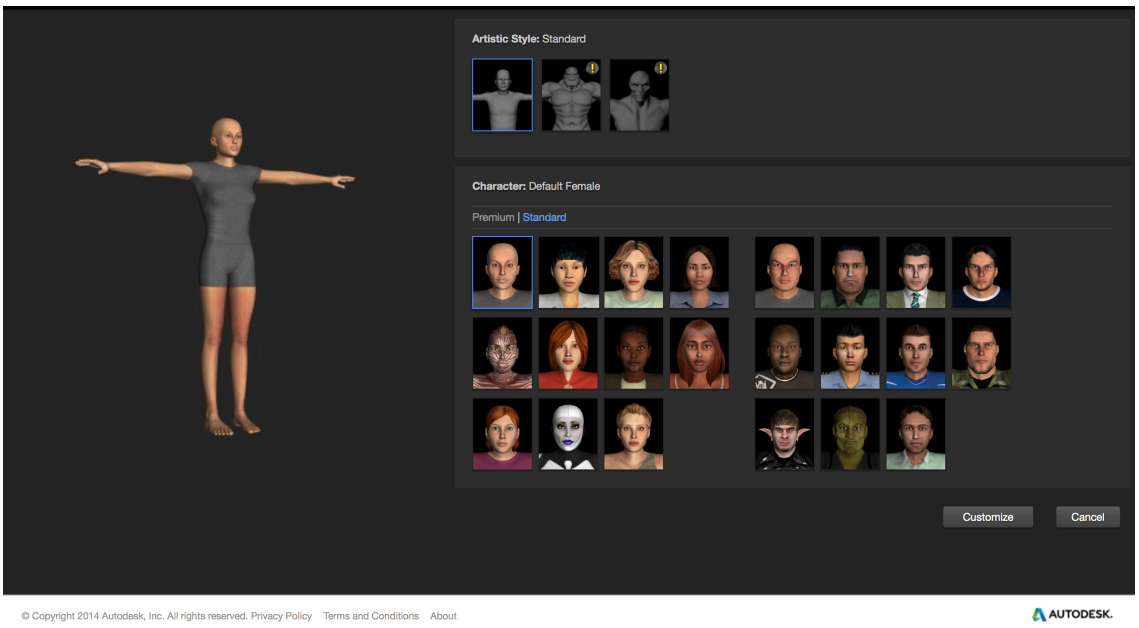


Figure 5. Screenshot of the used online tool (Autodesk Character Builder) for designing interviewees' virtual assistant

When the interviewees had designed their virtual assistant, they were asked to evaluate this agent. Afterwards, further interview questions were asked, and the socio-demographics of the interviewees were obtained. Each interview took about one hour. The interviewees were captured by a video camcorder and by the laptop's webcam. Also, the screen of the laptop was recorded to capture the design process of the interviewees' personal virtual agent.

6.2.2 Interview Guidelines and Analyses

Based on the categorization of the agent's appearance variables (as depicted above) a guide for the semi-structured interviews was constructed. The interviewer followed the manual consequently (the exact interview questions can be obtained from Table 1), but repeated the questions if they were not understood correctly. The interviewees' answers were not restricted nor did time constraints exist. During the whole interview, the interviewer was careful not to induce any answers and to stay neutral, but encouraged the interviewees to state their opinion and asked for explanations to gain a deep understanding of the participants' thoughts.

Interview questions about the different categories of an agents' appearance were asked to derive the interviewees' preferences. In order to get rid of (technical) constraints that limit interviewees' answers, interviewees were asked to imagine a future scenario where most of the households own a virtual assistant and where its usage is ubiquitous:

> "We are now 10 years in the future and virtual agents are so common that they can be found in every household. You also own a virtual assistant to support you in your everyday life. This agent helps you to design your everyday life and is on your TV, so you can talk to it. For example, you can schedule appointments with it, it also makes appointment suggestions and reminds you of important things. Furthermore, it can take on many other tasks or assist you with some things that might make your everyday life easier. You will actually talk to your virtual assistant every day and it is a great help."

When all this introductive information was given, the interviewer started with the actual interview questions. First of all, interviewees were asked some general questions about the visual appearance and the importance of virtual embodiment. Afterwards the questions became more precise and focused on the described categories (species, realism, rendering style and genre). Interviewees were asked what kind of species they felt most comfortable with. Further on, interviewees also stated and explained their preferences concerning the level of realism (e.g. realistic vs. line drawing, comic-like, unrealistic body characteristics like a blue skin or unrealistic big eyes etc.) and the rendering dimension (2D vs. 3D). Afterwards, interviewees had the chance to design their own virtual assistant with the Autodesk Character Builder. During the design process, interviewees were told to voice their thoughts to explain their choices. After their virtual assistant was designed, different questions to evaluate this constructed agent were asked. Additionally, further questions were asked that are not contributing to the current research questions and therefore are not presented here. When all main topics were discussed, interviewees were asked to summarize their thoughts so that their main preferences could be derived. In the end, socio-demographic data was obtained.

In order to analyze the data, all the conversations were transcribed into plain text based on the captured videos and the evaluation of the data was conducted deductively and inductively (Schreier, 2012) with reference to the research questions. Based on the interview guideline, which was constructed considering the research questions and hypotheses, a first version of the coding scheme was developed. This version was modified with regard to the interviewees' statements and some prior codes were excluded, since they became irrelevant for the current research questions (cf. Mayring 2010). Table 3.gives an overview of the final coding system of the interview questions with 11 main codes and 45 sub-codes. In order to guarantee the internal consistency of the constructed coding scheme, 2 interviews (one student interviewee and one elderly interviewee) were cross-coded by a second coder. Results

show substantial inter-rater reliabilities (Landis & Koch, 1977). Cohan's Kappa values are 0.69 for the student interview and 0.78 for the elderly interview, which is both assumed to be substantial.

Table 3.
Constructed Coding Scheme and concerning interview questions

Interview questions	**Coding Scheme**		
	Code	**Subcode**	**Definition**
How important is the virtual agent's embodiment for the users?			
Should your virtual assistant be an embodied character or could you imagine interacting with the voice only?	Preferred degree of embodiment (voice only vs. embodied agent)	No preference	Interviewees state no preference regarding the agent's embodiment
		Voice only	Interviewees state to prefer a voice only
		Embodied agent	Interviewees state to prefer an embodied agent
Why do you feel this way? Could you give me further explanations?	Reasons	Reasons why a voice only is preferred	All statements that contain a reason for a voice only
		Reasons why an embodied agent is preferred	All statements that contain a reason for an embodied agent
Do users want to customize the outer appearance of their virtual assistant?			
Would it be important to you to customize the assistant's appearance?	Preferred degree of customization	Wish customization	Interviewees like the idea to create the appearance of their virtual agent on their own
		Do not wish customization	Interviewees do not see any needs to create or change the appearance of their virtual agent on their own
		Choose from options	Interviewees would rather choose from a limited number of options than create their agent on their own
Why do you feel this way? Could you give me further explanations?	Reasons of customization	Pro Customization	All statements that contain a positive argument for the opportunity to customize the agent's visual appearance
		Contra Customization	All statements that contain a negative argument for the opportunity to customize the agent's visual appearance
Which factors would you like to customize?	Variables that should be customizable	Voice	Interviewees wish to customize the agent's voice
		Styling	Interviewees wish to customize the agent's styling
		Age	Interviewees wish to customize the agent's age
		Sex	Interviewees wish to customize the

			agent's sex
		Face	Interviewees wish to customize the agent's face
What kind of species do the users prefer?			
Please imagine all possible species your virtual agent could have. Which species would you like your virtual assistant to be?	Preferred species	Human	Interviewees preferred a humanoid agent
		Animal	Interviewees preferred a zoomorphic agent
		Robot	Interviewees preferred a machinelike agent
		Object	Interviewees preferred an object as their agent
		Mystical Character	Interviewees preferred a mystical character as their agent
Why do you feel this way? *Could you give me further explanations?* *Do you see strengths or weaknesses?*	Reasons of the preferred species	Pro Human	All statements that justify or contain a positive argument for a human as virtual agent
		Contra Human	All statements that reject or contain a negative argument for a human as virtual agent
		Pro Animal	All statements that justify or contain a positive argument for an animal as virtual agent
		Contra Animal	All statements that reject or contain a negative argument for an animal as virtual agent
		Pro Robot	All statements that justify or contain a positive argument for a robot as virtual agent
		Contra Robot	All statements that reject or contain a negative argument for a robot as virtual agent
		Pro Object	All statements that justify or contain a positive argument for an object as virtual agent
		Contra Object	All statements that reject or contain a negative argument for an object as virtual agent
		Pro Mystical Character	All statements that justify or contain a positive argument for a mystical character as virtual agent
		Contra Mystical Character	All statements that reject or contain a negative argument for a mystical character as virtual agent
Do users prefer 2D or 3D characters as their personal assistant?			
Would you prefer your virtual assistant	Preferred rendering dimension	2D	Interviewees prefer a 2-dimensional agent

to be 2-dimensional or 3-dimensional?		3D	Interviewees prefer a 3-dimensional agent
		No preference	Interviewees have no preference regarding the different dimensional rendering
Why do you feel this way? Could you give me further explanations? Do you see strengths or weaknesses?	Reasons	Pro 2D	All statements that justify or contain a positive argument for a 2-dimensional agent
		Contra 2D	All statements that reject or contain a negative argument for a 2-dimensional agent
		Pro 3D	All statements that justify or contain a positive argument for a 3-dimensional agent
		Contra 3D	All statements that reject or contain a negative argument for a 3-dimensional agent
Which degree of realism do users prefer?			
Should your virtual assistant be designed with regard to real facts? How important is a realistic appearance of the virtual assistant for you? Within the continuum of photorealistic and line drawing, how should your virtual assistant look like?	Preferred Realism	Comic stylized agent	Interviewees prefer a comic style agent
		Cartoon stylized agent	Interviewees prefer a cartoon style agent
		Line drawing agent	Interviewees prefer a line drawing figure as their agent
		Realistic agent	Interviewees prefer a photo-realistic agent
Why do you feel this way? Could you give me further explanations? Do you see strengths or weaknesses?	Reasons	Pro Realistic agent	All statements that justify a high degree of realism or contain a positive argument for a realistic agent
		Contra Realistic agent	All statements that reject a high degree of realism or contain a negative argument for a realistic agent
		Pro unrealistic agent	All statements that justify a lower degree of realism (comic, cartoon, line drawing) or contain a positive argument for a more unrealistic-looking agent
		Contra unrealistic agent	All statements that reject a lower degree of realism (comic, cartoon, line drawing) or contain a negative argument for a more unrealistic-looking agent

Additionally to the analysis of the statements, the interviewees' behavior during the design process was analyzed. The time that interviewees used to design different features of their virtual assistant was coded. The coding was done along the used online tool. As described above, only limited options existed and therefore, mostly choices with regard to feature specifications were obtained. Overall, there were seven main categories (initial selection, face, skin & make-up, eye color, hair style, body and clothing) and in regard to face, body and clothing, different subcategories could be varied (e.g. face- eyes, ears, chin, nose and cheeks).

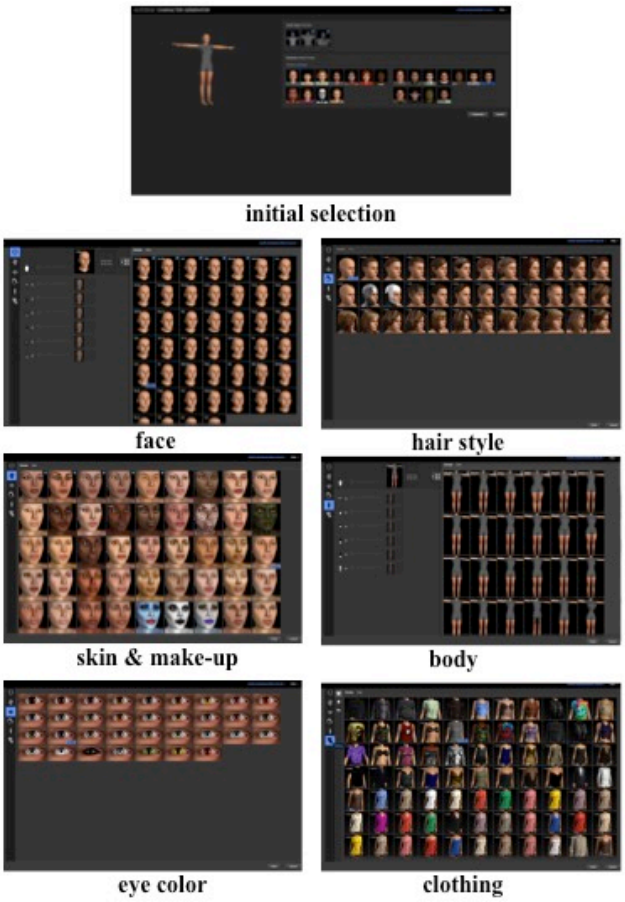


Figure 6. Screenshots of the used online tool (Autodesk Character Builder) with regard to all coded categories

Figure 6 presents the coded main categories and its options. For each category (including the subcategories), the active time (in seconds) an interviewee spent with the design was coded (starting with the first occurrence of the category and ending with the final decision or mouse movement within that category).

6.2.3 Sample

Since the influence of the users' age should be investigated, two different age groups were interviewed. Six students, aged between 20 and 39 years (*M* = 24.17, *SD* = 7.44), as well as five elderly people with an age between 61 and 74 years were interviewed (*M* = 67.00, *SD* = 6.12). Sex was equally distributed (45.45 % women). Most of the interviewees had no prior experiences with virtual agents (7 of 11), while four interviewees had interacted with a virtual agent before. Students were recruited via local Facebook groups and incentivized with course credits when desired. In order to recruit elderly people, posters in pharmacies, butcher shops and local bakeries were used. The elderly people received money for their participation. Interviewees' age, sex, professions and their prior experiences can be seen in Table 4 and Table 5

Table 4
Overview of the elderly participants and their experiences with virtual agents

Elderly interviewees						
				Prior experiences		
Abbreviation for subject	**Age**	**Sex**	**Profession**	**Knowledge about virtual agent**	**Seen/ used virtual agent before**	**Experiences with related technologies**
Senior 1	62	Male	Banker	Wrong conceptualization (non-autonomous program)	-	-
Senior 2	73	Female	Shop assistant	None	Seen Siri	-
Senior 3	74	Male	Professional expert	Wrong conceptualization (robot)	-	-
Senior 4	61	Male	Power plant employee	Wrong conceptualization (robot)	None	-
Senior 5	65	Female	Cosmetician	Little	None	Robots

Table 5
Overview of student participants and their experiences with virtual agents

Student interviewees						
				Prior experiences		
Abbreviation for subject	**Age**	**Sex**	**Profession**	**Knowledge about virtual agent**	**Seen/ used virtual agent before**	**Experience with related technologies**
Student 1	20	Female	Student	None	-	Robot
Student 2	20	Male	Student	Wrong conceptualization (speech-to-text program)	None	-
Student 3	20	Female	Student	Yes	Microsoft Clippy	-
Student 4	22	Female	Student	Wrong conceptualization (robot)	None	Avatars
Student 5	39	Male	Student	Wrong conceptualization (non-autonomous program)	Cortana, Siri	-
Student 6	24	Male	Student	Yes	Ikea Anna, Siri	-

6.3 RESULTS

In the following Section, the results of the eleven qualitative interviews are presented. The order of presentation is based on the hypotheses and research questions: First results refer to the agent's embodiment, followed by the results in regard to the different categories of appearance, afterwards statements concerning customization are described and in the end, statements and behaviors during the design process are presented. In order to get a better overview of the differences between the students and the elderly people and to answer research question *RQ3*, results of the sub-samples are presented separately.

6.3.1 Importance of a Virtual Agent's Embodiment

In hypothesis *H1* it was assumed that users prefer to interact with an embodied virtual assistant, since it was found to be beneficial in prior research. Therefore,

interviewee's were asked to state their thoughts and opinion with regard to embodiment. Moreover, the reasons and explanations were analyzed.

Student sample.

Referring to the agent's embodiment, only one student stated to prefer an embodied virtual agent as personal assistant, while four students preferred a speech-based system without an embodied character and one student interviewee did not state a preference. Some participants described problems with the appearance of a virtual agent and stated that embodied agents evoke feelings of *uncanniness* and therefore a non-embodied agent was preferred:

> "If I imagine my TV at home and there will be a huge head that talks to me… I don't know… I would feel scared." (Student 6, male, 24 years old).

Further, the students perceived an embodied agent to be a *simulation of a human*. They clearly rejected this simulation and perceived a non-embodied character as more fitting to the artificial nature of a virtual agent. Thus, students argued that the virtual assistant is no human and therefore should not simulate to be one:

> "I know that it is not a real human and this embodiment simulates one." (Student 2, male, 20 years old)

> "In my opinion, it is still a machine and nothing real. Therefore, I don't need such a humanoid character." (Student 4, female, 22 years old).

Besides the problems an embodied character evokes with regard to its appearance, students further see practical reasons to prefer a non-embodied agent. Interviewees who preferred a disembodied character justified this by referring to the embodiment as a potential *distraction*. They seem to perceive an embodied character to distract from the presented content, which lowers the usefulness:

> "I don't know, in that case [of a disembodied character] there might be more space for the information." (Student 6, male, 24 years old).

Furthermore, interviewees liked the idea of not being *restricted to one certain device*, since a voice can be present overall in the whole room or house. An embodied agent does need to be presented on a certain screen and therefore users can only interact at a specific place in their house with this agent:

> "If I would know the system is around me and I can use it (without any device), I would like that" (Student 5, male, 39 years old).

Although most students stated to prefer a non-embodied agent, they also saw advantages of an embodied agent. The main advantage they saw was the potential value of a virtual embodied character to be perceived as *interlocutor, who can be addressed* and where certain *impressions about the agent* can be made of:

> "I think having a contact person is more effective, even if it is virtual." (Student 5, male, 39 years old)

> "When I am talking to someone, it is nice to have an image of this person." (Student 1, female, 20 years old).

Elderly people

In contrast to the student sample, most of the elderly interviewees preferred an embodied character (n = 4), while only one elderly person stated that a speech-based system would be sufficient. The elderly person, who rejected an embodied agent used the same reason that was mentioned among the students, that an embodied character *simulates a human* and that she does not like this:

> "It is no human being. It is a computer voice in my opinion, and I don't need an embodied character. The voice only is enough." (Senior 5, female, 65 years old).

However, the majority of seniors stressed to appreciate an embodied character more. In line with some statements of the student interviewees, the elderly subjects explained that they would like to have a *contact person who they can address* during the interaction and that they therefore prefer an embodied agent:

> "There has to be some kind of reference point in the wider sense." (Senior 1, male, 62 years old)

> "I like to look in the people's eyes while I am talking to them. And I don't like it when there is only a voice present." (Senior 2, female, 73 years old).

Another person stated to be more *interested* in a conversation with an embodied agent and the character seemed to *attract attention* in the interviewee's opinion:

> "I think it is interesting if I look at it and the character talks to me." (Senior 3, male, 74 years old)

> "In the moment, when there is any kind of reference – figure, image whatever- for me there is another degree of attention." (Senior 1, male, 62 years old)

" One is probably more likely to look at it, when someone is talking to you." (Senior 4, male, 61 years old).

Additionally, the opportunity to *display nonverbal behavior* was perceived as beneficial. One person went on to attribute more *trust* to an embodied agent, while a non-embodied agent was perceived as impersonal and *unfamiliar*.

"I think, in that case you show more trust to it. I like that." (Senior 2, female, 73 years old)

"[talking about a non-embodied agent] something impersonal, something foreign, something mechanical" (Senior 1, male, 62 years old).

6.3.2 Preferences regarding Appearance Categorization

As described above, the constructed categorization was tested for the first time in this qualitative research. Therefore, research question *RQ1* aims to explore participants' preferences with regard to the different categories. It was for example asked how participants think about the degree of details and how they would explain their preferences. Moreover, in research question *RQ2* it was aimed to extract the categories of highest interest. Again, results are presented separate to gain a better understanding of age-related differences, which were examined with research question *RQ3*. The last category of the categorization "feature specification" was investigated using the design process described in Chapter 6.3.4, therefore participants' expressions and thoughts with regard to feature specifications are not described in this Chapter.

Student sample

With regard to *species* (Category 1), most of the students rejected a humanoid agent, since only one interviewee (n = 1) stated to prefer to have a human appearance as their personal virtual assistant. In contrast, most preferred non-humanoid creatures such as animals (n = 2), objects (n = 1) or machinelike agents (n =2). No one preferred a monster and all interviewees stated that they favored a likable appearance to be their personal assistant, which would not be the case with a monster character.

Although most rejected a human appearance, students expressed certain reasons to prefer a humanoid appearance. Reasons for liking a humanoid agent were the *natural interaction* and the *familiarity* a human appearance evokes:

"I think it would be most naturally." (Student 6, male, 24 years old)

> "Since it is a habit to talk to other humans, to communicate with them, to negotiate appointments… and therefore I would take a human." (Student 6, male, 24 years old).

Additionally, it was mentioned that a humanoid character would be the *most serious* one:

> "that [a human appearance] will be perceived as most serious" (Student 2, male, 20 years old).

As described above, most students rejected a human appearance, since they see clear disadvantages. People who disliked a humanoid appearance justified it with *missing variety:*

> "There are enough humans around me. I don't feel the need to interact with a human in this way in addition." (Student 5, male, 39 years old).

Again, similar explanations used with regard to embodied agents occur. Students stated to dislike the *simulation of a human,* which is obvious for human appearance, while non-human appearances are clearly distinct form real humans:

> "that is no person anyways" (Student 3, female, 20 years old)

> "You don't have to simulate that" (Student 2, male, 20 years old).

Furthermore, student participants assumed that the humanoid character would *trigger specific expectations*, which the virtual agent cannot fulfill:

> "I would like to interact normal with a humanoid character and do chit-chatting or something like that. And I cannot do that with a virtual agent, since he has no personality" (Student 2, male, 20 years old)

> "I don't know to what degree people will be influenced by the fact that there is a human… so they might think: Ok, it is a human, therefore I have to be polite or something like that" (Student 3, female, 20 years old).

While humans were described as serious, objects and animals were chosen because of their *funny appearance*:

> "Simple because it is funny" (Student 4, female, 22 years old).

Animals were preferred since there are *pet-like*. People are used to have pets in their houses and therefore an animal appearance is also seen to evoke familiarity and will be perceived as *likable*:

> "You have dogs and cats around you in your life and they are your pets… and it is not unnaturally, if they are with you in your home" (Student 6, male, 24 years old)

"I think an animal is likable … a little helper" (Student 3, female, 20 years old).

In addition, with regard to objects, some interviewees wished for something *neutral and task-related*. An object can easily be adapted to the task it is fulfilling:

"Yes, I like an object, because it is neutral (…) on the other hand it is related to the topic" (Student 3, female, 20 years old).

With regard to *realism* (Category 2), only one interviewee in the student sample preferred a realistic appearance of his virtual agent:

"I think that would be nice… realistic… I would appreciate it, if the appearance would be natural and not completely absurd" (Student 6, male, 24 years old).

The other students (n = 5) rejected a realistic appearance and stated not to wish for a high resolution with a high degree of details:

"I don't have to see any hair and pore of it, like it gets popular in recent graphical developments. That doesn't need to be. Depending on how the character is designed that could be deterrent to me." (Student 5, male, 39 years old).

Moreover, most of the interviewees stated that they did not like the fact that realistic-looking humans reminded them of humans from their real world and that they want to *distinguish between fictional and real worlds*. This thoughts fit to the aforementioned explanation that students dislike a *simulation of humans*:

"I don't know, but if you design your virtual agent on your own, I think one would link it to already existing people and therefore I like the cartoon version more." (Student 1, female, 20 years old)

"it is no realistic human, because it is no real human" (Student 6, male, 24 years old)

"In my opinion, I interact with a virtual person… so the distance between my computer and me will be still maintained. Because I have a fictional character I am still aware of it" (Student 5, male, 39 years old).

The majority of the student sample prefers a *3-dimentional* (Category 3) agent (n = 5), while only one person (n = 1) liked the *2-dimentional* versions more, because it appeared to be more simple and less distracting:

"I would prefer the 2-dimentional, because it is probably more simple, because it is not distracting" (Student 2, male, 20 years old).

The other interviewees liked the possibility to make movements and turns within the virtual space of a 3-dimentional agent. Further on, they appreciate the modern look of it:

> "I would create it with 3 dimensions, so that it is more modern-looking" (Student 1, female, 20 years old)

> "it [a 3-dimentional agent] looks better done" (Student 3, female, 20 years old).

Contradicting to the prior findings in regard to species and realism, interviewees stated that they liked the more realistic look of a 3-dimentional agent:

> "I think 3D is more realistic" (Student 6, male, 24 years old)

> "I think, it looks more likable, if it is more real" (Student 3, female, 20 years old).

Elderly sample

Referring to the agent's *species* (Category 1), seniors appeared to have a clear preference for a human appearance. Contradicting the statements of the student sample, almost all elderly persons tend to prefer a humanoid agent as their personal assistant (n = 4), while one person (n = 1) picked an animal.

The one person, who preferred an animal, could *not accept a humanoid agent*, which is in line with the expressions of the student sample:

> "Because it is a computer, I can not accept it as a human." (Senior 5, female, 65 years).

However, in contrast the majority preferred a virtual human to be their personal assistant. One main reason seems to be the familiarity to talk and interact with humans, which can also be applied to virtual humans. Seniors stated that a humanoid agent triggers a more *familiar* interaction compared to an animal:

> "… she talks to me… great I can talk to her… and if I see an animal...well, yes I can talk to it too, but it does not respond. Therefore, I would choose a human." (Senior 3, male, 74 years old).

This familiarity seems also to lead to a *higher interpersonal relationship* between the user and the agent, as one elderly stated that he would feel more related to a humanoid agent:

> "You might have a relation to it or a better relation." (Senior 4, male, 61 years old).

In contrast, the interaction with non-human appearances like objects or animals was clearly rejected from most seniors. For some people it even felt *absurd* to interact with an animal or object:

> "I imagine a better interaction (with a human) instead of such kind of puppet." (Senior 2, female, 73 years old)

> "With regard to the object I feel somehow stultified." (Senior 4, male, 61 years old).

With regard to *realism* (Category 2), nearly the same argument structure was used, that occurred above with regard to the species. A realistic appearance was preferred by most of the elderly (n = 3), while two of them (n = 2) indicated they would like to have an agent with an unrealistic appearance. Those, who preferred an unrealistic appearance, named reasons, which were similar to the ones mentioned by the students. Basically, they wanted to be able to visually distinguish between virtual and fictional characters:

> "which is clearly recognizable as fictional character" (Senior 1, male, 62 years old)

> "the comic-style… since it is no real human on the computer, I don't need a realistic image" (Senior 5, female, 65 years).

In contrast, the other three elderly, who preferred a realistic appearance, liked the fact that it came pretty close to a real human, as an interaction with humans is normal (and familiar) to them:

> "Yes, when I see this natural image… this is a human, but this figures are not (…) here I had the feeling of talking to a real human" (Senior 2, female, 73 years old)

> "A normal looking human, really normal." (Senior 3, male, 74 years old).

Moreover, with regard to a realistic appearance, they see benefits for the relation between them and the agent. In line with statements found regarding the species, seniors stated that they could *trust* a realistic-looking (human) agent and could even imagine some *bonding* to this agent version:

> "I can trust him, if I talk to him" (Senior 2, female, 73 years old)

> "That would be the best! In that case you could even build some kind of relation if you interact with it longer" (Senior 4, male, 61 years old).

Furthermore, an unrealistic look was rejected since it was perceived as not acceptable and even inappropriate. An unrealistic look was described as following:

"kid's stuff" and "abnormal" (Senior 3, male, 74 years old).

The elderly sample did not have a major preference in regard to *dimensionality* (Category 3). One interviewee evaluated the dimensionality as:

"irrelevant in the end" (Senior 1, male, 62 years old).

While two interviewees (n = 2) preferred the 2-dimentional agents, the other two would choose 3-dimentionality (n = 3). Elderly preferring the 3D agent justified this by perceiving higher *lifelikeness* and emphasized to prefer the higher degree of realism:

"I see more live in this one" (Senior 3, male, 74 years old)

"yes, because it is more realistic" (Senior 4, male, 61 years old).

On the other side, seniors, who preferred a 2-dimentional agent, liked the *simple look* of these appearances:

"This is sufficient… I prefer the simple design" (Senior 5, female, 65 years old).

6.3.3 Customization

In hypothesis *H2* it was assumed, that the customization of an agent's appearance will be perceived as promising, since on the one hand the user is able to adapt his or her own preferences and on the other hand people generally evaluate things more positively when they built it on their own. Therefore, participants were asked to state their feelings about a customization process and to give explanations for their feelings. Again, both sample groups are presented individually to extract the general opinion climate of the subsamples better and thus to investigate RQ4.

Student sample

Four student interviewees (n = 4) stated that they would *like to customize* the appearance of their personal assistant, one person (n = 1) stated *no preference to customize* the appearance and another person (n = 1) preferred *choosing from different options* instead of manipulating the appearance on their own. The two latter interviewees did not see any importance in the design of the appearance:

"I don't have any requirements, it doesn't matter how it looks like. " (Student 3, female, 20 years old)

"It is irrelevant as long as there is someone responding." (Student 5, male, 39 years old).

Furthermore, the design process of the appearance was described as too time-consuming:

"You can chance the character in many different ways and I think in this case it would be excessive and inappropriate." (Student 5, male, 39 years old).

In contrast, as most students tend to like the customization of an agent's appearance, they stated various advantages of this. Persons, who liked to customize their assistant, *liked the process of designing* itself and appreciated taking their *own preferences* into account:

"Yes, I think it would be fun to design it." (Student 1, female, 20 years old)

"Everybody has their own preferences" (Student 1, female, 20 years old)

"it is important that you like the appearance of it" (Student 2, male, 20 years old).

Furthermore, besides the adaption to the own preferences, interviewees liked to customize their agent in order to express individuality:

"If everybody owns one, you can distinguish between the different agents" (Student 1, female, 20 years old)

"It adds a personal touch… a very personal assistant" (Student 6, male, 24 years old).

Elderly sample

The majority of the elderly people felt *no need to customize* the virtual assistant (n = 3), while only one elderly person (n = 1) *liked to customize* the appearance. In addition, one male interviewee (n = 1) would have liked to *choose from variations* instead of designing it, because he did *not like the process* of designing itself:

"I don't feel like designing it completely (…) without spending half an hour to design" (Senior 1, male, 62 years old).

Similar to some statements of the student sample, the person who liked the idea of customization wanted to consider their own preferences:

"I would adjust it with regard to my preferences" (Senior 3, male, 74 years old).

The other elderly stated that the outer appearance is not as important for them and therefore it is irrelevant to them to customize the agent's appearance:

"For the moment it is irrelevant to me" (Senior 4, male, 61 years old)

"Appearance is not decisive." (Senior 2, female, 73 years old).

6.3.4 Design process

In the following, the results of the actual choices interviewees made during the design process of their own virtual assistant (done with the Autodesk Character Builder) are presented. Here mainly derivations about preferences regarding the agent's feature specifications can be made, which helps to answer the research questions *RQ1-3*.

Student sample

Overall, it took the students on average about 10 minutes to design their individual virtual agent. In order to explore the most important factors during the design process, the time students spend within the different design categories of the tool were coded (see Table 6).

Table 6
Time interviewees spend with the design of different design categories

Design categories (obtained from the Autodesk Character Builder)	**Design duration (time spend with designing) in seconds**	
	Student sample (Mean)	Elderly sample (Mean)
Initial selection	M = 60,234	M = 113,346
Face	M = 134,451	M = 86,584
Skin & Make up	M = 73,226	M = 41,418
Eye color	M = 36,320	M = 46,818
Hair style	M = 75,510	M = 70,065
Body	M = 54,613	M = 67,006
Clothing	M = 174,174	M = 102,361
Overall	M = 608,527	M = 527,598

The three categories with the longest design durations were clothing, face and hair, while the manipulation of the eye color and body shape was done faster and therefore seems to be less important.

When interviewees are limited to humanoid characters (due to the limitations of the customization tool), they tend to prefer character features, which are close to the reality (see Figure 3). Only two students chose an unnaturally skin tone (Student 1, female, 20 years old) or hair color (Student 5, male, 39 years old), although most students stated earlier in the

interviews, they would prefer a more unrealistic agent. Most students stated to intend the design of a *normal* or *natural looking* agent:

"This looks pretty normal to me" (Student 4, female, 22 years old)

"I take this one… this is the most natural one for me" (Student 4, female, 22 years old)

" I don't understand those extreme colors […] that looks strange and I would like to have a pretty neutral assistant" (Student 3, female, 20 years old)

"Yes, I don't know… this is…hardcore. Beside those extremes, I think this looks pretty natural" (Student 6, male, 24 years old).

Regarding clothing, most (n =4) students picked a more *casual look* with jeans, a simple top and sneakers. They stated to choose unobtrusive clothes and to aim for something *serious*:

"So, I try to choose something unobtrusive" (Student 4, female, 22 years old)

"Yes it's good… I don't change anything… don't know […] that's serious" (Student 6, male, 24 years old).

Against this background, they preferred for instance jeans and sneakers:
"a jeans… yes this is pretty normal… yea this looks pretty neutral and than she'll get some pair of shoes… yes a pair of sneakers" (Student 3, female, 20 years old).

Figure 7. Designed agents of the student sample

Elderly sample

The design process of the elderly people took on average about 9 minutes. Table 4 shows the specific times elderly spend within the different design categories of the customization tool. The three categories with the longest design durations were initial

selection, clothing and face, while it took the least time to design the eye color as well as skin and make-up.

Elderly people spend way more time than students to decide which character they choose first in order to manipulate it the way they want it to be. Probably because of the well-decided initial selection, the design duration for most of the other categories (e.g. clothing, face or hair) was smaller in comparison to the students' durations. Since elderly already stated they do not feel to customize their agent, it might be more beneficial for them to choose from different alternatives instead of designing every feature on their own.

The designed agents of the elderly people are presented in *Figure 8*. Just like the students the elderly sample mostly intended to design a serious and normal looking agent:

> "He looks serious"(Senior 2, female, 73 years old)

> "A realistic figure, when male or female, I would take female. I like serious clothing, a nice hairstyle…" (Senior 3, male, 74 years old)

> "yay that is normal… for me this looks more normal, while the other things wear to much make-up or the facial expression is not appealing to me" (Senior 4, male, 61 years old).

As a reason why they prefer a normal looking agent one interviewee stated the following:

> "so you can communicate better with it or I would say, because one isn't used to something else." (Senior 4, male, 61 years old).

Only one woman designed an unrealistic character with blue skin color, since she preferred to have a comic-style agent:

> "This is funny as well… I think she can stay blue now […] since I still like to have a comic character and don't like to have something really humanoid. In this sense… even when this is a humanoid face, it is still close to a comic character" (Senior 5, female, 65 years).

It has to be considered that this interviewee did not liked the idea to use a virtual agent at all and had a really negative attitude toward virtual agents:

> "So I don't like that some kind of computer character or a computer or a robot takes care for me in the future or something like that. I have strong concerns about it." (Senior 5, female, 65 years).

With regard to clothing, the elderly also mostly prefered casual clothing like jeans and a T-shirt:

"a bit more sporty" (Senior 2, female, 73 years old),

"yes, now he doesn't look that formal anymore" (Senior 2, female, 73 years old)

"I think this normal one is ok, since it is normal to me and not so extreme like these tight tops…" (Senior 4, male, 61 years old).

Figure 8. Designed agents of the elderly sample

6.4 DISCUSSION

In this Chapter the results described above are summarized and implications with regard to age-related differences will be made. Moreover, implications for further studies referring to the constructed categorization are highlighted. At the end, the study's shortcomings with hints for future studies as well as a conclusion compressing the contributions of the present study are presented.

6.4.1 Result Summary and Interpretation

The main goals of the presented approach were (a) to explore the constructed categorization of different appearance variables for the first time as a basis for further investigations and (b) to gather first insights into age-related user preferences with regard to a virtual assistant's appearance. For this purpose, a qualitative interview study with five elderly people and six students was conducted, where personal preferences with regard to the constructed categories of the agent's appearance were examined.

In hypothesis *H1* it was proposed that users would prefer an embodied character instead of a merely speech-based system. Results indicated that this is true for the elderly, who favored having an interlocutor to talk to and look at, while the students perceived more advantages in a solely speech-based system. Students looked at the system in a more technical way and stated that they would like the opportunity to use it without being restricted to certain

devices (such as a computer display or smart TV). This might be due to the current developments of personal assistants on smartphones. Most of the students might be used to Siri, Cortana or Google and the mobile usage of those systems. On the other side, the elderly, who are less experienced with technical devices, might wish for an embodied character as a reference. Although most of the students did not prefer having an embodied character, this does not necessarily mean, that an embodied character would not be beneficial to them; as prior research demonstrates the potential positive effects of an embodied agent (Yee et al., 2007). Here, the perceived preferences might differ from the behavioral outcomes an embodied agent might produce.

With regard to the research question *RQ1* and *RQ3,* different preferences along the appearance variable categories have been found for the students and seniors. The overall tendencies of both age groups appear to be consistent over the different categories: Students indicated to prefer a rather unrealistic non-humanoid appearance to prevent a mix up of virtual and real worlds, while seniors liked a realistic humanlike appearance, as it evokes familiarity and social processes like trust and bonding. Most of the currently used virtual agents are humanoid agents. It became evident, that this is appropriate for elderly people since they prefer to interact with a realistic-looking human; but students mostly rejected this kind of appearance. In line with the findings by Koda and Maes (1996), students stated that they prefer a more cartoon-like appearance in order to avoid distractions from important information. While students consider the social effects of a realistic humanoid appearance as negative, the elderly do see advantages of this appearance. Students are aware of the social expectations of a realistic humanoid appearance, while the elderly prefer to interact naturally with a “normal human”. Thus, for the elderly, who may use virtual agents as daily life support, a more realistic looking humanoid agent seems to be the most suitable. The elderly even stated that a realistic humanlike appearance will evoke trust and can help to build an interpersonal relationship. This might be especially promising, since seniors were found to be rather skeptical towards the usage of virtual entities (Scopelliti et al., 2005). Implications from this study might demonstrate that when a realistic humanlike appearance is used, the appearance will trigger trust and feelings of bonding, which as a consequence might enhance usage intentions and the overall frequency of interactions with the agent. In general, the overall preferences of students and elderly differed in most categories. Therefore, there are age-related differences in the preferred visual appearance of a virtual agent that need to be taken into account, when a virtual agent is designed.

Nevertheless, students and seniors did not differ that much in their preferences with regard to feature specifications. It became obvious that during the design process most participants aimed to design a normal looking humanoid appearance with a serious look and more casual clothes (jeans and T-shirt). Thus, although students stated to prefer a rather unrealistic look to distinguish virtual and real worlds, when they are limited to humanoid characters with a rather high realism, they aim to design a realistic looking agent. To conclude, when a human character (with high realism without any stylization) is designed, results indicate that realistic feature specifications are preferred regardless of the age group. Moreover, it has to be marked that in contrast one elderly participant designed her virtual agent with unrealistic feature specifications (e.g. blue skin). This participant stated a very negative opinion about virtual agents and therefore, (when she was forced to choose or design an agent's appearance) she preferred an unrealistic appearance with regard to all categories. These statements contrast with most of the thoughts of the other elderly participants and most likely caused by the general negative attitude toward virtual agents. Hence, in future research, the attitudes toward virtual agents of the users are worth investigating since they might moderate the evaluation and perception of appearance variables.

In reference to research question *RQ2*, the results indicate that the categories embodiment, species and realism are the one of most interest. Within those categories, on the one the hand the most age-related differences were found and on the other hand the statements within one age group were mostly consistent. Moreover, participants expressed clear opinions related to these categories, while for example seniors described the category "2D vs. 3D" as irrelevant. Additionally, only little variance in preferences regarding the agent's feature specifications occurred and those were highly depending on the former choices with regard to species and realism. Since the current dissertation mainly focuses on appearance aspects, it can be derived from this study that the species and the degree of realism are the categories which should be investigated in the following studies.

With regard to customization, it was assumed that most users would like the idea of customizing the appearance of their virtual assistant (*H2*). Students mostly liked to customize the appearance of the virtual agent since they liked enhancing the individuality of their virtual assistant. These findings appear to be in line with the findings by Marathe and Sundar (2011), which indicate that customization increases the sense of identity. When users create the appearance of their virtual assistant in regard to their own preferences and like to express individuality, their sense of their own identity might also be affected. Moreover, students see

advantages of the design process itself and of the option to take their own preferences into account and to design an appearance, which matches their expectations.

With regard to the research question *RQ4*, again age-related differences occur in the perception of the beneficial potential of a customization process. In contrast to the thoughts expressed by students, most of the elderly people abstained from the opportunity to customize the appearance of their agent. To some degree this might be caused by their lower technical experience. Most of the students have probably already designed a virtual character before (e.g. in video games), as some of them even made references to other customization tools ("similar to the possibility of a Wii", Student 1, female, 20 years old). Therefore, the customization and design of a virtual character might be more familiar to students than to the elderly. In this sense, the elderly might tend to overrate the effort of this design process and therefore avoid it. Thus, customization might be beneficial for elderly too, although they do not recognize the potential. Within the design process itself, it was obtainable that elderly spend way more time with the initial selection of an agent instead of designing specific features on their own. Therefore, the most effective opportunity for elderly people might be to choose from different appearance alternatives, instead of customizing every specific feature on their own. After this design process, elderly and students were equally satisfied with their constructed agents. Although no direct derivations about the effect of customization can be concluded by the presented findings, since only the thoughts about the idea to design their own agent's appearance were gathered, promising results occurred. Especially for seniors, who have fewer experiences with such a design process and the usage of technology in general, a selection of different options might be more promising than a full design process. This matches with the findings from Diehl et al. (2017), who found out that the choice of a agent had a more positive effect than the design process of an agent's appearance. Therefore, future studies should investigate age-related differences and the effect of choosing an agent's appearance for elderly people more deeply.

6.4.2 Limitations & Future Research

Due to the qualitative method, the sample size is very small and therefore the generalizability of the results is limited. But since the interviewees' statements are highly consistent, the presented research gives interesting first insights into the preferences of potential users of a virtual daily-life assistant. Those results should be investigated in more detail in the future by using quantitative methods in order to enhance generalizability.

Most of the examined appearance variables were investigated solely based on imaginations and without designing the virtual agent in real time. Therefore, it might have

been hard for the interviewees to imagine all possibilities within the different categories. In order to stimulate the imagination of the interviewees, different stimuli pictures were presented. This might have been beneficial to clarify the current appearance category during the interview, but in some cases it showed to be problematic. The stimulus pictures might have primed interviewees opinion, although the interviewer reminded them several times to see those stimuli as examples for the represented category rather than evaluation the pictures.

Moreover, it might be one thing to imagine the perfect assistant and another to interact with it. When the interviewees actually had interacted with a virtual agent, their preferences and evaluations might have been quite different than the imagined ones without any real interaction. Additionally, some responses to an agent's appearance might be unconscious (as assumed from the media equation theory (as assumed from the media equation theory; Nass & Moon, 2000)), so that participants cannot express a certain reaction, since they are not aware of it. Therefore, the main results regarding species and realism need to be examined in further experimental approaches, where users evaluate and interact with real stimulus material. Those quantitative experiments should compare the evaluation of the presented appearance categories systematically by producing different stimuli combinations of the various categories. As species and realism were found to be the categories of highest interest, experimental approaches should start with those two.

In future research, it might also be interesting to examine the effect of customization of the virtual agent's appearance in a real human-agent interaction with elderly people. Since the current results indicate that the elderly opt out of customizing their virtual assistant, it has to be investigated whether customization (or at least choosing from alternatives) can still enhance the interaction.

6.4.3 Conclusion

This study explored the previously constructed categorization of different appearance variables as well as age-related differences with regard to appearance preferences for the first time. Qualitative interviews revealed that the species as well as the degree of realism (especially stylization) seem to be the categories of most interest. Moreover, the presented findings indicate that it is highly important to study the aimed target group, since various differences between students and elderly people in regard to appearance preferences were found. For elderly people, the social and familiarity aspects in a virtual agent's appearance are the most important features. They like to stick as close to a real human appearance as possible, since they desire a more natural and social interaction. In contrast, students reject

those kinds of appearances, since they clearly want to distinguish between real and virtual humans. The agent should not pretend to be a real human and students would like to be aware that it is solely a computer system which should not simulate being more social than it is. To conclude, there are first hints that age-related differences in user preferences of a virtual agent's visual appearance exist.

7 STUDY 2: QUANTITATIVE EXAMINATION OF THE PERCEPTION AND EVALUATION OF DIFFERENT APPEARANCE VARIABLES WITH REGARD TO USER CHARACTERISTICS

Since the results of Study 1 gave valuable insights in the age-related differences with regard to users preferences of a virtual agent's appearance, Study 2 was conducted to examine these differences in more depth. More over the qualitative results from Study 1 become investigated using a systematically experimental approach. From Study 1 species and realism have been extracted as the appearance categories of most interest. Thus, the present study examines the evaluation and perception of an agent's species and realism with regard to user characteristics that might moderate these effects.

7.1 BACKGROUND & HYPOTHESES

There is a continuous development of virtual agents and their possible applications. Virtual agents are computer interfaces mostly represented through a virtual embodied character and are humanlike in the way they communicate by using verbal and nonverbal cues (Cassell, 2000). Based on these abilities, those agents are promoted as future personal everyday-life assistants, which might help scheduling and remembering appointments, inform about the weather, or assist with shopping (Kopp et al., 2018). Although the latest commercial agents (such as Siri or Alexa) are solely speech-based and represented by a voice only, an embodied character was found to enhance the human–agent interaction in a positive way (c.f. Yee et al., 2007). A meta-analysis revealed the impact of an embodied representation (Yee et al., 2007): a comparison of 46 papers showed that the representation of a humanoid face had, in most cases, a more positive outcome compared to the absence of a represented face (Yee et al., 2007). Hence, the human–agent interaction seems to benefit from an embodied character.

Furthermore, multiple studies demonstrate that the appearance of those embodied characters affect different social variables (e.g., motivation (Baylor, 2009b) or the agent's general evaluation (Ring et al., 2014)). Thus, a virtual agent's appearance has a high impact on the human–agent interaction and is of great importance.

Accordingly, it is important to adapt the appearance of the agent to the user's needs and preferences in order to enhance beneficial effects of human–agent interaction. With regard to the manifold application fields, the target groups and potential users are diverse and will differ in multiple variables. The application as a virtual assistant, for example, seems to be especially beneficial for people in need of support, such as elderly or cognitively impaired people (Yaghoubzadeh et al., 2013). Nowadays, those people are dependent on other people's help, but for some tasks, virtual agents can assist their daily life to enable a more self-

determined life. This special target group largely differs from digital natives, such as students, who are usually employed in user studies (e.g., regarding technical skills, prior experiences with technologies, or cognitive abilities). Therefore, it is important to analyze the specific needs and preferences of older and/or cognitively impaired users not only with regard to functions, but also with regard to appearance variables. Besides specific target groups, also users with special characteristics concerning attitudes, personality traits, or prior experiences will interact with a virtual assistant. These characteristics also need to be taken into account when the virtual agent's appearance will be analyzed. This might eventually help to enhance the acceptance of these new technologies and to enrich the social interaction.

The best opportunity to tailor the appearance to the user's needs seems to be the customization by the users themselves. However, this process is expensive, time-consuming, and mostly requires technical skills, which special target groups might be afraid of and not able to do. Similarly, results showed that the free customization and own design process of a virtual agent's appearance does not lead to higher user satisfaction compared to a choice out of different options (Diehl et al., 2017). Therefore, it can be more helpful to know the potential user group and its specific characteristics (e.g., personality traits such as tendency to anthropomorphize) and tailor the appearance or the choice of appearances to these needs. In order to do so, the impact that user characteristics such as age, sex, cognitive impairments, or personality traits have on the perception and evaluation of different appearance variables need to be analyzed. Until now, only little has been known about the preferences of these special target groups such as elderly or cognitively impaired people (c.f. Chattaraman et al., 2011; Straßmann & Krämer, 2017) or the influence of user characteristics such as personality traits on the perception of appearance variables. Prior research (Chattaraman et al., 2011; Straßmann & Krämer, 2017) used qualitative approaches to analyze the preferences of elderly people in regard to a virtual assistant's appearance. Although this gives interesting insights, quantitative research is needed to specify these findings more. Additionally, no study investigated the preferences of cognitively impaired people, who are a highly relevant user group (Yaghoubzadeh et al., 2013). The current approach aims to fill this research gap and explores the effect of user characteristics on the perception and evaluation of different appearance variables in two studies.

Based on previous research, the species of an agent and the degree of realism were identified as key variables to affect human–agent interaction. Accordingly, prior research demonstrates that especially for a specific target group such as elderly, those appearance

variables are the ones of most interest (Straßmann & Krämer, 2017) and thus need to be investigated in more depth.

Although most embedded characters are humanoid (Chattaraman et al., 2011), a humanoid character does not necessarily have to be the best choice, since animals, for instance, were found to enhance buying intention (Terada et al., 2015) or learning outcomes (Sträfling et al., 2010) more than humanoid characters. Therefore, it is highly important to take the species of the virtual agent into account more. However, most studies did not systematically investigate the effect of species, since most of the used stimulus materials did also differ regarding other variables such as realism. Realism, in itself, has been shown to be relevant not only in terms of uncanny valley effects (Mori et al., 2012), but it was also demonstrated that a cartoon stylization can have positive effects on the social interaction (Yee et al., 2007) and the agent's overall evaluation such as appeal or perceived friendliness (McDonnell et al., 2012). However, realism has not been investigated systematically in combination with the species of the agent. Thus, there is a lack of controlled, systematic research on the effect of species and realism and their potential interactions.

Therefore, the present studies aim to examine the perception and evaluation of these variables in more depth and additionally explore the effects of user characteristics.

7.1.1 The Effect of the Agent's Species

Although virtual agents are fictional characters and could easily take up different forms of species (e.g., animals, objects, or humans), Chattaraman et al. (Chattaraman et al., 2011) found that 91% of virtual agents used on e-commerce websites were humans or humanoid, and findings highlight a preference for humanoid faces (Forlizzi, Zimmerman, Mancuso, & Kwak, 2007; Koda & Maes, 1996; Parise, Kiesler, Sproull, & Waters, 1999). However, humanoid agents might not necessarily be the best choice for all occasions. Terada et al. (Terada et al., 2015) demonstrated that buying intention varies with regard to different species of the agent (Buddha, robot, real person, text, dog, virtual human). It has to be noted, however, that to some degree, the authors not only compared different species, but also their embodiment (text only vs. embodied character) and the perceived realism (real vs. virtual human). The results of that study indicate that an animal and a virtual human evoked the highest values of buying intention, while no embodiment, a robot, and a photorealistic human did not differ from each other and the object leads to the lowest buying intention. Although a virtual human achieved a more positive evaluation than a robot or an object, the animal was also evaluated as being equally positive to a cartoon-like human. Thus, humanoid agents are not necessarily the gold standard. In line with this, a nonhumanoid tutor has been found to be

more effective in an online learning environment. Sträfling et al. (Sträfling et al., 2010) compared the effect of an humanoid agent to a cartoon-stylized rabbit on learning outcomes. Their results showed that learners preferred a cartoon-like rabbit over a humanoid agent to be their tutor. The rabbit was rated as friendlier and learners exposed themselves longer to the material compared to those who received advice from a humanoid agent. Those results demonstrate that a humanoid character is not always the best solution, since objects or animals have been shown to evoke more engagement and a more positive outcome. However, those results are based on specific appearances, which have not been varied systematically, since the used agents not only differed in species, but also in other appearance variables such as realism. Although these findings provide initial hints that the species of the agent is important and that nonhumanoid appearances seem to have positive effects, further systematic research has to be made to scrutinize the effects of those variables on user responses more closely. Therefore, this research aims to answer the following research question in a more systematic manner:

> *Research Question 1 (RQ1)*: Which species will be perceived and evaluated as most positive?

7.1.2 The Effect of the Agent's Degree of Realism

As mentioned before, the realism of the agents was also found to have an impact in human–agent interaction. Yee and colleagues (Yee et al., 2007) conducted a meta-analysis and found that a humanlike representation with higher realism leads to more positive social interactions compared to representation with lower realism. However, the meta-analysis showed that this effect was only true for subjective measures and not for behavioral ones. It has to be noted that the authors only used division into high and low levels of realism, without any further distinction. As the categorization of appearance variables (pictured in Figure 1; (Straßmann & Krämer, 2017)) demonstrated, realism is more multilayered. Subcategories such as stylization, degree of detail, and resolution affect the overall perception of realism, and thus this concept is more complex than the distinction into high and low realism. Since different subcategories can affect the degree of realism (Straßmann & Krämer, 2017) and the meta-analysis gives no further information about these categories, no deductions about the exact effects of these variables can be made. However, other authors have investigated the effects of realism in greater detail and focused on specific realism variables. These studies stressed the effect of more specific realism variables, such as stylization (McDonnell et al., 2012; Ring et al., 2014; Zell et al., 2015). McDonnell et al. (McDonnell et al., 2012) compared 10 different rendering styles (ranging from line drawing to a highly realistic

representation) and found that those which were highly abstract or highly realistic were perceived as equally appealing. A rendering style that was neither abstract nor realistic was less appealing to the users. Furthermore, this study showed that a cartoon stylization evokes the attribution of friendliness. These findings are in line with the results of Sträfling et al. (Sträfling et al., 2010), who found that a cartoon-like rabbit was perceived as more friendly. Similarly, Ring et al. (Ring et al., 2014) examined the effect of a cartoon-looking shading style and character proportions with regard to different tasks of the virtual agent. Results revealed that a realistic shade was evaluated as being more realistic, while a cartoon shade was rated as being more friendly. In line with this, the cartoon proportions evoked more friendliness than human proportions. Moreover, they found interaction effects of appearance and the presented tasks. A realistic character seems to be more suitable for medical tasks, while the cartoon-like agents were preferred in social tasks (Ring et al., 2014). In contrast to this, in the context of health applications, a more realistic-looking agent was evaluated more positively with regard to likability, competence, and trustworthiness compared to a stylized agent (van Wissen et al., 2016). These findings might be caused by the specific application, since Ring and colleagues (Ring et al., 2014) showed that users prefer a realistic character for medical tasks.

In summary, prior research (McDonnell et al., 2012; Ring et al., 2014; Sträfling et al., 2010) showed that a cartoon stylization of a virtual agent evokes the perception of friendliness and that this is true for both the shade (or rendering style) of the agent and the proportions of the agent (Ring et al., 2014). At the same time little evidence exists to demonstrate that for medical and health-related tasks a more realistic appearance might be preferred (van Wissen et al., 2016).

Nevertheless, it is still unclear how realism variables affect the perception of a virtual agent, especially in the context of daily-life support. As the context of the application seems to be crucial (Ring et al., 2014), effects might be different for assistive technologies in everyday life. Hence, this research investigates the following research question:

> *Research Question 2 (RQ2)*: What degree of realism will be perceived and evaluated as being most positive in the context of daily-life support?

Since most of the prior studies either focused on species or realism or confounded both, no conclusion of possible interaction effects between both can be made. The stylization of humans might evoke effects that are different from the stylization of other species. Since

robots or talking animals are per se artificial and fictional, a stylization of humans might make more of a difference regarding the evaluation.

Prior studies mostly used humanoid agents (c.f. McDonnell et al., 2012; Ring et al., 2014; van Wissen et al., 2016), while Sträfling et al. (Sträfling et al., 2010) showed that a cartoon-like rabbit evoked more friendliness. Therefore, the beneficial effect of a cartoon stylization seems to not necessarily be limited to humanoid agents. However, it has to be mentioned that so far, to the authors' knowledge, no research exists where the effect of different realism variables for varying species has been investigated systematically. As Sträfling et al. (Sträfling et al., 2010) only compared a cartoon-like rabbit to a more realistic humanoid agent, it is not derivable whether the positive effect on friendliness is based on the cartoon-like stylization or the species of the agent. Therefore, studies investigating the effect of realism combined with the agent's species are needed.

To fill this research gap, the two studies of the current research aim to give the first insights on these potential interaction effects, by answering this research question:

Research Question 3 (RQ3): Are there any interaction effects of species and realism with regard to the person perception of the agent, users' liking, and usage intention?

7.2 STUDY A

7.2.1 Outline and Deviation of Hypotheses

Since various application fields result in different target groups and those target groups might differ in their preferences and needs, Study A focuses on target group-related differences (namely, the preferences of seniors and cognitively impaired people in comparison to students) in the perception and evaluation of appearance variables. Modern technologies give great opportunities to support elderly or cognitively impaired people in their daily life and to maintain their autonomous living as long as possible (Kopp et al., 2018). Virtual assistance systems are, for example, able to remind people, help them to structure their daily life, or support them with making appointments. It has already been demonstrated that people in need of support accept virtual assistants (Yaghoubzadeh et al., 2013). These findings are mainly focused on elderly people, but might also be valuable for other target groups, such as cognitively impaired people, as they have similar needs of support. Vardoulakis et al. (Vardoulakis, Ring, Barry, Sidner, & Bickmore, 2012) designed a social relational agent for elderly people (based on a Wizard of Oz setting) in order to care for their mental health and investigate different topics that users may like to talk about with a conversational virtual agent. After talking for one week with the agent on a daily basis,

participants stated high levels of companionship, support, and satisfaction and felt comfortable having a virtual agent to talk to at their homes. Moreover, it has been shown that even if interaction problems occur during interaction with a speech-based conversional agent, elderly people and people with cognitive impairments state good levels of acceptance (Yaghoubzadeh et al., 2013).

In summary, an embodied virtual agent is assumed to enhance the daily-life assistance of elderly people and people with special needs. However, those findings are based on interactions with agents that have one specific appearance and thus cannot be generalized for all agents. Since appearance was found to affect the perception and evaluation of a virtual agent in multiple ways, it needs to be investigated how these target groups perceive and evaluate different appearance variables. Thus, what remains open is the question of how a virtual agent for these target groups should look. There are some insights with regard to e-commerce. The results of focus groups with 25 elderly people indicate that abstract agents were preferred, because they are less distracting than humans or humanoid agents (Chattaraman et al., 2011). Although these results provide initial insights, it is unclear whether they are transferrable to other applications and tasks. Since in the context of e-commerce, the presented products need to be in the user's focus, a humanoid agent might be distracting; however, in the context of personal assistance, a humanoid agent could be perceived as more serious and engaging. A qualitative interview study in the context of daily-life assistance gave first hints that age-related differences with regard to appearance preferences exist (Straßmann & Krämer, 2017). In contrast to the findings of Chattaraman et al. (Chattaraman et al., 2011), seniors stated in these interviews that they would prefer a more realistic humanoid agent, while students rejected this kind of appearance (Straßmann & Krämer, 2017). It seems as if seniors, who are mostly less experienced with technologies, strive for higher familiarity than students in order to foster their trust in the agent and to remove boundaries of technology usage. However, both studies used qualitative methods and therefore their results are difficult to generalize. The current study aims to investigate these target groups and their preferences in more detail. Based on the findings of Straßmann and Krämer (Straßmann & Krämer, 2017), we hypothesize the following:

Hypothesis 1 (H1): Seniors evaluate humanoid agents more positively compared to nonhumanoid agents (animals and robots).

Hypothesis 2 (H2): Students evaluate nonhumanoid agents (animals and robots) more positively than humanoid ones.

Hypothesis 3 (H3): Seniors and students evaluate varying degrees of realism differently.

Since Yaghoubzadeh et al. (Yaghoubzadeh et al., 2013) showed that virtual assistant is also accepted by cognitively impaired people and as this technology might be highly beneficial for these people, it is important to investigate their special needs. To the authors' knowledge, there are no findings reported to date about the preferences of people with cognitive impairments regarding appearance variables. Nonetheless, it has been shown that people with intellectual disabilities preferred more simple visual representations of hyperlinks to browse more easily through the internet (Rocha, Bessa, Bastardo, & Magalhães, 2018). Consequently, it can be assumed that people with cognitive impairments do also prefer a more simple visual appearance (e.g., a cartoon stylization or reduced detail) to avoid distraction. At the same time, the same principles that are assumed for seniors might be applied, in the sense that this target group requires higher familiarity to enhance acceptance and trust. The current study explores the perception of appearance variables by people with cognitive impairments for the first time and intends to answer the following research question:

Research Question 4 (RQ4): How do cognitively impaired people perceive and evaluate different species and degrees of realism?

7.2.2 Method

The first study aims at investigating the effect of the target group on the evaluation of different appearance variables. Therefore, three different target groups (students, seniors, and cognitively impaired people) participated in this study. A mixed design with one between-subjects factor (3 target groups) and two within-factors (3 species × 5 realism) was used and all participants evaluated 30 different virtual agents, which differed mainly in regard to species and realism (see *Figure 9*). The order of the presentation was randomized to prevent sequence effects. The questionnaire was programmed with the tool SoSciSurvey (www.soscisurvey.de). Participants were invited to the lab, since not all target groups have internet access and to avoid high drop-out rates. The local ethics committee approved the study.

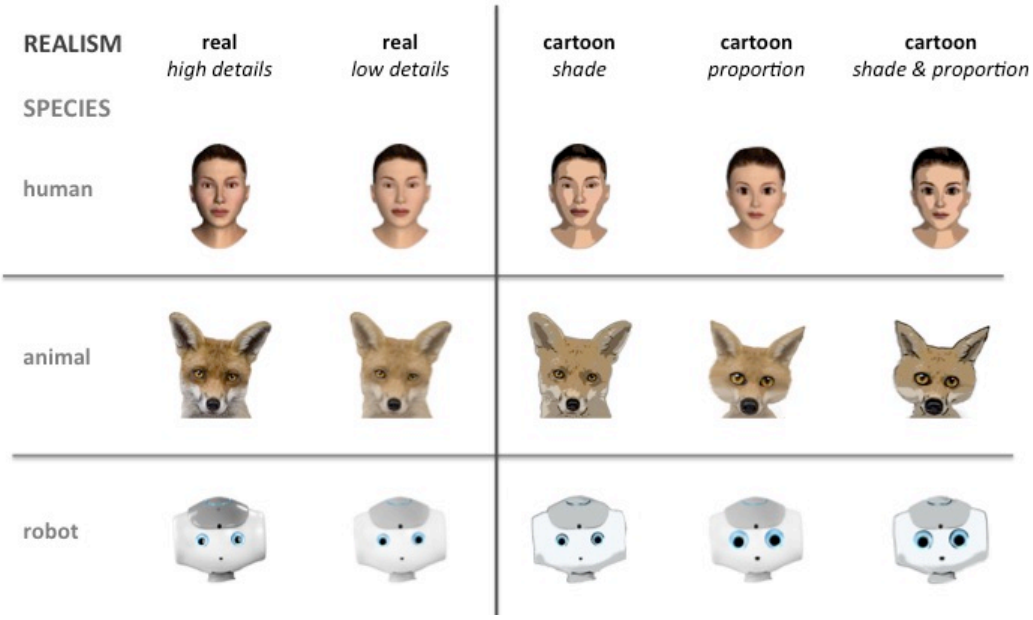


Figure 9. Examples of the used stimulus material

Sample

Overall, 59 participants from three different target groups completed the questionnaire; 22 students (12 female; age: $M = 21.45$, $SD = 3.99$), 21 seniors (14 female; age: $M = 68.14$, $SD = 8.42$), and 16 people with cognitive impairments (3 female; age: $M = 45.81$, $SD = 19.33$) took part. Half of the participants (31 participants: 11 students, 12 seniors, and 8 cognitively impaired people) knew what a virtual agent is, but only 13 persons (8 students, 4 seniors, and 1 cognitively impaired person) had interacted with one before. The students' data were collected at a large European university, while the cognitively impaired people were recruited in a European Health Care foundation and seniors participated partly in both places. Students were incentivized by course credits, while elderly and cognitively impaired people received 7 Euro as a compensation for their participation. All participants classified as seniors needed to be at least 65 years old. For the cognitively impaired sample, no age restriction was used, but participants were all associated to a psychosocial care facility and were chosen by skilled staff of the Health Care Foundation. No concrete diagnoses were provided for these participants, as on the one hand, this is often very difficult to derive even for skilled persons, and on the other hand, we wanted to explore the preferences of this generally heterogeneous group. Since a virtual assistant should be useful and effective for people with all kinds of impairments and needs of support, the goal was not to determine the preferences of people with a specific impairment. During the recruitment process, it was

checked that those people have a certain degree of cognitive impairments, but that the participants, on the other hand, were able to express their own opinion correctly.

Stimulus Material

As the stimulus material, 30 different static pictures were used. All pictures had the same size and showed the virtual characters' head, which varied in species and realism.

With regard to species, humans, animals, and robots were chosen. Since even within this subcategory, the range of possible appearances is enormous, a pretest with a total of 18 agents (6 for each species) was conducted first. Overall, 24 people (18 female, age: M = 32.46, SD = 12.12) who were not participants of the two main studies evaluated these agents in a between-subjects design on single items measuring perceived realism, likability of the agent, and willingness to interact. To guarantee generalizability of the results, two agents for each species have been chosen (for example pictures, see Figure 2) based on their likability scores from the pretest. Therefore, to ensure that obtainable differences in the main studies would be caused by the manipulation only, stimuli should not differ in perceived likability, realism, and willingness to interact. In addition, those stimuli with the highest likability scores were chosen. In the end, a woman, a man (both created with Autodesk's character builder), a fox, a giraffe, the Nao robot, and a more anthropomorphic robot were used. However, the results of the pretest demonstrated that the agents within the species were not evaluated radically differently. Therefore, both versions of the species were collapsed for the calculations.

Realism was manipulated in 5 different degrees based on prior research (Gulz & Haake, 2006b; Ring et al., 2014). As mentioned above, multiple subcategories influence the overall perception of realism (Straßmann & Krämer, 2017) and therefore are assumed to influence the overall perception of the agent. To investigate these effects in more detail, two subcategories were manipulated: degree of detail and stylization. Thus, the manipulation can be divided into realistic appearances and stylized appearances with a more cartoon-like look. While the realistic appearances are further divisible into high detail and low detail (comparable to the concepts of Gulz and Haake (Gulz & Haake, 2006b)), the stylization was either applied to the proportions, to the shading, or both (see (Ring et al., 2014)). As realism and its subcategories can be seen as a continuum ranging from high realism to low realism, the created stimuli can be sorted along this continuum: (1) realistic high detail, (2) realistic low detail, (3) stylized shade, (4) stylized proportion, and (5) stylized shade and proportions. With decreasing detail and increasing stylization, the perceived realism of the agent is

assumed to decrease as well. The (based on the pretest) chosen appearance pictures all scored medium–high on their resolution. In addition, the appearances had a high level of detail (1), since one can obtain details such as skin or fur properties (for the humanoid agents, wrinkles, rashes, and freckles are displayed, while for the machinelike agents, reflections were obtainable and the fur of the animals was pictured in great detail). To create an appearance with a lower degree of detail (2), those properties were smoothened out. Based on the stimuli with lower detail, the shading (3) was manipulated in a cartoon-like look. Therefore, the outlines were thickened, a soft focus was applied, and the shading itself was manipulated. In order to stylize the agent's proportions (4), facial properties were varied. In dependence of the species, key features such as the eyes and mouth became bigger, while less important parts such as the chin got smaller. Both manipulations (shading and proportions) were applied to create the last degree of realism (5). All manipulations of the degree of realism have been done manually with Photoshop Elements using the same criteria for all species. Figure 2 presents an example of the used stimulus material. In the end, 30 different pictures of agents were created systematically.

Measurements

As a dependent variable, the person's perception was measured repeatedly for each of the 30 stimuli. Five different subscales with an overall 12 items were used: likability, uncanniness, realism, willingness to use, and appeal. Most items have been selected from prior research (Bartneck, Kulić, Croft, & Zoghbi, 2009; McDonnell et al., 2012), and self-constructed items to measure uncanniness were added. With regard to the within-subjects design and the high number of stimuli, we aimed to keep the measurements short. Therefore, realism, willingness to use, and liking were queried using single items. The scale of likability contained 4 items (not attractive (reversed), unlikable (reversed), reliable, and pleasant) and showed good internal consistency (Cronbach's alpha = 0.81). Furthermore, two items (uncanny and negative) were used to measure uncanniness (Cronbach's alpha = 0.84). Participants rated their agreement to these items on a 5-point Likert scale ranging from 1: "strongly disagree" to 5: "strongly agree". Some additional variables were measured, but since these are not in the focus of the study's research aim, they were not reported in this paper. Since a proportion of the sample had cognitive impairments, the questionnaire was adapted to their special needs in order to guarantee that these participants understand what we intended to measure. The language was adapted in a way that the instructions were simplified,

but the items themselves were the same for all three groups to ensure comparability of the results.

Procedure

The experimenter welcomed the participants to the lab and instructed them to fill in the questionnaire at the computer. Before the actual study began, the experimenter informed the participants about the aim and procedure to gain proper informed consent. As done for the measurements, the whole introduction and debriefing material was also adapted to simple language in correspondence with schooled people from the Health Care Foundation. Therefore, proper informed consent was also ensured for people with cognitive impairments. Furthermore, the experimenter assisted the participants (especially those of need in support) whenever it was necessary. However, participants were asked to answer the questionnaire autonomously and informed they should only rely on the experimenter for comprehension problems. After an initial introduction, participants were asked about their prior experiences with virtual agents. Regardless of their answer, a short description of virtual agents was presented to guarantee that all participants had the same definition of a virtual agent in mind. Further on, participants' usage intention was queried. Before the stimulus material was presented, participants were asked to imagine a scenario 5 years in the future, where virtual agents are widely spread. This future scenario was used to ensure that participants were free of thinking about technical restrictions while they were evaluating the presented stimuli. When the main evaluation began, all 30 stimuli were presented in a randomized order. After each stimulus, the dependent variables were assessed. Toward the end, sociodemographic variables were queried. Finally, a debriefing informed the participants about the main research questions.

7.2.3 Results

To investigate the presented hypotheses and research questions, multiple mixed (Analysis of variance) ANOVAs with a three-component between-subjects factor of target group (students, seniors, and cognitively impaired people) and the two within-subject factors of species (three factors: human, animal, and robot) and realism (high detail, low detail, stylized shade, stylized proportions, and stylized shade with stylized proportions) have been calculated. When the assumption of sphericity was violated, the Greenhouse–Geisser correction is reported. Interaction effects between the within-subject factors and the between-subject factors were further analyzed by calculating the repeated ANOVAs again separately

for each target group. In these cases, post-hoc tests using the Bonferroni correction are presented.

Perceived Realism

First, we examined whether the manipulation was successful and investigated the effect of the manipulation on perceived realism. A significant main effect for species occurred; $F(2, 55) = 37.110$, $p < 0.001$, $\eta_p^2 = 0.399$. Post-hoc analyses revealed that all three groups differed significantly from each other. While humans were rated as the most realistic ($M = 3.07$, $SE = 0.11$), animals ($M = 2.77$, $SE = 0.12$) were less realistic and robots ($M = 2.00$, $SE = 0.11$) were least realistic. Further on, a significant main effect for realism emerged; $F(4, 53) = 50.59$, $p < 0.001$, $\eta_p^2 = 0.475$. The post-hoc analyses showed that there was no difference between a realistic agent with high and low detail, while the three stylized agents were significantly rated as less realistic than the realistic agent with high and low detail. Furthermore, there was no difference in perceived realism between an agent with a stylized proportion and an agent with a stylized proportion and shading, but a stylized shading was rated as being significantly more realistic compared to both the other stylized agents (see Table 7 for means and standard error).

However, no main effect for target group could be found, indicating that ratings from students, seniors, and cognitively impaired people were, in general, the same regarding the perception of realism.

Table 7
Overall perceived realism ratings for degrees of realism

Appearances	*M*	*SE*
Realistic high detail	$3.09_{a,b,c}$	0.10
Realistic low detail	$2.98_{d,e,f}$	0.08
Cartoon proportions	$2.51_{a,d,g}$	0.11
Cartoon shades	$2.28_{b,e}$	0.11
Cartoon proportions and shades	$2.22_{c,f,g}$	0.11

Note: Means in columns sharing subscripts are significantly different from each other.

Likability

With regard to likability, no main effect of species was found, while a main effect of realism occurred; $F(2.555, 153.093) = 9.861$, $p < 0.001$, $\eta_p^2 = 0.150$. Post-hoc analyses revealed that agents with both nonstylized degrees of realism were perceived as significantly more likable than agents with stylized proportions or agents whose shade and proportions were stylized (see Table 8).

Overall, there was no significant difference in the target groups' likability evaluation. However, significant interaction effects of the species and target group ($F(3.962, 110.938) = 11.151$, $p < 0.001$, $\eta_p^2 = 0.285$) as well as of realism and target group ($F(5.110, 143.093) = 16.881$, $p < 0.001$, $\eta_p^2 = 0.376$) were found. Post-hoc analyses revealed the following patterns: Students perceived no differences between the three species in likability, while seniors and cognitively impaired people did. Seniors rated the robot ($M = 2.16$, $SE = 0.12$) as significantly less likable than humans ($M = 3.25$, $SE = 0.16$) and animals ($M = 2.94$, $SE = 0.18$). In contrast, cognitively impaired people evaluated robots ($M = 3.53$, $SE = 0.22$) as being more likable than animals ($M = 2.77$, $SE = 0.16$), while they perceived no differences concerning likability between robots and humans ($M = 3.09$, $SE = 0.24$) or humans and animals. With regard to realism, students rated agents with both nonstylized degrees of realism as being significantly more likable than the three stylized versions, while seniors only saw differences in likability between the realistic version with low detail and the stylized proportions. Cognitively impaired people perceived agents with all degrees of realism as being equally likable. (Consult Table 8 for means and post-hoc analyses.)

Table 8 *Likability ratings for agents' degrees of realism of all target groups*

Target Group	Realistic High Detail		Realistic Low Detail		Cartoon Proportions		Cartoon Shades		Cartoon Proportions and Shades	
	M	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
Students	$3.28_{a,b,c}$	0.09	$3.20_{d,e,f}$	0.1	$2.52_{a,d}$	0.1	$2.67_{b,e}$	0.09	$2.31_{c,f}$	0.1
Seniors	2.96	0.1	2.95_{a}	0.11	2.62_{a}	0.14	2.77	0.12	2.62	0.13
Cognitively impaired	2.94	0.14	2.93	0.16	3.28	0.21	3.18	0.21	3.33	0.24
Overall	3.06_{a}	0.06	3.02_{b}	0.07	2.88	0.08	$2.81_{a,b}$	0.08	$2.76_{a,b}$	0.09

Note: Means in rows sharing subscripts are significantly different from each other.

Uncanniness

Further calculations showed significant main effects of species ($F(2, 112) = 3.735$, $p = 0.027$, $\eta_p^2 = 0.063$) and realism ($F(2.746, 153.767) = 252.843$, $p < 0.001$, $\eta_p^2 = 0.820$) on perceived uncanniness. Post-hoc analysis demonstrated that animals ($M = 2.12$, $SE = 0.09$) were evaluated as being significantly less uncanny compared to robots ($M = 2.38$, $SE = 0.09$), while humans ($M = 2.26$, $SE = 0.08$) did not differ significantly from animals or robots. Further on, agents with a stylized shade (also true for the combination with a stylized proportion) were perceived as being less uncanny compared to those with a realistic shading (high detail, low detail, and stylized proportions) (see Table 9).

Again, no main effect of target group emerged as being significant. However, significant interaction effects of the species and target group ($F(3.959, 110.853) = 7.792$, $p < 0.001$, $\eta_p^2 = 0.218$) as well as of realism and target group ($F(5.492, 153.767) = 11.292$, $p < 0.001$, $\eta_p^2 = 0.287$) were found. Students again perceived no differences in uncanniness between the species. The same pattern that was found for likability with regard to robots occurred for seniors and cognitively impaired people: seniors perceived robots ($M = 2.64$, $SE = 0.14$) as being more uncanny than humans ($M = 1.90$, $SE = 0.13$) and animals ($M = 2.01$, $SE = 0.14$), while cognitively impaired people evaluated robots ($M = 2.13$, $SE = 0.23$) as being least uncanny (humans: $M = 2.39$, $SE = 0.20$; animals: $M = 2.71$, $SE = 0.16$). Moreover, cognitively impaired people evaluated appearances without any stylization as being more uncanny than students and seniors did (Table 9).

Table 9
Uncanniness ratings for agents' degrees of realism of all target groups

Target Group	Realistic High Detail		Realistic Low Detail		Cartoon Proportions		Cartoon Shades		Cartoon Proportion and Shades	
	M	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
Students	$2.24_{a,b,c}$	0.12	$2.45_{d,e,f}$	0.15	$3.06_{a,d,g,h}$	0.15	$1.46_{b,e,g}$	0.06	$1.60_{c,f,h}$	0.07
Seniors	$2.67_{a,b}$	0.13	$2.58_{c,d}$	0.14	$2.79_{e,f}$	0.13	$1.41_{a,c,e}$	0.07	$1.47_{b,d,f}$	0.08
Cognitively impaired	$3.18_{a,b}$	0.22	$3.12_{c,d}$	0.23	2.86	0.24	$1.47_{a,c,e}$	0.12	$1.42_{b,d,f}$	0.12
Overall	$2.70_{a,b}$	0.09	$2.72_{c,d}$	0.1	$1.45_{a,c,e}$	0.05	$2.90_{e,f}$	0.1	$1.50_{b,d,f}$	0.05

Note: Means in rows sharing subscripts are significantly different from each other.

Liking of the Agent

Further on, we found significant differences between the species with regard to the participants' liking of the agent ($F(2, 112) = 5.232$, $p = 0.007$, $\eta_p^2 = 0.085$). Post-hoc analyses revealed that humans ($M = 2.68$, $SE = 0.11$) were more liked than robots ($M = 2.25$, $SE = 0.11$), while animals ($M = 2.50$, $SE = 0.13$) did not differ from robots or humans. The degree of realism also affected the participants' evaluation of liking ($F(2.957, 165.615) = 19.350$, $p < 0.001$, $\eta_p^2 = 0.257$). The lower the degree of realism, the less people liked the appearance. Agents with a realistic rendering (high and low detail) were significantly more liked than appearances with a cartoon stylization (all three conditions), while an appearance with cartoon-stylized proportions evoked significantly higher ratings of liking than the combination of a stylized shading and stylized proportions (see Table 10).

Table 10
Liking of the agent for degrees of realism from all target groups.

Target Group	Realistic High Detail		Realistic Low Detail		Cartoon Proportions		Cartoon Shades		Cartoon Proportion and Shades	
	M	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
Students	$2.92_{a,b,c}$	0.14	$2.74_{d,e}$	0.17	$2.12_{a,d}$	0.14	$2.37_{b,f}$	0.12	$1.97_{c,e,f}$	0.12
Seniors	2.22	0.16	2.12	0.16	1.9	0.17	2.07	0.15	1.94	0.16
Cognitively impaired	3.04	0.2	3.08	0.2	2.99	0.23	2.83	0.19	2.77	0.25
Overall	$2.73_{a,b,c}$	0.1	$2.65_{d,e,f}$	0.1	$2.43_{a,d,g}$	0.08	$2.34_{b,e}$	0.1	$2.23_{c,f,g}$	0.1

Note: Means in rows sharing subscripts are significantly different from each other.

Target groups did not differ in their evaluation of liking. However, again, interaction effects were obtained for target group and species ($F(3.809, 106.662) = 4.980$, $p < 0.001$, $\eta_p^2 = 0.151$) as well as for realism ($F(5.815, 165.615) = 4.964$, $p < 0.001$, $\eta_p^2 = 0.151$). Again, students' liking of the species was not significantly different, and seniors liked robots ($M = 1.52$, $SE = 0.12$) significantly less than humans ($M = 2.37$, $SE = 0.26$) and animals ($M = 2.26$, $SE = 0.21$). In contrast to prior findings, cognitively impaired people liked animals ($M = 3.38$, $SE = 0.17$) more than robots ($M = 2.64$, $SE = 0.25$), while there was no significant difference between robots and humans ($M = 2.81$, $SE = 0.24$) or humans and animals. Furthermore, for

seniors and cognitively impaired people, the degree of realism had no effect on whether they liked the agent. Students liked more realistic agents more than ones with stylized proportions and proportions and shade (see Table 10).

Willingness to Use

Moreover, the participants' willingness to use the agent was examined. The mixed ANOVA showed significant main effects for species ($F(2, 112) = 3.798$, $p = 0.025$, $\eta_p^2 = 0.064$) and realism ($F(2.916, 163.274) = 16.062$, $p < 0.001$, $\eta_p^2 = 0.223$). Post-hoc tests showed that participants were more likely to use a humanoid agent ($M = 2.62$, $SE = 0.11$) compared to a machinelike one ($M = 2.27$, $SE = 0.10$). Similar to participants' liking behavior, the usage intention decreased with the degree of realism. Post-hoc tests revealed that appearances with a realistic stylization evoked the highest usage intention values, while participants stated that they were least likely to use an agent with stylized shading combined with stylized proportions (see Table 11).

Table 11
Participant's usage intention of agents for degrees of realism from all target groups

Target Group	Realistic High Detail		Realistic Low Detail		Cartoon Proportions		Cartoon Shades		Cartoon Proportion and Shades	
	M	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
Students	$2.69_{a,b,c}$	0.15	$2.51_{d,e,f}$	0.16	$2.01_{a,d}$	0.13	$2.06_{b,e,g}$	0.12	$1.80_{c,f,g}$	0.11
Seniors	2.06	0.16	2.1	0.16	1.91	0.17	2.02	0.17	1.93	0.16
Cognitively impaired	3.19	0.21	3.19	0.24	2.99	0.22	3.08	0.21	2.97	0.24
Overall	$2.65_{a,b,c}$	0.1	$2.60_{d,e,f}$	0.11	$2.39_{a,d,g}$	0.1	$2.30_{b,e}$	0.1	$2.23_{c,f,g}$	0.1

Note: Means in rows sharing subscripts are significantly different from each other.

Further on, a significant main effect of target group was found ($F(2, 56) = 735.755$, $p < 0.001$, $\eta_p^2 = 0.929$). Cognitively impaired people ($M = 3.08$, $SE = 0.17$) stated higher values of usage intention compared to seniors ($M = 2.01$, $SE = 0.15$) and students ($M = 2.21$, $SE = 0.15$). Additionally, interaction effects of the manipulations and the target groups were found (species: $F(3.73, 104.54) = 4.047$, $p = 0.005$, $\eta_p^2 = 0.126$; realism: $F(5.83, 163.27) = 4.913$, p

< 0.001, $\eta_p^2 = 0.149$). The same patterns that were shown with regard to liking behavior occurred for usage intention: for students, the species had no effect on their usage intention, while seniors liked to use robots (*M* = 1.52, *SE* = 0.13) significantly less than humans and animals (humans: *M* = 2.23, *SE* = 0.27; animals: *M* = 2.27, *SE* = 0.21) and cognitively impaired people showed higher usage intention for animals (*M* = 3.38, *SE* = 0.18) compared to robots (*M* = 2.88, *SE* = 0.23). With regard to realism, students stated a higher usage intention for agents with a realistic stylization, while seniors and cognitively impaired people showed no differences with regard to realism in their usage intention (see Table 11).

7.2.4 Interim Conclusions

In line with prior research, the presented findings give hints that species and realism affect the agent's evaluation and user's usage intention of the agent. While there was no main effect of species on the perceived likability, robots were evaluated as being more uncanny compared to animals. However, humanoid agents were more liked and evoked higher usage intention compared to machinelike agents. Although appearances with stylized shading were perceived as being less uncanny than those with realistic shading, a higher degree of realism was rated as being more likable than a stylized appearance. Overall, the liking of the agent and user's usage intention decrease with the degree of realism.

Results of Study A suggest that overall, there are differences between the target groups. For students, the species of the agent seems to be of lower importance, since they evaluated the agent's person perception of the different species equally. Overall, seniors rejected robots and rated them more negatively than humans and animals. Cognitively impaired people rated robots as being more likable and less uncanny than other species, but liked animals more and stated higher usage intention for animals. In contrast, the degree of realism has mostly no effect on person perception for seniors and cognitively impaired people, while students evaluated a more realistic appearance as being more positive. Thus, results indicate that students seem to rely more on realism, while the species is more important for seniors and cognitively impaired people.

7.3 STUDY B

7.3.1 Outline and Deviation of Hypotheses

Results of Study A give the first insights into the effect of species and realism on an agent's perception and evaluation, especially in regard to target group-related differences. Nevertheless, this study has some limitations that make it difficult to derive generalizable

implications. Due to the different target groups and the high effort of recruiting this group of people, the sample size of Study A was rather small and a within-subjects design was used. Therefore, these characteristics might have influenced the results and produced more differences between stimuli.

Study B therefore employs a between-subjects design in order to complement the findings of Study A. Instead of focusing on differences between specific user groups, the current study mainly investigates the effects of species (*RQ1*), realism (*RQ2*), and their potential interaction effect (*RQ3*) in a more controlled manner with a larger sample. In addition, to further contribute to the question of user characteristics, different moderating variables are taken into account to investigate the influence of personality traits. This is based on research which shows that personality traits have an impact on the evaluation of an artificial entity (Isbister & Nass, 2000).

One user characteristic of interest is the tendency to anthropomorphize, which can be defined as "the tendency to apply human characteristics (i.e., emotions, motivations, and goals) to nonhuman animals, objects, and natural entities" (Neave, Jackson, Saxton, & Hönekopp, 2015) (p. 214). It has already been shown that anthropomorphism tendency is positively related to the perception of uncanniness (Meullemans & Abraham, 2017). However, these insights were found in the context of the perception of readers of fiction, and thus it has to be investigated whether they are also valid for the perception of artificial entities. However, since both are fictional, the same principles might well be applicable, and it can be therefore assumed that the tendency to anthropomorphize affects the perception of different appearance variables. People who have a higher tendency to anthropomorphize might feel more comfortable to interact with a fictional character, since they are more able to attribute humanoid characteristics to nonhumanoid characters. Therefore, we assume the following:

Hypothesis 5 (H5): A higher tendency to anthropomorphize leads to more liking of nonhumanoid agents.

In the context of human–robot interaction, negative attitudes and anxieties towards these technologies were found to affect the interaction (Nomura, Kanda, Suzuki, & Kato, 2008). We assume that these effects are also transferable to virtual agents. People with negative attitudes or anxieties toward agents might evaluate varying appearance variables differently. For instance, people who are afraid of social harm caused by virtual agents might

be more likely to wish for a clear distinction between the virtual and real world and therefore prefer a higher degree of stylization and more artificial species.

Hypothesis 6 (H6): Negative attitudes and anxieties towards virtual agents moderate the effect of species and realism on liking and usage intention.

7.3.2 Method

For Study B, an online study with a 3 (species: human, animal, and robot) × 5 (degree of realism: high detail, low detail, cartoon proportions, cartoon shade, and cartoon proportions and shade) between-subjects design has been conducted. The questionnaire was accepted for the recruitment of participants by the SoSciSurvey panel (Leiner, 2016). This panel has about 65,000 active members. While the age is almost balanced, most of those people in the panel have a high level of education. The local ethics committee approved the study.

Sample

Overall, 792 people filled in the questionnaire. Sex was not equally balanced, since 304 (39%) men and 471 (59%) women participated. In addition, 17 people (2%) did not want to state their sex. Age ranged from 15 to 80 years, with an average age of 39 years (M = 38.63, SD = 14.77). When participants had interacted with an agent before, most of them had talked to an agent without an embodiment (78%).

Stimulus Material

Appearance was manipulated in the same way as in Study A. In this way, species were varied between humans, animals, and robots (with two variations of each species), which were manipulated in five different degrees of realism (high detail, low detail, stylized shade, stylized proportions, and stylized shade and proportions). Thus, the same 30 pictures (see. Figure 2) were used, but this time, in a between-subjects design. Every participant saw and evaluated only one picture.

Measurements

As dependent variables, participants evaluated the stimulus material regarding person perception, liking, perceived usefulness, usage intention, and trust. Person perception was measured with five different subscales: perceived realism, likability, appeal, trustworthiness, and competence. These were five-point semantic differentials and most of the items originated from the Goodspeed Questionnaire (Bartneck et al., 2009) and the measures of

McDonnell et al. (McDonnell et al., 2012). Perceived realism (e.g., unreal–real or machinelike–humanoid) and likability were measured with six pairs of adjectives, while appeal (e.g., appealing–not appealing or attractive–unattractive), trustworthiness (e.g., trustworthy–not trustworthy or reliable–not reliable), and competence (e.g., competent–incompetent, intelligent–not intelligent) contained four item pairs. Liking was a self-constructed scale with five items (e.g., if I had a personal virtual agent, I would wish that the agent would look exactly like this) to measure how much participants liked the presented appearance. To measure usage intention, perceived usefulness, and trust, a scale from Heerink et al. (Heerink et al., 2010) was transferred to the application of virtual agents. Participants rated the items of those scales on a five-point Likert scale ranging from 1: "strongly disagree" to 5: "strongly agree". All dependent variables showed good or excellent internal consistency (see Table 12).

Table 12
Dependent variables' reliability values

Concept	Number of Items	Cronbach's Alpha
Perceived Realism	6	0.836
Likability	6	0.950
Appeal	4	0.916
Trustworthiness	4	0.810
Competence	4	0.875
Liking	5	0.887
Usage intention	3	0.914
Perceived usefulness	3	0.926
Trust	2	0.920

To measure the users' tendency to anthropomorphize things, a scale with 10 items (e.g., I sometimes wonder if my computer deliberately runs more slowly after I have shouted at it) was used (Neave et al., 2015). From the original scale with 20 items, only those 10 items focusing on the present behavior and feelings were used. The internal consistency was good (Cronbach's alpha = 0.82). The users' attitude toward agents (three items, e.g., I think it's a good idea to use a virtual agent) and anxiety towards agents (four items, e.g., if I should use a virtual agent, I would be afraid to make mistakes with it) originates from Heerink et al. (Heerink et al., 2010). Theses scales showed an acceptable reliability (Cronbach's alpha =

0.84 and 0.78, respectively). Additionally, the scale of negative attitude towards robots (Nomura, Kanda, & Suzuki, 2006) with its three subscales was transferred to the context of virtual agents. Although the Cronbach's alpha values were not sufficient (ranging from 0.59 to 0.69), this scale was used, since it was very relevant for the presented research questions and is a well-established scale in the realm of human–robot interaction. Again, participants used a five-point Likert scale to state their agreement for all moderating variables. Furthermore, participants stated their age, sex, educational background, and prior experiences with virtual agents.

Procedure

In the beginning, participants were welcomed and thanked for their willingness to participate. After this initial introduction, their tendency to anthropomorphize was queried. Thereafter, general questions about virtual agents such as prior experiences or the participant's attitude towards this technology were assessed. Before the picture of the agent was presented, participants were asked to imagine the same future scenario as the one used in Study A. Each participant saw one of the 30 different agents and rated it with regard to the dependent variables. After the participants' sociodemographics were measured, participants were debriefed. At the end, participants had the chance to take part in a lottery.

7.3.3 Results

Person Perception

To investigate the effect of species and realism on person perception, a (multivariate analysis of variance) MANOVA with two between-subject factors (species and realism) on perceived realism, likability, appeal, trustworthiness, and competence has been calculated. Main effects of species on all variables except for competence were found (see Table 13). All three species differed with regard to perceived realism; while animals scored highest, humans reached moderate values and robots were perceived as being least realistic (Table 14). With regard to likability, appeal, and trustworthiness, animals differed significantly from humans and robots: animals were perceived as more likable, more appealing, and more trustworthy than humans and robots (see Table 14).

Table 13
Results of MANOVA with regard to person perception

Dependent Variable	Source	*df*	*F*	*p*	$\eta_{part.}^2$
Perceived Realism	Species	2	88.593	<0.001	0.280
	Realism	4	7.459	<0.001	0.061
	Species X Realism	8	4.611	<0.001	0.075
	Error	456			
Likability	Species	2	18.389	<0.001	0.075
	Realism	4	1.745	0.139	0.015
	Species X Realism	8	1.267	0.259	0.022
	Error	456			
Attractiveness	Species	2	21.383	<0.001	0.086
	Realism	4	3.493	0.008	0.030
	Species X Realism	8	1.349	0.217	0.023
	Error	456			
Trustworthiness	Species	2	9.506	<0.001	0.040
	Realism	4	1.267	0.282	0.011
	Species X Realism	8	1.008	0.429	0.017
	Error	456			
Competence	Species	2	1.978	0.139	0.009
	Realism	4	1.947	0.102	0.017
	Species X Realism	8	1.651	0.108	0.028
	Error	456			

Table 14
Means of person perception for species

Species	Perceived Realism		Likability		Attractiveness		Trust-worthiness		Competence	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Human	2.21_a	0.71	2.54_a	0.86	2.22_a	0.84	2.40_a	0.76	2.91	0.81
Animal	2.93_a	0.95	$3.18_{a,b}$	1.00	$2.90_{a,b}$	0.91	$2.81_{a,b}$	0.09	2.95	0.82
Robot	1.87_a	0.62	2.69_b	1.06	2.35_b	0.97	2.50_b	0.87	3.09	0.81
Overall	2.32	0.89	2.80	1.02	2.49	1.02	2.57	0.87	2.99	0.82

Note: Means in columns sharing subscripts are significantly different from each other.

Table 15
Means of person perception for realism

Degree of Realism	Perceived Realism		Likability		Attractive-ness		Trust-worthiness		Competence	
	M	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
High detail	$2.55_{a,b}$	0.08	2.86	0.10	2.60	0.10	2.58	0.09	2.92	0.08
Low detail	$2.53_{c,d}$	0.07	2.99	0.10	2.74_{a}	0.09	2.72	0.08	3.18	0.08
Cartoon shade	$2.20_{a,c}$	0.08	2.71	0.11	2.37	0.11	2.51	0.09	2.90	0.09
Cartoon proportion	2.30	0.08	2.68	0.10	2.41	0.10	2.47	0.09	2.95	0.09
Cartoon shade and proportion	$2.09_{b,d}$	0.08	2.74	0.10	2.30_{a}	0.10	2.55	0.09	2.97	0.08

Note: Means in columns sharing subscripts are significantly different from each other.

In addition, a significant main effect of realism occurred for perceived realism ($F(4, 456) = 7.459$, $p < 0.001$, $\eta_p^2 = 0.061$) and appeal ($F(4, 456) = 3.493$, $p = 0.008$, $\eta_p^2 = 0.030$). Post-hoc analyses with Bonferroni correction revealed that the stylization of the shade leads to lower perceived realism, since a stylized shade (also in combination with a stylized proportion) differed significantly from both versions without any stylization (high and low detail) (see Table 15). In addition, the realistic style with lower detail was perceived as being more appealing than the complete cartoon stylization.

Interaction effects of both independent variables were only found for perceived realism (Table 13). Animals were seen as more realistic than humans and robots for most of the realism degrees, but when the cartoon stylization was applied to both proportion and shade, no difference between the species occurred (*Figure 10*).

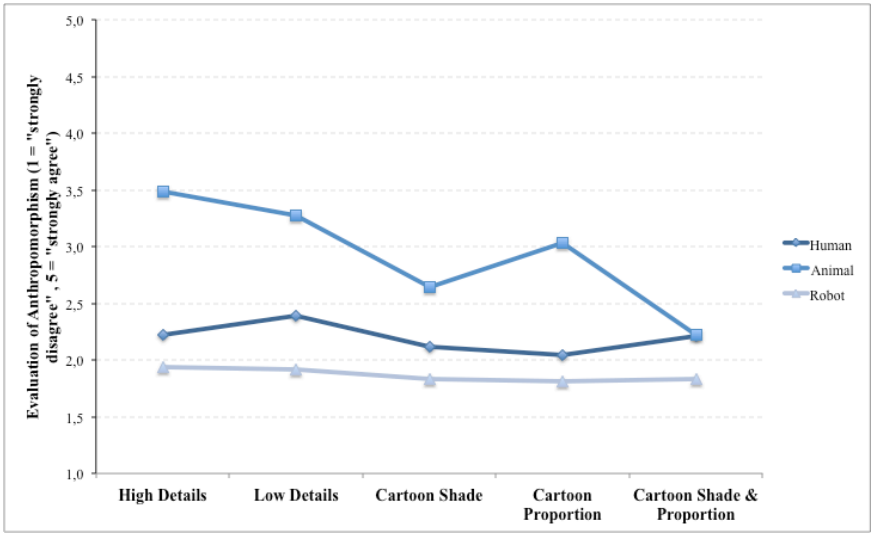


Figure 10. Interaction effect of species and realism on perceived realism

Liking of the Agent

Moreover, a two-factorial ANOVA with species and realism as independent variables and liking as the dependent variable showed a main effect only for species ($F(2, 734) = 7.578$, $p = 0.001$, $\eta_p^2 = 0.020$), indicating that animals ($M = 1.96$, $SD = 0.96$) and robots ($M = 1.94$, $SD = 0.99$) were more liked than humans ($M = 1.68$, $SD = 0.74$). However, no main effect for realism occurred, nor was an interaction effect found.

Usage Intention, Perceived Usefulness, and Trust

A second MANOVA has been calculated with usage intention, perceived usefulness, and trust towards virtual agents as dependent variables and species and realism as independent variables. We found significant main effects of species (usage intention: $F(2, 744) = 1.769$, $p = 0.008$, $\eta_p^2 = 0.013$ and perceived usefulness: $F(2, 744) = 6.899$, $p = 0.001$, $\eta_p^2 = 0.018$), while no effect of realism nor an interaction effect was revealed. Usage intention was higher for robots than for humans and animals, while robots also were perceived as being more useful than animals (Table 16).

Table 16
Participant's usage intention, perceived usefulness, and trust ratings for all agent species

Species	Usage Intention		Perceived Usefulness		Trust	
	M	*SD*	*M*	*SD*	*M*	*SD*
Human	1.92_{a}	1.04	2.31	1.12	1.91	0.93
Animal	1.91_{b}	1.03	2.18_{a}	1.05	1.87	0.94
Robot	$2.16_{a,b}$	1.11	2.53_{a}	1.13	1.97	0.94
Overall	2.00	1.06	2.34	1.10	1.92	0.94

Note: Means in columns sharing subscripts are significantly different from each other.

Moderating Variables

To investigate the effect of user characteristics such as age, sex, anthropomorphic tendency, attitudes toward agents, anxieties toward agents, and the negative attitude toward agents on usage intention and liking of the agent, two hierarchical regression analyses were calculated separately. Predictors were entered in the following order: age and sex (step 1); the tendency to anthropomorphize (step 2); and attitude toward agents, anxieties toward agents, and all three subscales of negative attitude toward agents (step 3). As presented in Table 17, the anthropomorphism tendency and participants´ attitude toward agents significantly explains the variance of usage intention and liking of the agent. The greater the tendency to anthropomorphize or the more positive the participant's attitude toward agents, the higher the participant's usage intention and liking of the agent. Moreover, the anxiety toward agents and negative attitudes toward the social influence of agents significantly contribute to the explanation of the variance of usage intention (Table 17). The usage intention increases with anxiety towards agents and decreases with increasing negative attitudes towards the social influence of agents.

To investigate a possible moderating influence of anthropomorphism tendency, attitudes toward agents, and anxieties toward agents on the effects of both independent variables (species and realism) on participants' liking and usage intention, multiple moderation analyses using the Hayes process were calculated separately. The effects of the potential moderating variables are presented in Figures 4 and 5.

Table 17
Hierarchical multiple regression analyses.

Steps of the Regression Analyses	Usage Intention					Liking of the Agent				
	b (SE_b)	*β*	*t*	*p*	ΔR^2	*b* (SE_b)	*β*	*t*	*p*	ΔR^2
Step 1					.000					.005
Sex	0.01 (0.11)	0.00	0.08	.934		0.09 (0.10)	0.05	0.90	.368	
Age	0.00 (0.00)	0.02	0.40	.690		0.00 (0.00)	0.06	1.10	.271	
Step 2					.037					.038
Anthropomorphism Tendency	0.38 (0.10)	**0.19**	3.72	<0.001		0.35 (0.10)	**0.20**	3.71	<.001	
Step 3					.280					.070
Anxiety toward agents	0.19 (0.08)	**0.16**	2.53	.012		0.15 (0.08)	**0.14**	1.97	.050	
Attitude toward agents	0.43 (0.07)	**0.40**	6.61	<.001		0.17 (0.07)	**0.17**	2.45	.015	
Negative attitude toward situations of interaction with agents	−0.15 (0.09)	−0.11	−1.66	.097		−0.11 (0.09)	−0.09	−1.17	.244	
Negative attitude toward social influence of agents	−0.19 (0.08)	**−0.14**	−2.55	.011		−0.14 (0.08)	−0.12	−1.87	.062	
Negative attitude toward emotions in interaction with agents	−0.11 (0.08)	−0.09	−1.45	.148		−0.05 (0.08)	−0.04	−0.56	.579	
Total R^2					.317					.112
Step 1	$F(2, 358) = 0.08$, $p = 0.924$					$F(2, 353) = 0.84$, $p = 0.431$				
Step 2	$F(3, 357) = 4.68$, $p = 0.003$					$F(3, 352) = 5.18$, $p = 0.002$				
Step 3	$F(8, 352) = 20.46$, $p < 0.001$					$F(8, 347) = 5.48$, $p < 0.001$				

Note: Values in bold indicate significant relationships.

Against our predictions, all 12 moderation analyses revealed no moderation effects of anthropomorphism tendency, attitudes toward agents, and anxieties toward agents in the effect of species or realism on liking and usage intention.

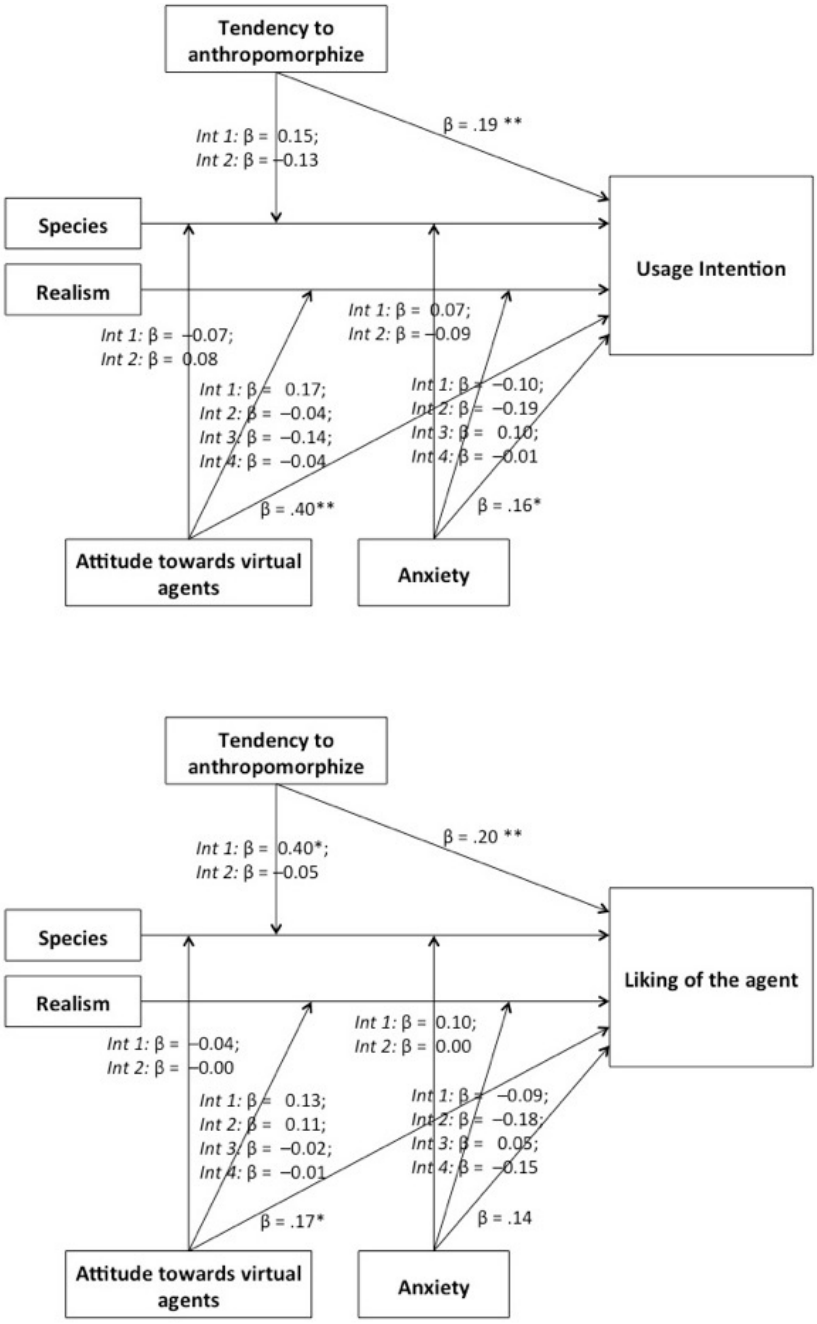


Figure 11. Conceptual models of the moderating variables' effects on usage intention and liking of the agent. Standardized regression coefficients are presented (* indicates a p value < 0.005 and ** <0.001). For the moderation analyses, independent variables were dummy coded (species: Int 1 = human vs. animal, Int 2 = human vs. robot; realism: Int 1 = high detail vs. low detail, Int 2 = high detail vs. cartoon proportions, Int 3 = high detail vs. cartoon shade, Int 4 = high detail vs. cartoon proportions with cartoon shade) and data were standardized before running the analyses.

7.3.4 Interim Conclusion

All three species vary in their perceived degree of realism, in the sense that animals were perceived as most realistic, humans reached moderate perceived realism ratings, and robots were evaluated as least realistic. Additionally, animals were seen as more likable, more appealing, and more trustworthy than humans and robots. While animals and robots were more liked than humanoid agents, participants' usage intention was higher for robots than for animals and humans.

The stylization of the shade leads to lower perceived realism ratings, while a realistic agent with low detail was evaluated as being more appealing than an agent with stylized shades and proportions. The degree of realism had no effect on likability, liking of the agent, or usage intention.

Regression analyses revealed that participants' tendency to anthropomorphize as well as their general attitude towards agents predicts the liking of the agent and participants' usage intention. However, no moderating effects were found for user characteristics.

7.4 DISCUSSION

7.4.1 Results Summary and Interpretation

In the context of human–agent interaction, the appearance of the agents was found to have main effects. Prior research demonstrates that the species (e.g., human or animal) and realism (e.g., cartoon shade) of the agent affect the agent's evaluation (McDonnell et al., 2012; c.f. Ring et al., 2014; Sträfling et al., 2010; Terada et al., 2015) and is therefore important for the outcome of the human–agent interaction. However, until now, both factors had not been examined in a controlled and systematic way. Therefore, we varied three different species (human, animal, and robot) and five degrees of realism (realistic-style high detail, realistic-style low detail, cartoon-stylized shade, cartoon-stylized proportions, and complete cartoon stylization) within two different studies. We investigated how different species (*RQ1*) and degrees of realism (*RQ2*) were evaluated and whether there is an interaction effect of both variables (*RQ3*). Since many applications have special target groups (e.g., people in need of support), we further investigated the preferences of different user groups (*H1, H2, H3*, and *RQ4*) and the influence of further user characteristics such as personality traits (*H4* and *H5*). When the influences of the target group and user characteristics are examined, the appearance can be tailored to the needs of the users. This will enhance the interaction and acceptance of virtual agents. In the following, the results of both studies are summarized and discussed based on prior research.

With regard to species and research question 1 (*RQ1*), both studies present contradicting findings. While Study A showed that users preferred humans to robots and showed higher usage intention for humans, this was not the case in Study B. People seem to prefer humanoid agents in a direct comparison (as in the within-subjects design of Study A), but this is not true when only one agent is presented (as in the between-subjects design of Study B). The between-subjects design showed that animals were perceived as more likable, appealing, and trustworthy than humans and robots. Moreover, participants liked animals and robots more than humans, and those species evoked higher usage intention. As the results of Study B represent a wider sample (no specific target groups and higher sample size) and these results are more generalizable than those of Study A, nonhumanoid agents seem to be evaluated as being more positive than humanoid ones. However, most of the agents employed in current systems are humanoid. Based on our findings, the gold standard of providing a humanoid agent needs to be reconsidered. While for specific target groups such as senior people, humanoid agents might be more appropriate, the results of Study B emphasize that at the same time, a majority were found to like nonhumanoid agents more than humanoid ones.

The investigation of research question 2 (*RQ2*) also led to contradictory findings. Study A demonstrates that participants liked appearances with a higher degree of cartoon stylization less than those with a realistic style. The same pattern was obtained for the usage intention of the participants. Furthermore, agents with a realistic stylization were evaluated as being more likable. However, these findings were not replicated in our second study, since the degree of realism had no effect on likability, liking of the agent, or participants' usage intention. Previous studies mostly showed that lower degrees of realism evoke a more positive perception of the agent (J. Li et al., 2015; McDonnell et al., 2012; Terada et al., 2015). However, there is prior research that supports findings of Study A, since Van Wissen et al. (van Wissen et al., 2016) showed that participants preferred a realistic-looking agent over a cartoon-stylized agent to be their virtual nurse. The most reasonable explanation for these results seems to be the type of application, since Ring et al. (Ring et al., 2014) and also van Wissen et al. (van Wissen et al., 2016) showed that realistic agents are more appropriate for medical tasks. In line with this, Robertson et al. (Robertson et al., 2015) found that people even stated that they get angry about the use of cartoon-stylized agents in medical applications, since it somehow stultifies a very serious task. In our studies, the task itself was not explicitly defined, but participants were asked to imagine the agents as their own personal assistant in everyday life. Thus, this application field might underlie the same principles as a medical task, since both applications aim to provide support and help. Therefore, an

appearance with a higher degree of realism seems to be more appropriate. The discrepancy between both studies might be caused by the different designs, since Study A used a within-factor design, while participants in Study B only evaluated one agent. Realism seems to be more important when users are able to compare different appearances. Since the different realism degrees were manipulated in five small steps (relying on different subcategories), the differences might be much clearer in a direct comparison. While one can easily obtain the agent's species (and the perception associated therewith), the realism of the agent is more complex and multilayered and its degrees are more subtle. Therefore, differences might be stronger in a within-subject design.

Furthermore, the present studies examined the interaction effects of species and realism for the first time in a systematic manner (*RQ3*). However, nearly no interaction effects were found (with the exception of perceived realism). It can be concluded that the stylization of an agent has similar effects on all kinds of species (such as humans, robots, and animals). In line with this, prior research showed that animals with atypical features such as enlarged eyes were rated as less familiar (Schwind et al., 2017), which is also true for humanoid agents with stylized proportions. While in Study A, the effects of realism degrees were bigger (higher effect sizes), in Study B, effects of the agent's species were stronger. Thus, as described above, the degrees of realism seem to be more important in a direct comparison (e.g., when users are able to choose from alternatives), while the species was more decisive in a between-subjects design. When people only see one agent, the species of the agent is more decisive for the impression-building process.

This was based on initial hints that indicated that there are age-related differences with regard to appearance preferences (Straßmann & Krämer, 2017). Based on these findings, we assumed that seniors evaluate humanoid agents more positively than nonhumanoid ones (*H1*), while students might rate nonhumanoid agents more positively (*H2*). This assumption was found partly to be true with regard to the evaluation of robots. While students liked robots the most and showed highest values of usage intention for robots, seniors clearly rejected machinelike agents. These findings are in line with the results of prior qualitative research (Straßmann & Krämer, 2017). In interviews, seniors stated that they would prefer a humanoid agent, since the interaction with it is more familiar as they talk to humans all the time, while animals or other species cannot answer appropriately (Straßmann & Krämer, 2017). However, these results cannot be generalized to all people in need of support, since cognitively impaired people evaluated robots more positively and stated higher usage intention and liking for animals; this group seems to prefer nonhumanoid agents more. While for seniors, a high

familiarity seems to be of importance and only a humanoid appearance is appropriate as their assistant, cognitively impaired people are less restricted to humans. The cognitively impaired people are probably used to receiving assistance in their daily life, while seniors mostly were more afraid to eventually be in need of support. Therefore, for seniors, it might be more important to have something familiar and serious as a humanoid agent, while cognitively impaired people are a little more open-minded in the sense that they are willing to accept assistance from different species. Differences between those target groups were also found in relation to realism. Thus, hypothesis *H3* was confirmed. However, in contrast to the findings of Straßmann and Krämer (Straßmann & Krämer, 2017), seniors evaluated a cartoon stylization more positively, while students rated realistic agents more positively. The overall patterns of Study A indicate that for people in need of support (seniors and cognitively impaired people), the species is more important for their evaluation, liking, and intention to use. In contrast, for students, the degree of realism was more important than it was for seniors and cognitively impaired people. Results emphasize that there is no universal appearance factor that is appropriate for various user groups.

Overall, no moderating effects of personality or attitudes were found. Therefore, hypothesis *H4* and *H5* need to be rejected. Nevertheless, anthropomorphism tendency and participants' attitude toward virtual agents predict the liking of the agent and usage intention. While those variables were found to not moderate the effects of appearance variables, they influence whether people are willing to interact with a virtual agent and whether they like it. People with a high tendency to anthropomorphize and a more positive attitude toward virtual agents liked the agent more. Furthermore, the attitude and anxiety towards virtual agents affect the willingness to use it. Thus, to enhance the interaction with virtual agents, people´s attitudes need to be improved. Although we found no moderating effects with regard to appearance, future research should investigate whether appearance variables can have a positive effect on people's attitudes or decrease anxieties toward agents in general.

7.4.2 Limitations and Future Work

For both studies, several limitations have to be taken into account. One major limitation is caused by the presentation of pictures only. Since participants evaluated the agent based on static pictures, only limited generalizations to user behavior in real human–agent interactions can be made. Although the used method was also beneficial, since participants could not get distracted by the agent's behavior or interaction characteristics and therefore evaluated appearance factors only, studies with agents that interact with the users are needed.

Thus, future studies should investigate whether the presented findings are transferable to user behavior in real interactions and whether there is still an influence of appearance on perceptions.

Even though the stimulus material was prepared in a more systematic way than in previous studies and was based on a pretest, the stimuli are still characterized by other appearance variables beside species and realism (e.g., styling of humanoid agents, chosen animals, or colors). To minimize the effect of those variables, two versions for each species have been chosen and evaluations have been collapsed, which leads to a higher generalizability of the findings. In addition, as mentioned above, the degrees of realism that were manipulated were relatively similar. Although the very systematic manipulation of the appearance factors is a big strength of the present research, with regard to our findings, it might also be a limitation. As only differences in realism in a within-subjects design have been found, it has to be mentioned that differences are very subtle and might only be obtainable in a direct comparison. Despite the fact that the participants were instructed to imagine a future scenario in order to reduce mental restrictions, e.g., with regard to technical implementations, participants might still have been limited in their imagination.

Study A aimed to investigate the preferences of different target groups and especially focused on people in need of support. Therefore, elderly and cognitively impaired people participated. It has to be noticed that some participants of the cognitive impairment condition could also be seen as seniors, since their age was over 65. However, the distinction has been made based on participants' cognitive skills, which were evaluated by schooled employees of an established Health Care Foundation. Since people with different sorts of impairments—which had not been specified, except for the fact that those people had low cognitive skills—participated, findings are limited and could not be generalized to specific impairment groups (e.g., autistic people). Since Study A indicates that people with cognitive impairments, in general, have specific preferences, future studies could focus on specific impairments to obtain deeper insights into the preferences and perceptions of this target group. In addition, the sample size of the cognitively impaired subsample was relatively low, since the recruitment of these participants is not easy. Therefore, the findings of this study can rather be seen as initial hints. However, since virtual assistance is highly beneficial for those users, our research is a valuable start in the investigation of the needs of this target group.

The within-subjects design of Study A also has to be reflected critically, since the evaluation of all 30 pictures took almost one hour for some of the participants. Especially for people with cognitive impairments, the participation most probably was very exhausting,

although the language of the questionnaire was adapted to the participants' skills to lower the cognitive effort. In order to avoid fatigue, the experimenter helped the participants and gave them a break whenever it was needed. However, it might be beneficial to replicate the effects of the target group in a between-subjects design.

In the present research, two studies have been conceptualized to supplement each other, but the results of both studies are contradictory rather than supporting. Possible explanations have been addressed in the discussion part, but nevertheless, further studies are needed to clarify these contradicting findings. Although such contradictions might decrease comprehensiveness of the implications, with regard to open science and the need for replications of scientific findings, contradicting results might occur more often. Overall, these results highlight that multiple (even more than two) studies are always needed to achieve reliable insights.

7.4.3 Conclusions

This research stresses the importance of different appearance variables, such as species and realism. Both factors were found to influence the users' person perception and preferences with regard to usage intention and perceived usefulness. Effect sizes revealed that in a direct comparison of different appearance alternatives, the realism seems to be more important. However, for the impression perception of one agent without any comparisons, the agent's species is more decisive. Results emphasize that a majority evaluated nonhumanoid agents more positively and therefore might be highly beneficial in human–agent interaction. Although animals or robots evoked a more positive perception in general when a wider sample was gathered, they might not be suitable for all kinds of user groups, as seniors were found to clearly reject robots as their personal assistant. These findings help to understand the preferences of specific target groups more deeply and therefore will be beneficial when designing appropriate appearances of daily-life assistants.

8 STUDY 3: LAB EXAMINATION OF SOCIAL EFFECTS OF DIFFERENT VIRTUAL AGENT'S APPEARANCES IN A HUMAN-AGENT INTERACTION

Since the first two studies investigated the influence of appearance variables on the perception and evaluation of the agent solely based on static stimuli (or pure imagination) and the mere perception of the appearance, these effects need to be examined in an actual human–agent interaction. Consequently, Study 3 aims to transfer the prior findings to a real human–agent interaction and investigates the effects of species, realism, and embodiment as well as age-related differences.

8.1 BACKGROUND AND HYPOTHESES

On the basis of the results of the qualitative interviews (Study 1), species and realism were extracted to be the variables of most interest. Therefore, based on this assumption, the effect of species and realism was investigated in a systematic manner in Study 2. However, it remained an open question whether these findings are also transferable to a real human–agent interaction. Thus, Study 3 was conducted to examine the effects of species and realism on the perception and evaluation as well as on social outcomes such as trust and bonding when users interact with a virtual agent. Since modern assistance systems are mostly non-embodied, the effect of an agent's embodiment should be tested further in a human–agent interaction. Because the first two studies demonstrated the importance of taking specific target groups into account, Study 3 aimed to close this research gap. In particular, the investigation of age-related differences in the perception of appearance variables in a real human–agent interaction has been neglected to date.

The aforementioned literature overview (see Chapter II3.2) and the results of the first two studies demonstrated the effect of appearance variables. Since participants in Study 2 evaluated humanoid agents less positively, the following hypotheses are assumed:

Hypothesis 1 (H1): Users will evaluate an agent with a machinelike appearance more positively than one with a humanoid appearance.

H1 a): Users will evaluate the person perception of an agent with a machinelike appearance more positively than one with a humanoid appearance.

H1 b): Users will like an agent with a machinelike appearance more than one with a humanoid appearance.

H1 c): Users will show higher usage intention for an agent with a machinelike appearance compared with one with a humanoid appearance.

There is a large body of research demonstrating that cartoon stylization has a positive effect on the evaluation of the agent (e.g., McDonnell, Breidt, & Bülthoff, 2012; Zell, Aliaga, Jarabo, & Mcdonnell, 2015). By contrast, with regard to medical applications, higher realism was appreciated and perceived as more positive (Tsiourti et al., 2014; van Wissen et al., 2016). These findings are also supported by the results of Study 2. Against this background, the following is hypothesized:

Hypothesis 2 (H2): An appearance with realistic shades and proportions will be evaluated more positively than a cartoon-stylized appearance.

H2 a): An agent with an appearance with realistic shades and proportions will be evaluated more positively regarding its person perception than an agent with cartoon-stylized appearance.

H2 b): An agent with an appearance with realistic shades and proportions will be more liked than an agent with cartoon-stylized appearance.

H2 c): Users will show higher usage intention toward an agent with realistic shades and proportions compared with an agent with cartoon stylization.

An embodied agent has been shown to be beneficial for human–agent interactions, since a meta-analysis by Yee, Bailenson, and Rickertsen (2007) demonstrated that an embodied character evokes a more positive social interaction than a non-embodied one. Thus, the following hypothesis is stated:

Hypothesis 3 (H3): An agent with an embodied character will lead to a more positive evaluation compared with a non-embodied agent.

H3 a): An agent with an embodied character will lead to a more positive evaluation of the agent's person perception compared with a non-embodied agent.

H3 b): Users will like an agent with an embodied character more than a non-embodied agent.

H3 c): An agent with an embodied character will lead to higher usage intention compared with a non-embodied agent.

As seniors are one main target group where virtual assistants are highly beneficial (Kopp et al., 2018; Yaghoubzadeh et al., 2013), the present study investigates age-related differences. Although seniors in general are more skeptical toward the use of technologies such as virtual agents (Scopelliti et al., 2005), they seemed to be less critical when they were

asked after an actual interaction with these technologies. Prior research with seniors in a human–agent interaction showed that seniors evaluated the agent more positively compared with students in general (Rosenthal-von der Pütten, Straßmann, & Krämer, 2018). This might be because they have fewer experiences with such technologies and thus have lower expectations. Thus, seniors' expectations might be outperformed, while students are more used to virtual agents (e.g., from the use of Siri or Alexa) and therefore have higher expectations and are less likely to be impressed by the interaction. Thus, the following can be hypothesized:

Hypothesis 4 (H4): Seniors will evaluate the agent and the interaction therewith as more positive compared with students.

H4 a): Seniors will evaluate the agent's person perception more positively compared with students.

H4 b): Seniors will like the agent more than students do.

H4 c): Seniors will show higher usage intentions than students do.

H4 d): Seniors will state a higher perceived usefulness than students do.

H4 e): Seniors will enjoy the interaction with the agent more than students do.

Nevertheless, as mentioned earlier, younger people are more familiar with virtual agents and their use. This familiarity with the usage of these technologies may most likely enhance the perceived ease of use. Therefore, the following hypothesis can be assumed:

Hypothesis 5 (H5): Students will rate the agent as easier to use than seniors do.

As described, to date, to the author's knowledge, no research exists that has investigated the effect of appearance on the preferences of different age groups in an actual human–agent interaction. Therefore, the current study tested age-related differences in the evaluation and perception of an agent's appearance variables during a human–agent interaction for the first time.

However, there is research that indicates age-related differences exist in the evaluation of an agent's appearance. Chattaraman et al. (2011) conducted qualitative interviews to investigate seniors' preferences of the design of a virtual agent applied in e-commerce contexts. In regard to appearance, the authors report that seniors prefer an abstract appearance, since it is less distracting than a human and even a human who use movements. They further show that their participants preferred an animal appearance over a human and that a humanoid agent is too distracting and that they do not like technical entities to simulate

a human ("I hate humanoid things that are fake human," p. 289, Chattaraman et al., 2011). However, as the authors investigated several parameters of a virtual agent, they did not systematically describe and distinguish between different appearance variables. As it was presented in Chapter II3.4 and Study 2, it is important to distinguish systematically between variables such as species, realism, or feature specifications. With regard to the results of Chattaraman et al. (2011), it is still unclear what is meant by an abstract agent. This can depend on the species (as it was compared to human and humanoid agents) and to the degree of realism as well. Besides these methodological inaccuracies, the findings contradict the statements of the interviews conducted in Study 1.

Here, seniors clearly preferred a humanoid and realistic agent to other species, since they are more familiar with interacting with humans. By contrast, they stated feeling stultified by non-humanoid characters and that cartoon-stylized agents are kids' stuff. Thus, it can be deduced that seniors can only take humanoid agents seriously and therefore will evaluate them more positively. The contradictions between the findings of Chattaraman et al. (2011) and the results of Study 1 (Straßmann & Krämer, 2017) are most likely explainable through the applied context that was used in the studies. It has been demonstrated that the task and application context is highly relevant (Ring et al., 2014) and that in e-commerce the presented product instead of personal communication and assistance is in focus, and therefore users might prefer differently designed agents. Since in the present dissertation the application of virtual assistance is used, the results of Study 1 are better applicable to the human–agent interaction aimed for (where the agent is also applied as virtual assistant). Building on these findings, Study 2 complements the qualitative statements and demonstrates that for seniors the species is more important in the evaluation process of a virtual agent's appearance and that seniors evaluate machinelike agents less positively than humanoid ones. Therefore, the following hypotheses based on the results of Study 1 and 2 are assumed:

Hypothesis 6 (H6): Seniors will evaluate a humanoid agent more positively than an agent with a machinelike appearance.

H6 a): Seniors will evaluate a humanoid agent's person perception more positively compared with an agent with a machinelike appearance.

H6 b): Seniors will like a humanoid agent more than an agent with a machinelike appearance.

H6 c): Seniors will show a higher usage intention for a humanoid agent compared with an agent with a machinelike appearance.

Furthermore, the seniors' interview statements (Study 1) demonstrated that they prefer to interact with an embodied agent, as they like to have someone to address during the communication. Seniors further argued that an interaction with an embodied agent is more interesting and attracts more attention and trust. Overall, seniors' statements stressed that they prefer an embodied agent to a non-embodied one. By contrast, students highlighted the advantages of a non-embodied agent, as non-embodied agents are not restricted to one device or a certain screen. Furthermore, embodied agents are assumed by students to simulate humans and to be potentially more distracting. Therefore, students are expected to evaluate non-embodied agents more positively than embodied agents. Based on these results of Study 1, the following hypotheses can be formulated:

Hypothesis 7 (H7): Seniors will evaluate an embodied agent more positively than a non-embodied agent.

H7 a): Seniors will evaluate an embodied agent's person perception more positively compared with a non-embodied agent.

H7 b): Seniors will like an embodied agent more than a non-embodied agent.

H7 c): Seniors will show a higher usage intention for an embodied agent compared with a non-embodied agent.

Hypothesis 8 (H8): Students will evaluate an agent represented through a voice only more positively compared with an embodied agent.

H8 a): Students will evaluate the person perception of an agent represented through a voice only more positively compared with that of an embodied agent.

H8 b): Students will like an agent represented through a voice only more than an embodied agent.

H8 c): Students will show greater usage intention toward an agent represented through a voice than toward an embodied agent.

Building on the technology acceptance model (Davis et al., 1989) and its adaptation to the acceptance of assistive social agent technologies (Heerink et al., 2010), the perceived usefulness, ease of use, and enjoyment during the interaction has an influence on the user's usage intention. The more useful, easy to use, and enjoying the agent and the interaction with it will be perceived, the higher will be the usage intention. Thus, the following hypothesis can be assumed:

Hypothesis 9 (H9): High ratings of perceived usefulness, ease of use, and perceived enjoyment increase the user's usage intention.

Prior studies demonstrated that users show bonding with virtual agents (Bickmore, Caruso, Clough-Gorr, et al., 2005; Ren, Schulman, Jack, & Bickmore, 2014). In interpersonal relationships and attraction, physical attractiveness and similarity are the key variables. The appearance of an agent, especially its species and realism, can affect the perceived similarity and of course the perceived attractiveness. Therefore, how these appearance variables influence participants' bonding and trust needs to be investigated. In line with this assumption, seniors stated in Study 1 that they felt more trust toward embodied agents and bonded more with humanoid agents. Students did not mention these perceptions and seemed to try to avoid such social processes (see statements of the qualitative interviews [Study 1] where students disliked the fact that humans are simulated and want to clearly distinguish between real and virtual persons). Nevertheless, with regard to the media equation theory (Reeves & Nass, 1996), such process are seen to occur automatically for all human beings. To sum up, whether the appearance of a virtual agent affects social processes such as bonding and trust and whether these are influenced by the users' age should be investigated. Therefore, the following research question is posed:

Research Question 1 (RQ1): How is bonding and trust affected by the agent's appearance and the users' age group?

Prior research demonstrated that affiliation and intimacy are correlated with spending time, for example, having conversations (McAdams & Constantian, 1983). Thus, it can be assumed that greater bonding is associated with a higher willingness to interact with the interlocutor. In the context of human–agent interactions, willingness to interact can also be described as intention to use. With regard to the usage of technical devices, the term *intention to use* is applied (e.g. see TAM (Davis, 1989)); nevertheless, since social processes are ongoing and technical entities can be seen as social actors (Nass et al., 1994), usage intention equals willingness to interact, for example, having conversations. Consequently, the following is assumed:

Hypothesis 10 (H10): Users' bonding with the agent affects users' usage intention positively.

As people bond more to people they like and they trust, and based on the previous hypothesis, bonding is assumed to mediate the effect of liking and trust on the users' usage intention. Thus, the following hypothesis is stated:

Hypothesis 11 (H11): Users' bonding mediates the effect of users' liking of the agent (*a*) and users' trust toward the agent (*b*) on their usage intention of the agent.

Furthermore, the effects of other user characteristics than age should be addressed in this research. In Study 2 it was found that attitude toward virtual agents, the tendency to anthropomorphize, and anxieties toward agents were significant predictors of participants liking the agent and of participants' usage intention (Straßmann & Krämer, 2018), and thus the following research question is posed:

Research Question 2 (RQ2): In which way do user characteristics such as attitude and personality traits influence (a) users' liking of the agent and (b) their usage intention?

8.2 METHOD

This study aims to investigate the effect of appearance and embodiment in a human–agent interaction and further examine the moderating effect of age, since a possible target group might be the elderly in need of support. Therefore, a Wizard-of-Oz study with a 2 (age group) × 4 (appearance) between-subjects design was used.

8.2.1 Sample

Overall, 130 people participated in this study. To investigate age-related differences, two different age groups were invited to participate in this study. In total, 84 students (M = 23.65, SD = 3.84; 18–38 years old) and 46 seniors (M = 70.93, SD = 9.05; 51–89 years old) interacted with a virtual agent and evaluated it afterwards. As intended, both groups significantly differed with regard to age, $F(1, 129) = 1{,}732.19$, $p < .001$, $\eta_{part.}^2 = .931$. Unfortunately, sex was not balanced, with more women (62.3%) than men (37.7 %) participating. However, there were no differences in sex distribution between the two age groups, $\chi^2(1) = 1.59$, $p = .206$. Of the senior group, 14 participants (30.4%) were male and 32 were female (69.6%), while 35 men (41.7%) and 49 women (58.3%) comprised the student group. Students and seniors were further equally distributed in all four experimental conditions, $\chi^2(3) = 0.32$, $p = .957$; Table 18). Nevertheless, both groups differed in their prior experiences with virtual agents, $\chi^2(1) = 31.40$, $p < .001$. Whereas among the students more people had interacted with a virtual agent in the past (n =54; 64.3%), only six seniors (13%) stated having had prior interaction with a virtual agent and, thus, the majority of seniors had no prior experiences with this technique.

Table 18
Distribution of students and seniors among the four experimental conditions

		Billie	Vince	Character	Voice only	Overall
seniors	*N*	12	13	9	12	46
	%	26,1%	28,3%	19,6%	26,1%	100,0%
students	*N*	21	22	20	21	84
	%	25,0%	26,2%	23,8%	25,0%	100,0%
overall	*N*	33	35	29	33	130
	%	25,4%	26,9%	22,3%	25,4%	100,0%

In addition, 18 cognitively impaired people participated in this study. But since this target group is not as focus of the present study, they were not included in the analyses here.

8.2.2 Stimulus material

In the present study, a real human–agent interaction was tested in a Wizard-of-Oz setting. Participants interacted with an agent that was manipulated with regard to its appearance. The behavior and interaction content were held constant between all conditions and participants were randomized into one of the four conditions. Owing to the between-subjects design, participants interacted with one out of four agents with different appearance.

Experimental Manipulation: Appearance

Since this study aims to examine the effect of the agent's appearance, four different appearances were used: two humanoid agents, one machinelike, and one without an embodiment. To investigate the effect of species in an actual interaction study, humanoid appearances and a machinelike appearance was used. Further, the influence of realism was tested, with a cartoonlike human and a more realistic character being used. Additionally, this study aims to investigate the influence of embodiment, in comparing embodied character with a voice-only version of the agent.

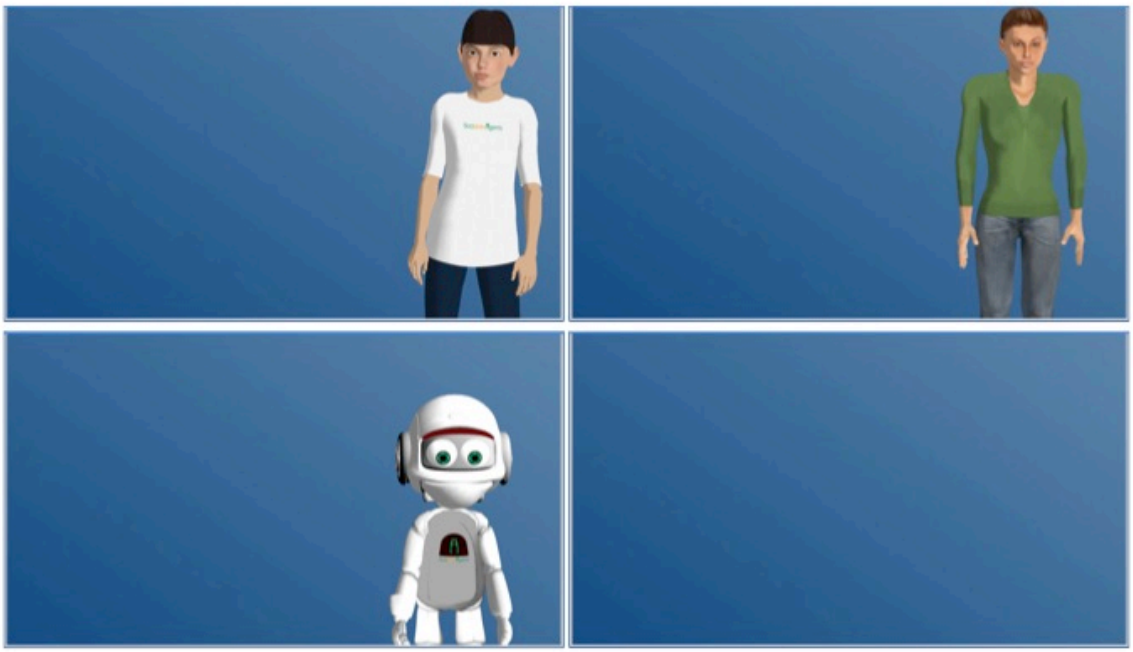


Figure 12. Overview over the used appearances: Billie (left upper corner), Character (right upper corner), Vince (left lower corner) and VoiceOnly (right lower corner)

All appearances are displayed in *Figure 12* and detailed descriptions for each one are listed in the following.

- *Billie*: Billie is a humanoid agent (species: human) with a childish look. It is somehow androgynous, while more people would describe it as male than

female. Its hair is dark brown and the eyes are also brown. Based on its white skin, it looks Caucasian. It is wearing a white shirt and blue trousers. Its overall degree of realism is rather low (realism: low). Although the degree of details is higher – since, for example, freckles are obtainable – it has a cartoon-stylized shade. The proportions are rather natural and not stylized. However, the resolution is more unrealistic than realistic owing to the material and texture.

- *Character*: The second humanoid agent (species: human) is a female adult and was created with Autodesk's Character Builder. She is wearing a green top and jeans. She can also be described as Caucasian, as her skin is white, her hair is brown and her eyes are blue. The degree of realism is rather high (realism: high), since many details are obtainable, no stylized shade or proportions are used and the resolution is also more realistic (although not completely photorealistic).
- *Vince*: Vince is a machinelike agent; it is a white humanoid robot (species: robot). Its degree of realism is comparable (a bit lower) to Billie, since it has a more cartoonlike look (realism: low). The degree of details is low and its shading and proportions are more cartoon stylized. Again, the resolution is more unrealistic.
- *Voice only*: In the voice-only condition, participants did not see any embodied character. While the agents are represented in the other three conditions, in this condition only a blue background is shown. Participants interacted with the voice only and had no further representation of the agent. Consequently, this condition cannot be described along the appearance variables (e.g., species and realism), since no embodied character appeared.

Since participants interacted with the agents, an actual agent with an underlying skeleton (that can be animated to move and talk) was needed. Owing to these technical restrictions, the possible design decisions and usable appearances were limited. Thus, the manipulation used is not as controlled as in the online study in which only pictures were used and could be manipulated more easily. Nevertheless, it is a big advantage to investigate the effect of appearance in an interaction situation, where participants are able to communicate with the virtual agent. As described earlier, the appearances can be used to explore the effects of species, realism, and embodiment in an actual human–agent interaction.

Interaction Procedure

The participants' task was to fill in a health diary and to schedule appointments in a calendar. That is a possible scenario in the application of daily-life assistance, since a virtual agent might be able to help with health choices, give reminders in regard to this, and help plan and structure the day. Within this scenario, different sections exist:

- *Introduction*: The agent welcomed the participant, introduced itself, and asked for the participant's name. Subsequently, the agent gave some information about itself and the task they would be solving together.
- *User Calibration*: In the first part of the diary, the participant's data were filled in. The agent therefore asked for the participant's sex, age, height, weight, and body type.
- *Preferences, Sports*: Afterwards, the participant's preferences with regard to sports were queried. Thus, their favorite sport was filled in; participants were asked to make decisions between different sports (e.g., swimming or running) and whether they prefer to do their exercises alone or together in a group.
- *Preferences, Food*: In the same manner, the participant's preferences with regard to food were recorded. They could state their favorite food and indicate whether they have any special diet.
- *Diary Entries*: Participants were asked to fill in their habits of the past week with regard to exercises, sleep behavior, drinking behavior, and food choices.
- *Calendar Task*: Within the calendar setting, participants were able to schedule different appointments in a calendar. Participants were free in their choices of which appointments they could enter in the calendar and whether these were actual appointments or fictional ones. Participants filled in four appointments, while one of them was misunderstood by the agent on purpose to analyze how people react toward such mistakes. In addition, the agent made two offers in back-reference to the participants' entries in their health diary (e.g., "You told me that you love to go swimming [referring to the participants favorite sport]. Would you like to schedule this in your calendar?").
- *Goodbye*: In the end, the agent thanked the participants for their participation and asked them to call the experimenter back in.

The diary entries were presented in tables and were supplemented by matching icons, to adapt to people in need of support who might have difficulties to read and understand the textual inputs. *Figure 13* presents an example of how the entries looked.

Figure 13. Presented tables during the interaction with the agent

Agent's behavior

During the interaction, the agent used the same voice for all conditions. Its nonverbal behavior was also held constant as much as possible. Owing to the different body types and skeletons, small differences might exist, for example, in the body posture. However, as the nonverbal behavior of the agent is not in focus, only small movements (e.g., head nod, small pointing gestures, smiling) were used, to let the agent appear more lifelike.

8.2.3 Measurement

As measures, mainly self-reported data queried via questionnaires were used. The questionnaire was divided into two parts. One was filled in before the experimental manipulation and the second one afterwards. Table 19 gives an overview of the measures used sorted according to their order of occurrence.

Table 19
Measures used in the study with the number of items, example items and reliability values

Scales	Subscales	Number of items	Example items	Cronbach's Alpha
Questionnaire Part I: *Control Variables*				
Anthropomorphism Tendency	(only general subscale)	10 items	*"I sometimes wonder if my computer deliberately runs more slowly after I have shouted at it"*	*.796*
Prior Experiences	Knowledge about virtual agents	Single item	*"Do you know what a virtual agent is?"*	-
	Definition of a virtual agent		*"A virtual agent is an autonomously acting, animated character that is represented on a computer-screen. This character has artificial intelligence and can communicate humanlike. It is therefore, able to communicate verbal and nonverbal and can understand human speech and gestures. Because of this, a natural way of communication is given."*	
	Prior interactions	4 items	*"Have you interacted with a virtual agent before?" "In what context?" "Had this agent an embodiment?" "How did the agent looked like?"*	-
Attitudes and Anxieties	Anxiety	4 items	*"If I should use the virtual agent, I would be afraid to make mistakes with it."*	*.725*
	General Attitude	3 items	*"I think it's a good idea to use the virtual agent."*	*.752*
Negative Attitudes toward agents (NARS)	S1	6 items	*"I would feel uneasy if I was given a job where I had to use a virtual agent."*	*.649*
	S2	5 items	*"I would feel uneasy if virtual agents really had emotions."*	*.635*
	S3	3 items	*"I feel comforted being with virtual agents that have emotions (R)."*	*.618*
Expectations	Preferred functions	Single item	*"Which function should such a virtual assistant have in your daily-life?"*	-
	Preferred Appearance	Single item	*"How should your ideal virtual assistant look-like?"*	-
Experimental Manipulation: *Interaction with the virtual agent*				
Questionnaire Part II: *Dependent variables*				

Person Perception	Realism	7 items	*"fake – natural"*	*.894*
	Likability	7 items	*"unfriendly – friendly"*	*.912*
	Attractiveness	4 items	*"unattractive – attractive"*	*.874*
	Trustworthiness	5 items	*"not trustworthy – trustworthy"*	*.808*
	Competence	5 items	*"incompetent – competent"*	*.852*
Liking	Liking scale	5 items	*"I think, I would like this agent."*	*.869*
	Open answers	3 items	*"What do you liked most about the agent's appearance?"*	-
Usage	Usage intention	3 items	*"I think I'll use the agent during the next few days."*	*.934*
	Perceived usefulness	3 items	*"I think the agent is useful to me."*	*.939*
	Trust	2 items	*"I would trust the robot if it gave me advice"*	*.930*
WOI	Bonding	12 items	*"I feel uncomfortable with the agent."*	*.842*
Interaction Evaluation	Enjoyment	5 items	*"I enjoyed the agent talking to me."*	*.877*
	Ease of use	5 items	*"I think I will know quickly how to use the agent."*	*.544*
	Sociability	4 items	*"I consider the robot a pleasant conversational partner"*	*.799*
Intention Health Behavior		8 items	*"Eat a well-balanced diet"*	*.831*
Socio-demographics	Age	Single item	*"How old are you?"*	-
	Sex	Single item	*"Which sex do you have?"*	-
	Education	Single item	*"What is your highest educational level?"*	-
	Profession	Single item	*"What is your current profession?"*	-
	Comment	Single item	*"Do you have any comments about the study?"*	-

Dependent Variables

In order to examine the effect of appearance on the evaluation of the virtual agent, the person perception of the agent was queried. Five different concepts were measured with overall 28 items on a 5-point semantic differential: perceived realism, likability, trustworthiness, competence, and attractiveness. The items were adapted from prior person perception measurements (Goodspeed Questionnaire; Bartneck, Kulić, Croft, & Zoghbi, 2009), (Heerink et al., 2010; McDonnell et al., 2012). See table X for example items and the exact number of items per subscale. All scales showed a good reliability (Cronbach's alpha < .808).

Supplementing this measure, participants' liking of the agent was measured with an ad hoc scale comprising five items (e.g., "I think, I would like this agent") rated on a 5-point

Likert scale (ranging from 1= *totally disagree* to 5= *totally agree*). The internal consistency of the scale was good (Cronbach's alpha = .869).

Additionally, the usage behavior of the participants was assessed. The perceived usefulness was measured with three items (e.g., "I think the agent is useful to me") and participants' intention to use the virtual agent was measured with three items (e.g., "I think I'll use the agent during the next few days"). All items were rated on a 5-point Likert scale (ranging from 1= *totally disagree* to 5= *totally agree*). Both scales showed an excellent internal consistency (Cronbach's alpha < .934).

Furthermore, trust in the virtual agent was queried with two items (e.g., "I would trust the robot if it gave me advice") rated on a 5-point Likert scale (ranging from 1= *totally disagree* to 5= *totally agree*). The reliability of this scale was excellent (Cronbach's alpha = .930).

Since the relationship between user and agent is the focus of this work, participants' bonding with the agent was measured using the bonding subscale of the Working Alliance Inventory (WAI; Horvath & Greenberg, 1986). The 12 items were rated on a 5-point Likert scale ranging from 1 = *totally disagree* to 5 = *totally agree,* and showed good reliability (Cronbach's alpha = .842).

To evaluate the interaction, enjoyment (five items, e.g., "I enjoyed the agent talking to me"), ease of use (five items, e.g., "I enjoyed the agent talking to me"), and sociability (four items, e.g., "I consider the robot a pleasant conversational partner") were measured. Participants evaluated all items on a 5-point Likert scale ranging from 1 = *totally disagree* to 5 = *totally agree*. Cronbach's alpha values demonstrated an acceptable internal consistency for enjoyment and sociability (Cronbach's alpha < .799), while the consistency of ease of use was not acceptable (Cronbach's alpha = .544). Nevertheless, since this concept is of high relevance for this study, it is included in further analyses, but the results have to be discussed cautiously.

In addition, participants' intended health behavior was measured, to see whether the virtual agent influenced them. This was measured with five items (e.g., "Eat a well-balanced diet" or "Eat fresh fruits and vegetables"; Moorman, Matulich, Moorman, & Matulich, 1993). Since this scale mostly contains behaviors regarding diet, three items concerning physical exercises from Cunningham and Kwon (2003) were added (e.g., "I'm planning to be physically active on a regular basis next week"). Participants rated all items on a 5-point Likert scale ranging from 1 = *totally disagree* to 5 = *totally agree*. The reliability of this scale was good (Cronbach's alpha = .831).

Control Variables

Several control variables were measured in the first part of the questionnaire. For example items, number of items, and internal consistency values, please see Table X. Participants' tendency to anthropomorphize was measured via a subscale with 10 items (Neave et al., 2015) rated on a 5-point Likert scale ranging from 1 = *totally disagree* to 5 = *totally agree*. The internal consistency of this scale is acceptable (Cronbach's alpha = .796).

Moreover, participants reported on their prior experiences by indicating whether they knew what a virtual agent is (binary answer: *yes / no*), if they had used one before (binary answer: *yes / no*) and if *yes*, in which context, whether the agent was embodied and how it looked (all open questions).

Additionally, participants' attitude, anxiety, and expectation toward a virtual assistant were queried. On the basis of prior research (Heerink et al., 2010), participants' anxiety toward virtual agents was measured with four items and their general attitude with three items. These items were adapted to the context of virtual agents, since the items originated from a scale for social robots. Again, both scales showed an acceptable reliability (Cronbach's alpha < .725). In addition, the Negative Attitudes Toward Robots Scale (NARS; Nomura et al., 2006) was also transferred to the application of virtual agents and used to examine participants' attitude in more detail. Participants rated all scales on a 5-point Likert scale ranging from 1 = *totally disagree* to 5 = *totally agree*. Cronbach's alpha values of the NARS subscales are questionable (Cronbach's alpha < .618). Therefore, these measures have to be analyzed and discussed with caution. Two open questions were used to determine participants' expectations of a virtual agent's functions and appearance.

In the end, sociodemographic variables such as age and sex were measured.

8.2.4 Procedure

When participants came into the laboratory, the experimenter welcomed them. To obtain informed consent, they were informed about the background and procedure of the study. After participants were introduced to the study and agreed to participate, the study part began. All participants first filled in questionnaires about personality traits, prior experiences, and other control variables. While most of the participants filled in the questionnaires on a computer, seniors who were not familiar with using computers could also choose to use a paper–pencil version of the questionnaire. Whenever needed, the experimenter assisted the participants and answered questions. When participants finished the first part of the questionnaire, the experimental part of the study began. Participants were asked to interact

with the virtual assistant in the context of a health diary. During the interaction, the experimenter left the room and participants were alone with the virtual agent. They were told that the agent interacts autonomously and can understand and react to their speech and behavior, but in fact a Wizard-of-Oz setting was used, where a confederate controlled the agent from a room next door. The confederate could hear and see the participant through a camera and was able to control the answers of the virtual agent. Since the context of the study is a daily-life assistant, participants' tasks were to fill in a health diary and the virtual agent gave some advice regarding participants' habits. Participants were asked to start the interaction with "Hello Billie." The confederate replied to this and the interaction began. In the end of the interaction, the agent told the participants that the interaction was over and that they could call in the experimenter again. After the interaction with the agent was finished, participants called the experimenter back in and the second questionnaire part started, where the dependent variables was retrieved. In the end, the experimenter debriefed the participants, explained the goals and manipulation of the study, thanked them for their participation, and offered the incentive (either money or course credits).

8.3 RESULTS

In the following results are presented along the dependent variables. The study aims to investigate the effect of species, realism and embodiment more closely in a real human-agent interaction. Therefore, planned contrasts are used, to analyze the data with regard to the specific comparisons and assumed hypotheses:

- Embodiment (Contrast 1): Billie, Vince and Character vs. VoiceOnly
- Species (Contrast 2): Billie and Character vs. Vince
- Realism (Contrast 3): Billie vs. Character

Furthermore, when significant interaction were found, according to Field (2018) those effects were further investigate using simple effects, where the effect of age groups at individual levels of the different appearances were looked at. Results are presented along the dependent variables and the decisions about hypotheses are summarized in Chapter 8.3.7.

8.3.1 Person Perception

In order to test the influence of the different appearances and the age group on the person perception of the virtual assistant, a 2-way MANOVA with appearance (4 factors: Billie, Character, Vince and Voice only) and age group (2 factors: students and seniors) as

independent variables and perceived realism, likability, attractiveness, trustworthiness and competence as dependent variables was conducted.

Using Pillai's trace, there was a significant effect of the age group ($V = 0.21$, $F(5, 118) = 6.17$, $p < .001$) and appearance ($V = 0.21$, $F(15, 360) = 1.84$, $p < .001$) on person perception.

To test hypothesis *H7a*, the univariate tests was consulted. Results indicate that seniors and students differ significantly in their evaluation of perceived realism ($F(1, 129) = 9.07$, $p = .003$, $\eta_p^2 = .069$) and attractiveness ($F(1, 129) = 21.16$, $p < .001$, $\eta_p^2 = .148$) of the agent. Seniors evaluated the agent in general as more realistic and more attractive than students did. No significant differences between the age groups were found for likability ($F(1, 129) = 1.12$, $p = .293$, $\eta_p^2 = .009$), trustworthiness ($F(1, 129) = 2.34$, $p = .129$, $\eta_p^2 = .019$) and competence ($F(1, 129) = 3.75$, $p = .055$, $\eta_p^2 = .030$). Refer to Table 20 for descriptive values. Overall, hypothesis *H7a* was only partly supported for attractiveness and has to be rejected.

Table 20
Person perception evaluation of the different age groups in general

	Seniors		**Students**		**Overall**	
	M	*SD*	*M*	*SD*	*M*	*SD*
realism	3.22	1.01	2.70	0.81	2.89	0.91
likability	4.11	0.76	3.95	0.71	4.01	0.73
attractiveness	3.79	0.91	3.01	0.87	3.29	0.96
trustworthiness	3.45	0.80	3.19	0.86	3.28	0.85
competence	4.13	0.61	3.86	0.76	3.95	0.71

Univariate tests further revealed a significant difference between appearances in their perceived competence, $F(3, 129) = 2.98$, $p = .034$, $\eta_p^2 = .068$. Running the planned contrast, results of the analyses showed significant effect of embodiment ($t(126) = -2.21$, $p = .029$, r = .19), where an embodied character was evaluated as less competent than the VoiceOnly condition. Consult Table X for means and standard deviations. Furthermore, a marginal significant effect of appearance on likability was found ($F(3, 129) = 2.63$, $p = .053$, $\eta_p^2 = .061$). Again, planned contrasts were used to explore this effect more deeply. Here a significant difference in species was found ($t(126) = -2.52$, $p = .013$, r = .22), where Vince

was evaluated as more likable than Billie and Character (Table 21). The tests yielded no significance for realism ($F(3, 129) = 2.20$, $p = .092$, $\eta_p^2 = .051$), attractiveness ($F(3, 129) = 1.99$, $p = .120$, $\eta_p^2 = .047$) and trustworthiness ($F(3, 129) = 0.60$, $p = .616$, $\eta_p^2 = .015$). Although for likability results support the assumption with regard to species (*H1a*), that a machinelike agent will be evaluated more positive compared to humanoid agents, no differences in the other variables like attractiveness, trustworthiness and competence were found. With regard to hypothesis *H3a* results even support an opposite finding, since a non-embodied character leads to higher ratings of competence than an embodied one. However, no such differences in the other person perception variables where found. As mostly no differences between the appearances in the agent's person perception occurred, the hypotheses *H1a*, *H2a* and *H3a* have to be rejected.

Table 21
Person perception evaluation for the different appearances regardless of the age group

	Billie		**Vince**		**Character**		**Voice Only**		**Overall**	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
realism	2.89	0.90	2.76	0.78	2.77	0.84	3.12	1.09	2.89	0.91
likability	3.82	0.77	4.20	0.58	3.82	0.72	4.15	0.78	4.01	0.73
attractiveness	3.14	0.95	3.49	0.98	3.03	0.99	3.47	0.88	3.29	0.96
trustworthiness	3.16	1.00	3.46	0.78	3.21	0.77	3.28	0.83	3.28	0.85
competence	3.78	0.76	4.04	0.66	3.81	0.71	4.18	0.67	3.95	0.71

According to Pillai's trace, no significant interaction effects of both dependent variables occurred on person perception, $V = 0.09$, $F(15, 360) = 0.74$, $p = .747$. In contrast, the univariate tests showed a significant interaction effect of appearance and age group with regard to perceived realism ($F(3, 129) = 2.71$, $p = .048$, $\eta_p^2 = .063$), but not for likability ($F(3, 129) = 1.18$, $p = .320$, $\eta_p^2 = .028$), attractiveness ($F(3, 129) = 0.47$, $p = .707$, $\eta_p^2 = .011$), trustworthiness ($F(3, 129) = 0.73$, $p = .536$, $\eta_p^2 = .018$) and competence ($F(3, 129) = 1.33$, $p = .268$, $\eta_p^2 = .032$). To analyze the interaction effect for perceived realism in more details, simple effects where used and revealed that seniors (Billie: $M = 3.40$, $SD = 0.83$; VoiceOnly: $M = 3.76$, $SD = 1.15$) evaluated Billie ($F(1,122) = 6.76$, $p = .010$) and VoiceOnly ($F(1,122) =$

7.74, p = .002) as more realistic than students (Billie: M = 2.59, SD = 0.86; VoiceOnly: M = 2.76, SD = 0.88) did (see Figure X). As only an interaction for realism was found, hypotheses *H4a*, *H5a* and *H6a* cannot be supported by the current data.

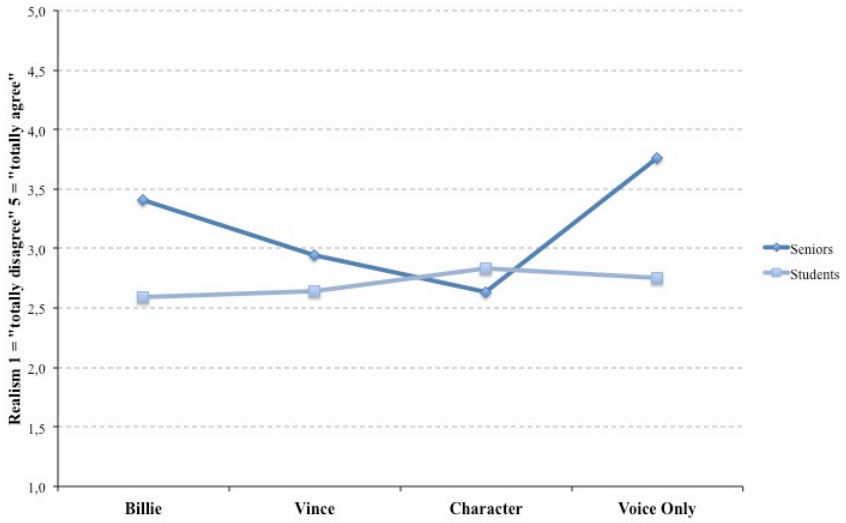


Figure 14. Interaction effect of appearance and age group for perceived realism

8.3.2 Liking of the agent

Since effects age and appearance on the user's liking of the agent were assumed, a 2-way ANOVA with appearance and age group as independent variables and liking of the agent as dependent variables was calculated. A significant difference between the age groups was obtainable (F(1, 129) = 10.71, p = .001, η_p^2 = .081) where seniors (M = 3.01, SD = 0.98) like the agents more in general than students (M = 2.39, SD = 1.03). Thus, hypothesis *H7b* is supported. However, only a marginal significant effect of appearance (F(3, 129) = 2.41, p = .070, η_p^2 = .056) and no significant interaction effect (F(3, 129) = 0.99, p = .400, η_p^2 = .024) occurred. Planned contrasts for the appearance effect, revealed an influence of species (t(126) = -2.71, p = .008, r = .23). Users stated to like the agent more in the condition where Vince (M = 2.93, SD = 1.12) was presented than in both humanoid conditions (Billie: M = 2.39, SD = 0.92; Character: M = 2.28, SD = 0.87). Those results support hypothesis *H1b,* while it has to be kept in mind that the main effect was only marginal significant. In addition differences in the participant's liking regarding realism (*H2b*), embodiment (*H3b*) and more specific interaction effects between age and those variables (*H4-6b*) have been hypothesized, but the presented findings cannot support these hypotheses.

Moreover, to investigate the influence of user characteristics on liking of the agent and to find an answerer to research question *RQ2a*, a hierarchical regression analysis with four steps was done. In Step 1 age and sex of the participants were included. In addition to user

characteristics based on technology acceptance models (see Chapter II4.2) participants' enjoyment, ease of use and usefulness were included in Step 2. For Step 3 the users' tendency to anthropomorphize was entered in the model and Step 4 tested the influence of anxiety, attitude and negative attitude towards virtual agents on the liking behavior. The results of the analysis are presented in Table 22. Here the presented main effect for age resulting from the ANOVA was again supported, as age significantly positive predicts participant's liking. Furthermore, enjoyment can be obtained as significant predictor of liking of the agent. The more people enjoyed the interaction with the agent, the more they liked the agent.

Table 22
Hierarchical regression analysis with liking as dependent variable

	Liking				
	b (*SE*$_b$)	*ß*	*t*	*p*	ΔR^2
Step 1					**.083**
Sex	-.04 (.19)	-.02	-0.19	.850	
Age	.01 (.00)	**.29**	3.34	.001	
Step 2					**.199**
Enjoyment	.39 (.12)	**.33**	3.29	.001	
Ease of Use	.02 (.12)	.02	0.20	.844	
Usefulness	.16 (.08)	.18	1.90	.060	
Step 3					.001
Anthropomorphism Tendency	.04 (.12)	.03	0.36	.719	
Step 4					.031
Anxiety toward agents	-.01 (.12)	-.01	-0.06	.954	
Attitude toward agents	-.14 (.12)	-.12	-1.22	.224	
Negative attitude toward situations of interaction with agents	-.14 (.15)	.10	0.91	.365	
Negative attitude toward social influence of agents	.13 (.13)	.09	0.97	.333	
Negative attitude toward emotions in interaction with agents	-.09 (.13)	-.07	-0.65	.515	
Total R^2					.314
Step1	$F(2,123) = 5.60$, $p = .005$				
Step 2	$F(3,120) = 9.43$, $p < .001$				
Step 3	$F(1,119) = 7.82$, $p < .001$				
Step 4	$F(5,114) = 4.74$, $p < .001$				

However, no effects of ease of use and usefulness have been found and the former opinion towards agents (anxiety and attitudes) did also not explain participants' liking of the agent significantly. Thus, the only user characteristic that was found to be a significant predictor of liking is age.

8.3.3 Ease of Use, Perceived Usefulness and Usage Intention

Using again a 2-way MANOVA with appearance (4 factors: Billie, Character, Vince and Voice only) and age group (2 factors: students and seniors) as independent variables and ease of use, perceived usefulness and usage intention as dependent variables, it was tested how age groups differ from each other and how these variables are influenced by the different agent's appearances. Results indicate based on Pillai's trace significant main effects for age group ($V = 0.98$, $F(3, 120) = 3.86$, $p = .011$) and appearance ($V = 0.14$, $F(9, 366) = 1.98$, $p = .041$) and additional a significant interaction effect of both ($V = 0.16$, $F(9, 366) = 2.25$, $p = .019$).

Referring to the univariate tests, both age groups differ in their evaluation of the agent's ease of use ($F(1, 129) = 4.07$, $p = .002$, $\eta_p^2 = .073$) with students rating the agent as easier to use than seniors (see Table 23). This finding is in line with the assumed hypothesis *H8*, where students were expected to state higher ease of use values than seniors. Usage intention ($F(1, 129) = 0.40$, $p = .594$, $\eta_p^2 = .002$) and perceived usefulness ($F(1, 129) = 0.76$, $p = .387$, $\eta_p^2 = .006$) did not turn significant for age differences. Thus, hypotheses *H7c* and *H7d* were not supported by the data, as students and seniors seem to state same levels of perceived usefulness and usage intention.

Table 23
Usage intention, ease of use and perceived usefulness evaluation of the different age groups in general

	Seniors		**Students**		**Overall**	
	M	*SD*	*M*	*SD*	*M*	*SD*
Usage intention	2.78	1.27	2.54	1.27	2.63	1.27
Ease of use	3.77	0.63	4.17	0.69	4.03	0.69
Usefulness	2.96	1.27	2.80	1.25	2.86	1.25

Looking at the univariate tests no significant differences between the appearances in participants' usage intention ($F(3, 129) = 1.20$, $p = .313$, $\eta_p^2 = .029$), ease of use ($F(3, 129) = 1.50$, $p = .219$, $\eta_p^2 = .035$) and usefulness ($F(3, 129) = 1.49$, $p = .220$, $\eta_p^2 = .035$) occurred. Consult Table 24 for descriptive values. These results contradict the hypotheses *H1c*, *H2c* and *H3c*, where differences between the appearances in regard to participants' usage intention were assumed.

Table 24
Usage intention, ease of use and perceived usefulness evaluation for the different appearance regardless of the age group

	Billie		**Vince**		**Character**		**Voice Only**		**Overall**	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Usage intention	2.70	1.25	2.72	1.45	2.33	1.26	2.72	1.10	2.63	1.27
Ease of use	3.81	0.65	4.02	0.72	4.09	0.64	4.19	0.71	4.03	0.69
Usefulness	2.65	1.32	2.98	1.19	2.56	1.27	3.20	1.20	2.86	1.25

However, a significant interaction effect of age groups and appearances was found for perceived usefulness ($F(3, 129) = 5.06$, $p = .002$, $\eta_p^2 = .111$) and usage intention ($F(3, 129) = 3.38$, $p = .021$, $\eta_p^2 = .077$), but not for ease of use ($F(3, 129) = 1.96$, $p = .123$, $\eta_p^2 = .046$). Consult *Figure 2* for this interaction effect.

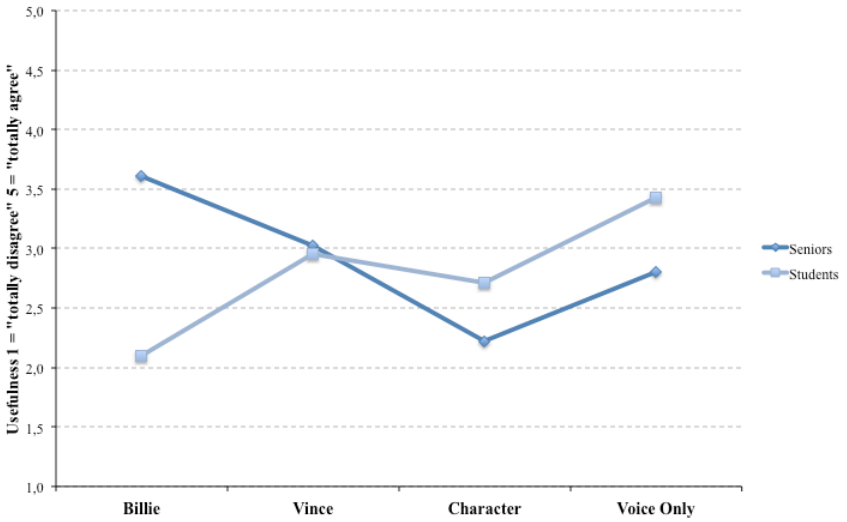


Figure 15. Interaction effect of appearance and age group for perceived usefulness

Again, simple effects were used to investigate the interaction effect. For perceived usefulness, seniors ($M = 3.61$, $SD = 1.32$) evaluated the usefulness of Billie higher than students ($M = 2.10$, $SD = 0.97$) did ($F(1,122) = 12.42$, $p = .001$).

In line with this, seniors ($M = 3.58$, $SD = 1.20$) stated higher usage intention compared to students ($M = 2.19$, $SD = 0.99$) after the interaction with Billie ($F(1,122) = 9.63$, $p = .002$). Consult *Figure 16* to obtain the interaction effect.

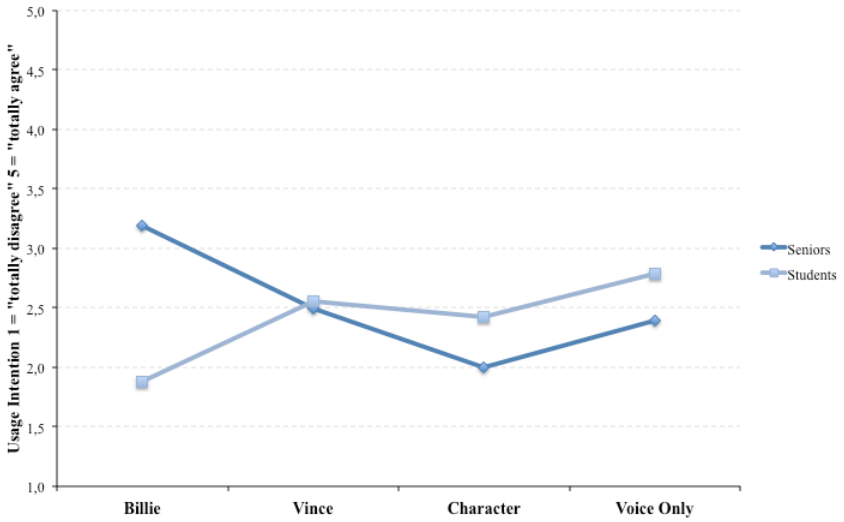


Figure 16. Interaction of appearance and age group for usage intention

Based on these results, the hypotheses *H4b* is partly supported, since it was assumed that seniors show greater usage intention for a humanoid agent. However, this was only true for a cartoon-stylized human. No such differences between both target groups in the evaluation of embodied compared to non-embodied agents were found, therefore the hypotheses *H5b* and *H6b* have to be rejected.

Since hypothesis H11 assumed bonding to mediate the effect of liking and trust on usage intention, two mediation analyses were carried out to investigate this relationship. For this propose the INDIRECT macro (Preacher & Hayes, 2008) was used with 5,000 bootstrap resamples (percentile-based 95% confidence interval).

In the first model participants liking of the agent was used as independent variable and bonding as mediator, while usage intention serves as dependent variable. Here a significant effect of liking on participants' bonding was found, but no significant direct effect on participants' usage intention, while the indirect effect of liking on usage intention mediated by bonding turned significant. Although no significant direct effect exists, it can still be assumed that the effect of liking on usage intention decreased when the mediator is considered in the model as methodology experts (e.g. Hayes, 2013; MacKinnon, 2008) suggest that indirect effects can also exist with the absence of a statistically significant direct effect. Consult *Figure 18* for exact values and influences. Hence, the data supports hypothesis *H11a*. In the second model the influence of trust on usage intention mediated by bonding was analyzed, where participants' self-reported trust was used as independent variable and their stated usage intention as dependent variable. Results and values are presented in *Figure 18*.

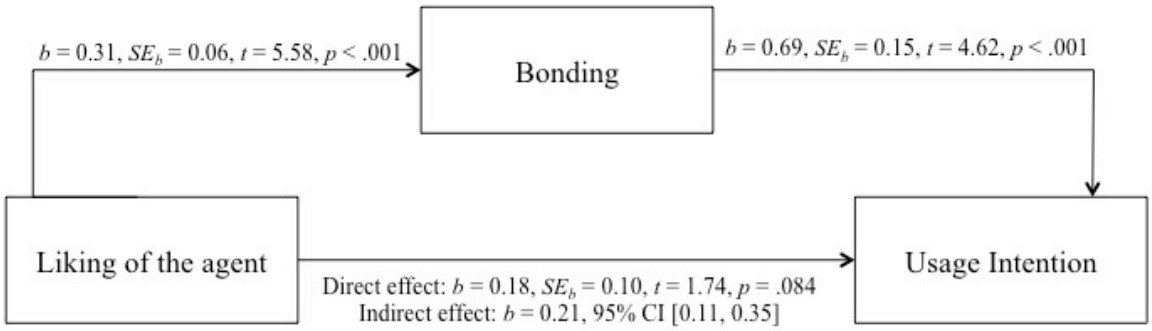


Figure 17. Model of liking of the agent as a predictor of usage intention, mediated by user's bonding towards the agent. The confidence interval for the indirect effect is a BCa bootstrapped CI based on 5000 samples. Model summary: $R^2 = 0.33$, $F(1,125) = 15.54$, $p = .001$.

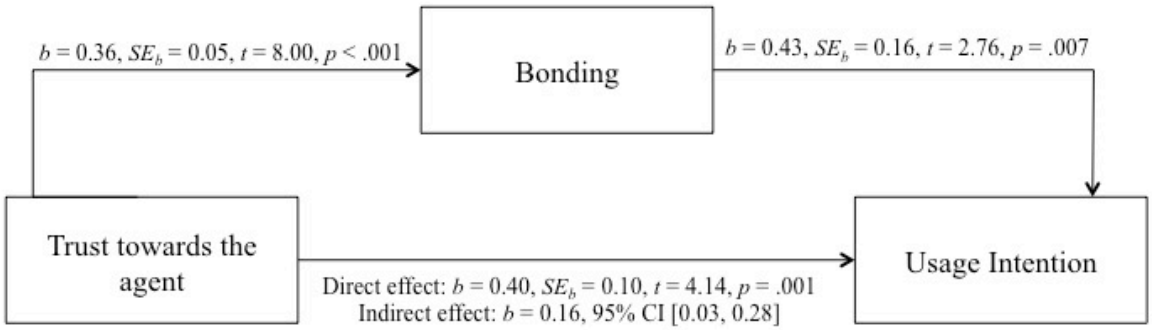


Figure 18. Model of trust towards the agent as a predictor of usage intention, mediated by user's bonding towards the agent. The confidence interval for the indirect effect is a BCa bootstrapped CI based on 5000 samples. Model summary: $R^2 = 0.52$, $F(1,125) = 47.42$, $p < .001$.

Here a significant effect of trust on bonding was found and participants' bonding significantly influences the usage intention in a positive way. Furthermore, both a significant direct effect of trust towards the agent on participants' intention to use the agent and a significant indirect effect indicating bonding to mediate the effect of trust on usage intention (see *Figure 17*). Again the data supports the assumed hypothesis *H11*, since bonding mediates the effect of liking (*H11a*) and trust (*H11b*) on usage intention.

Furthermore, to further explore predicting variables of usage intention a hierarchical regression analysis with four steps was calculated. With this analysis the influence of the perception of the interaction (Step 2; *H9*) was tested and the influence of user characteristics on usage intention (Step 1 & 3-4; *RQ2b*) was investigated. Similar to the hierarchical regression on liking the following variables were entered in the model: Step 1: Sex and Age; Step 2: Enjoyment, Ease of Use and Usefulness; Step 3: Anthropomorphism Tendency; Step

4: Anxiety toward agents, Attitude toward agents and the three subscales of negative attitude toward agents. Consult Table 25 to obtain results and values of the analysis. In line with the analysis of variance above no significant of age was found. In addition, participants' sex did also not predict the usage intention significantly. While enjoyment and ease of use did also not influence usage intention, the perceived usefulness is a significant predictor. Participants, who perceive the agent to be useful, also show higher intention to use the agent. Hence, hypothesis *H9* is only partly supported for usefulness. The only user characteristic that was found to predict usage intention was the general attitude towards virtual agents. Participants, who have a more positive attitude towards virtual agents before they interacted with the agent, tend to show higher usage intentions after the interaction. All other variables (anthropomorphism tendency, anxiety toward agents and negative attitude toward agents) cannot explain participants' usage intention.

Table 25
Hierarchical regression analysis with usage intention as dependent variable

	Usage intention				
	b (SE_b)	*ß*	*t*	*p*	ΔR^2
Step 1					.010
Sex	-.14 (.23)	-.05	-0.60	.550	
Age	.01 (.01)	.09	1.03	.307	
Step 2					**.708**
Enjoyment	.16 (.09)	.11	1.79	.075	
Ease of Use	.09 (.09)	.05	0.97	.334	
Usefulness	.79 (.06)	**.78**	12.94	<.001	
Step 3					.000
Anthropomorphism Tendency	-.04 (.09)	-.02	-0.44	.660	
Step 4					.023
Anxiety toward agents	.05 (.09)	.03	0.51	.609	
Attitude toward agents	.17 (.09)	**.12**	2.04	.044	
Negative attitude toward situations of interaction with agents	-.19 (.11)	-.11	-1.74	.085	
Negative attitude toward social influence of agents	.18 (.10)	.11	1.91	.059	
Negative attitude toward emotions in interaction with agents	.09 (.10)	.06	0.92	.360	
Total R^2					.741
Step1	$F(2,123) = 0.65$, $p = .522$				
Step 2	$F(3,120) = 61.19$, $p < .001$				
Step 3	$F(1,119) = 50.68$, $p < .001$				
Step 4	$F(5,114) = 29.72$, $p < .001$				

Note. Values in bold indicate significant relationships.

8.3.4 Bonding, Trust & Sociability

As research question *RQ1* aims to investigate the effect of appearance and age group on bonding and trust, multiple 2-way ANOVAs have been calculated to answer this research question.

The influence of age group and appearance on participants' bonding was analyzed using a 2-way ANOVA with appearance (4 factors: Billie, Character, Vince and Voice only) and age group (2 factors: students and seniors) as independent variables and bonding as dependent variable. Analyses revealed a significant main effect for age group ($F(1, 126) = 11.46$, $p = .001$, $\eta_p^2 = .088$) and a significant interaction effect of both independent variables ($F(3, 129) = 3.67$, $p = .014$, $\eta_p^2 = .085$), while the different appearances did not differ in participants' bonding ($F(3, 129) = 1.09$, $p = .356$, $\eta_p^2 = .027$). Referring to the descriptive values seniors ($M = 3.40$, $SD = 0.79$) stated higher feelings of bonding toward the agent than students ($M = 2.94$, $SD = 0.67$) did.

Table 26
Means and Standard deviation for all Appearances along the age groups for bonding, trust and sociability

	Billie				**Vince**				**Character**				**Voice Only**			
	Seniors		**Students**		**Seniors**		**Students**		**Seniors**		**Students**		**Seniors**		**Students**	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Bonding	3.57	0.94	2.58	0.45	3.20	0.67	3.08	0.75	3.02	0.78	3.11	0.76	3.73	0.67	2.98	0.58
Trust	3.79	0.94	2.19	1.04	2.62	1.14	3.07	1.28	2.44	1.10	3.18	1.16	3.00	1.28	3.12	1.13
Sociability	4.13	0.81	3.00	.742	4.04	0.72	3.81	0.88	3.36	0.99	3.44	0.84	4.38	0.85	4.05	0.83

To explore the interaction effect between age group and appearance more, simple effects were used. Results indicate that students and seniors show different bonding behavior in the Billie ($F(1,119) = 13.89$, $p < .001$) and VoiceOnly ($F(1,122) = 9.18$, $p = .003$) condition with seniors showing higher bonding in both conditions than students did (see Table 26 and *Figure 19*).

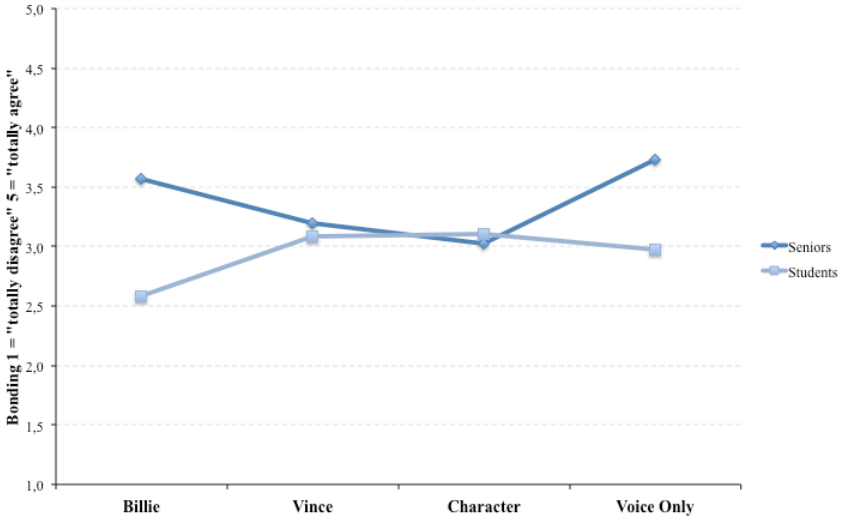


Figure 19. Interaction of appearance and age group for participants' bonding

A second 2-way ANOVA with same independent variables (age group and appearance) and trust as dependent variables was calculated. While no significant main effects of age group ($F(1, 129) = 0.13$, $p = .724$, $\eta_p^2 = .001$) and appearance ($F(3, 129) = 0.32$, $p = .814$, $\eta_p^2 = .008$) were found for trust, analyses yield a significant interaction effect of both variables ($F(3, 129) = 6.19$, $p = .001$, $\eta_p^2 = .132$). Simple effects further revealed that seniors showed more trust for Billie than students ($F(1,122) = 14.93$, $p < .001$), while both age groups did not differ in trust for the other appearances (see *Figure 2* and Table 26).

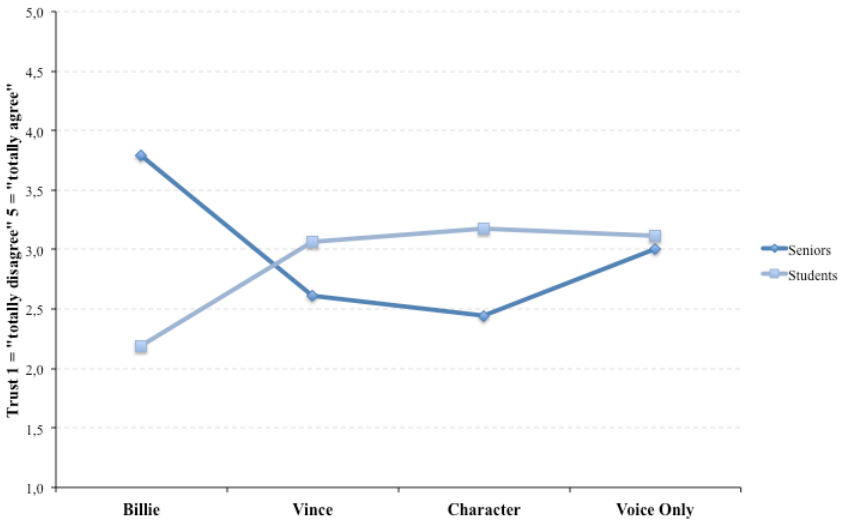


Figure 20. Interaction of appearance and age group for participants' trust towards the agent

In addition a 2-way ANOVA with both independent variables and perceived sociability of the agent has been run to analyze how the different appearances were perceived as sociable interaction partner and whether the age group has an effect on it. The analysis revealed a significant difference between age groups ($F(1, 129) = 8.88$, $p = .003$, $\eta_p^2 = .068$),

where seniors (M = 4.02, SD = 0.88) in general evaluated the agent as more sociable than students (M = 3.53, SD = 0.87) did.

Furthermore, a significant influence of appearance was found, $F(3, 129) = 4.57$, $p = .005$, $\eta_p^2 = .101$. The planned contrasts reveal a significant difference in perceived sociability between embodied and non-embodied agents ($t(126) = -2.72$, $p = .007$, r = .24) and a significant difference between machinelike and humanoid agents ($t(126) = -2.64$, $p = .009$, r = .23). Participant rated the agent in the VoiceOnly condition (M = 4.05, SD = 0.83) as more sociable than participants in the three embodied conditions (Billie: M = 3.41, SD = 0.93; Vince: M = 3.89, SD = 0.82 and Character: M = 3.41, SD = 0.87). Additionally, Vince the machinelike agent was found to evoke more sociability than both humanoid agents (Billie and Character).

In addition, a significant interaction effect of both variables occurred, $F(3, 129) = 2.80$, $p = .043$, $\eta_p^2 = .064$. The same pattern that was found before is also obtainable for sociability, since referring to simple effects seniors and students differed in their evaluation of Billie ($F(1,122) = 14.48$, $p < .001$), but not in the evaluation of the other appearances. Again, seniors rated Billie as more sociable than students did (see *Figure 21* and Table 26), while no differences between both groups regarding the other appearances occur.

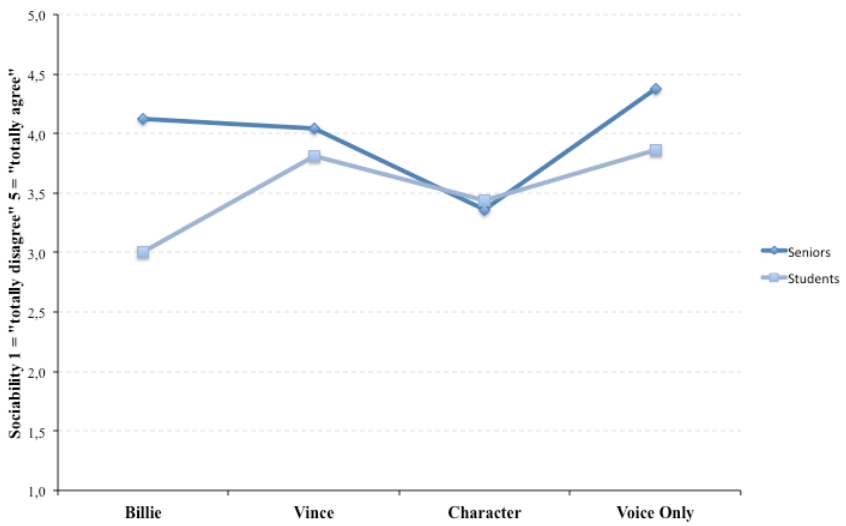


Figure 21. Interaction effect of appearance and age group for perceived sociability

8.3.5 Enjoyment

As seniors were generally assumed to perceive the agent and the interaction therewith more positive (*H7*), Hypothesis *H7e* claims that seniors will enjoy the interaction with the agent more than students. Results of a 2-way ANOVA with age group and appearance as independent variables and enjoyment of the interaction as dependent variable reveal a significant age effect, $F(1, 126) = 7.26$, $p = .008$, $\eta_p^2 = .057$. In line with prior findings,

seniors (M = 4.20, SD = 0.82) evaluated the interaction as more enjoyable than students (M = 3.72, SD = 0.89) did. Thus, the hypothesis *H7e* was supported by the presented findings. However, analysis yield no significant effect ($F(3, 126) = 1.61$, $p = .191$, $\eta_p^2 = .039$) of appearance nor a significant interaction effect, $F(3, 126) = 2.40$, $p = .071$, $\eta_p^2 = .057$.

8.3.6 Health Behavior

As explanatory investigation the persuasive effect of the agent on the participant's health behavior was considered. Thus, participants' intention to live healthy (e.g. drink water, be physical active, have a healthy diet) was analyses using a 2-way ANOVA with both independent variables (age group and appearance) and health behavior as dependent variable. Again, a significant difference between seniors and students was found ($F(1, 129) = 19.39$, $p < .001$, $\eta_p^2 = .137$) with seniors (M = 4.30, SD = 0.58) showing higher intention for a healthy lifestyle than students (M = 3.72, SD = 0.72). Differences between appearances did not turn significant $F(3, 129) = 0.84$, $p = .474$, $\eta_p^2 = .020$ and additionally no significant interaction between both variables was found, $F(3, 129) = 0.29$, $p = .836$, $\eta_p^2 = .007$.

8.3.7 Summary of the results and decision about the hypotheses

Following, the empirical findings become highlighted and the consequences for the assumed hypotheses and posed research questions will be summarized. Therefore, Table 27 presents an overview over all presented findings and decisions about the hypotheses and research questions.

While the species (human vs. robot; *H1*) varied in perceived likability and sociability and machinelike agents are more liked than humanoid ones, no differences in other person perception variables and usage intention occurred. Furthermore, the realistic and cartoon-stylized agents (*H2*) as well as embodied and non-embodied (*H3*) seems to be evoke equal person perceptions, were equally liked by participants and does not differ in participants' usage intention.

Regarding interaction effects between appearance and age group (*H4, H5, H6*), overall the same pattern occurred: Seniors evaluated the humanoid agent more positive than students did. While no differences in person perception and liking were found, seniors stated higher usage intentions than students after an interaction with the humanoid agent Billie. In line with this, Billie was perceived as more useful, evoked higher feelings of bonding and trust and was perceived as more social by seniors compared to students.

Moreover, seniors evaluated the agent and the interaction therewith more positive than students did (H7). The agent was rated as more realistic, attractive and social and was more liked by seniors. Additionally, seniors perceived the interaction with the agent as more joyful

and showed more bonding compared to students. Furthermore, seniors stated after the interaction with the agent a higher intention to live healthy than students. In contrast, as assumed in hypothesis *H8*, students perceived the agent to be easier to use than seniors.

With regard to research question RQ1, it can be summarized that seniors show higher social outcomes towards the agent. While the appearances overall did not affect bonding and trust, an interaction effect occurred, with seniors showing more trust and bonding towards the humanoid agent Billie.

While usefulness, ease of use and enjoyment were assumed to affect the participants' usage intention positively (*H9*), only perceived usefulness of the agent does lead to a significant increase of usage intention. In terms of the relevance of social processes in a human-agent interaction, bonding does predict the usage intention of the agent (*H10*) and further mediates the effect of trust and liking on usage intention (*H11*).

Based on the findings of Study 2, it was investigated in which way user characteristics affect the liking and usage intention (*RQ2*). Results show that the only user characteristic that affects liking is age, while usage intention is positively affected by prior general attitudes towards virtual agents. The older users are the more do they like the agent and when persons think more positive over virtual agents in general, they state higher usage intentions after the interaction with a virtual agent.

Table 63
Summary over all results and decisions for hypotheses and research questions

Hypotheses & Research Questions	**Independent Variables**	**Dependent Variables**	**Results**	**Decision**
Hypothesis *H1*				
H1a	Attractiveness	Person Perception	Attractive agents were evaluated as more attractive (successful manipulation)	Partly supported
H1b	Attractiveness	Liking	Attractive agents are more liked than unattractive	Supported
H1c	Attractiveness	Usage Intention	No differences	Not supported
H1 d	Attractiveness	Heath Intention	Attractive agents evoke higher health intention	Supported
Hypothesis *H2*				
H2a	Attractiveness	Bonding & Closeness	No differences	Not supported
H2b	Attractiveness	Trust	No differences	Not supported
Hypothesis *H3*				
H3a	Humanlikeness	Person Perception	Humanlike agents are perceives as more realistic than machinelike agents	Not supported
H3b	Humanlikeness	Liking	No differences	Not supported
H3c	Humanlikeness	Usage Intention	No differences	Not supported
Hypothesis *H4*				
H4a	Humanlikeness	Bonding & Closeness	No differences	Not supported
H4b	Humanlikeness	Trust	No differences	Not supported
Hypothesis *H5*	Number of Interactions (Time)	Bonding & Closeness	No differences for bonding Linear increase of closeness over time	Partly supported
Hypothesis *H6*	Number of Interactions (Time)	Trust	Quadratic trend, with trust decreasing over time	Not supported
Hypothesis *H7*				
H7a	Number of Interactions (Time)	Person Perception	Quadratic trend of attractiveness (overall decrease) Linear increase of trustworthiness	Partly supported
H7b	Number of Interactions (Time)	Liking	Cubic trend with an overall enhancement of liking	Supported

H7c	Number of Interactions (Time)	Usage Intention	No differences	Not supported
H7d	Number of Interactions (Time)	Ease of Use	Linear enhancement of ease of use	Supported
Hypothesis *H8*	Usefulness, Ease of Use Enjoyment	Usage Intention	Usefulness enhances user intention	Partly supported
Hypothesis *H9*	Bonding	Usage Intention	Direct effect of bonding on usage intention found	Supported
Hypothesis *H10*				
H11a	Liking (Mediated by Bonding)	Usage Intention	No effect of liking on bonding nor on usage intention→ no mediation	Not supported
H11b	Trust (Mediated by Bonding)	Usage Intention	Bonding mediates the effect of trust on usage intention	Supported
Hypothesis *H11*	Usage Intention	Usage Online Version	Usage Intention did not predict the usage of the online version	Not supported
Research Question *RQ1*	Number of Interactions X Appearance	Evaluation & Social Outcomes	Overall no interaction effects	
Research Question *RQ2*				
RQ2a	Attitude Anxiety Anthropomorphism Tendency	Liking	With increasing enjoyment liking increases With decreasing ease of uses liking increases No effects of user characteristics	
RQ2b	Attitude Anxiety Anthropomorphism Tendency	Usage intention	Negative attitudes towards social agents predict usage intention	

8.4 DISCUSSION

In the present Chapter, the aforementioned results are interpreted and discussed against the background of prior research. Implications for the design of virtual agents and the consequences with regard to social outcomes and for a human–agent relationship will be presented. Additionally, limitations and shortcomings of the study will be presented and open questions for future studies are posed. Finally, a conclusion outlines the study goals and outcomes.

8.4.1 Results Summary and Interpretation

This study aimed to investigate the effects of species, realism, and embodiment on the perception and evaluation of agents in a human–agent interactions. Moreover, age-related differences in this perception and evaluation should be explored, since to date no study tested the impact of age on the perception and evaluation of appearance variables in a human–agent interaction. To close this research gap, a laboratory study was conducted in which four different appearances (cartoon-stylized human, realistic human, cartoon-stylized robot, and voice only) were tested in a between-subjects design ($N = 130$) with two different age groups (students and seniors).

Evidence regarding the effect of different appearance variables has already been presented in prior research (see Chapter II3.2) and in Study 1 and 2. Despite the contradictory findings of prior empirical research (e.g., Bergmann, Eyssel, & Kopp, 2012; Krumhuber, Kappas, Hume, Hall, & Aylett, 2015; Li, Kizilcec, Bailenson, & Ju, 2015; Terada, Jing, & Seiji Yamada, 2015), on the basis of the results of Study 2, participants were assumed to evaluate machinelike agents more positively than humanoid ones (H1). The results presented can partly support this hypothesis because for person perception, more precisely likability (H1a) and liking of the agent (H1b), the assumed effect was found. The machinelike appearance of Vince was evaluated as more likable and participants liked the agent more after interaction with this appearance compared with the humanoid appearances. Additionally, the machinelike appearance was perceived as more sociable than the humanoid one. However, no differences in the participants' usage intention (H1c) were noted. Overall, the results support the findings of Study 2 with regard to likability and liking. Once again, it was demonstrated that humanoid appearances are not necessarily worthwhile. The findings contradict the design of currently applied agents in commercial contexts, as a meta-analysis showed that the most commonly used agents are humanoid (Chattaraman et al., 2011). Therefore, when designing a

virtual agent, it should be kept in mind to consider machinelike agents. Nevertheless, as significant interaction effects with the age group were found, these results have to be interpreted with caution and the target group needs to be determined first before the agent is designed.

Realism is an often-discussed feature of an agent's appearance. While most research with regard to the uncanny valley (Mori et al., 2012) reports that realistic appearances will be evaluated less positively (e.g., McDonnell et al., 2012; Zell et al., 2015; Zibrek & McDonnell, 2014), under some circumstances and specific applications it was found that realistic appearances were evaluated more positively (Ring et al., 2014; Robertson et al., 2015; van Wissen et al., 2016). Since these applications match the present interaction topic and based on the findings of Study 2, it was assumed that a realistic appearance would be evaluated more positively than a cartoon-stylized appearance (*H2*). This assumption was not supported by the data and overall no differences in the perception and evaluation of the agent based on realism were found. Because perceived realism was one part of the person perception measurement, it might function for this specific comparison as a manipulation check. But since no differences between the two humanoid agents with regard to perceived realism were found, it can be argued that the manipulation was not successful, so that the assumed differences were not observed for the participants. As a consequence, realism cannot affect the perception and evaluation if no differences were observed. Moreover, it might be that participants evaluated the realism of the agent based on its behavior and that the behavioral realism suppressed the effects of appearance realism. Since in all conditions the agent presented the same behavior and interaction, no differences in behavioral realism were produced. Prior studies have already demonstrated the importance of behavioral realism (Bailenson et al., 2005; von der Pütten et al., 2010). Although behavioral realism was not the focus of the present work, it should be investigated further in future. On the one hand, because no differences in appearance realism were perceived, the manipulation might have been too subtle. On the other hand, large differences in appearance realism are needed to impact the perception and evaluation of a virtual agent in a human–agent interaction.

Regarding hypothesis H3, significant differences between embodied and non-embodied agents were found for the perceived competence of the agent (H3a) and sociability, but not for participants' liking (H3b) and usage intention (H3c). Additionally, the differences found contradict the assumption of H3 that embodied agents were evaluated more positively than non-embodied agents (c.f. Study 1), and the exact opposite was found. Participants rated the agent as more competent and sociable after an interaction with Voice-Only compared with

the other embodied agents. Overall, hypothesis H3 has to be rejected. The results may indicate that adding an embodied character lowers the perceived competence and sociability. One possible explanation might be the increasing familiarity with solely speech-based technologies (e.g., speech-based assistants like Siri or navigation systems). Users are more familiar with interacting with a voice only and this might affect competence and sociability. A second plausible explanation is that people idealize the agent in their heads when only a voice is presented. In line with this assumption, Riegelsberger and colleagues (2006) found that, in contrast to a voice recording, adding a picture to a profile of an online gamer lowered the overall evaluation made by other gamers and that they rejected potential partners based on the picture. This was not the case in the voice-only condition. In line with the findings of the present study, the authors emphasize that pictures in some circumstances can lower the evaluation of the interlocutor (Riegelsberger et al., 2006). However, if participants are assumed to idealize the agent in the voice-only condition then the perceived attractiveness of the agent and likability should also be higher in the voice-only condition. As this is not the case, the presence of the embodied characters seems to specifically lower competence and sociability. Again, these findings need to be consulted with caution, because interaction effects were found that could confound the main effects (Field, 2018).

Concerning these interaction effects, several target group with specific differences in participants' evaluations were assumed. Because there is scant prior research that tests the preferences of senior users with regard to an agent's appearance, based on the evidence of Study 1 and 2 (Straßmann & Krämer, 2017, 2018) it was assumed that seniors prefer a humanoid appearance over a machinelike one (*H6*). Moreover, with reference to the statements of the qualitative interviews in Study 1 where seniors stated they preferred an embodied agent, because an interaction with it is more familiar and they would like to have an interlocutor they can address (e.g., looking in the eyes) during the interaction, it was assumed that seniors evaluate an embodied agent more positively than a non-embodied agent (*H7*). Again, in line with students' interview statements, the opposite was hypothesized for students, since they instead reported preferring a solely speech-based system that is more ubiquitous and not restricted to a specific screen (*H8*).

Overall, the results show the same pattern for multiple dependent variables: Seniors evaluate Billie (the cartoon-stylized humanoid agent) more positively than students do, while no differences between the target groups with regard to the other appearances were shown.

Seniors perceive Billie and Voice-Only as more realistic than students do. These findings contradict the intended manipulation of realism, as it was assumed that Character

would be perceived as more realistic than Billie. In Study 2 it was shown that seniors rely more on the species of the agent and that realism is not a crucial variable for them. Therefore, the differences in perceived realism might be positive side effects of the generally more positive perception of Billie. Still, it is questionable why no differences in likability or the overall liking were found, while several other positives effects of a humanoid agent for seniors were obtainable. It might further be possible that participants rely more on the interaction and the agent's behavior for their realism evaluation and that in some way the interactions in the Billie and Voice-Only conditions were perceived as more realistic. However, the interaction itself was designed in a maximally controlled way, where only specific pre-scripted answers based on a decision tree were chosen from the wizard. But still small differences may have necessarily occurred as the agent was asked to respond to the specific answers of the participants. This limitation cannot be ruled out during an actual interaction study, where the agent should be perceived as responsive and relational. The findings cannot support hypotheses *H6a* and *H6b*, since only differences in perceived realism and not in any other variables of the agent's person perception were noted and no interaction effects of the target group and appearance for participants' liking of the agent were found.

In addition, Billie was perceived as more useful and more sociable by seniors and evoked higher usage intentions, trust, and bonding in the senior sample compared with students. Hence, hypothesis *H6c* was supported, as seniors showed higher usage intention after the interaction with the humanoid agent Billie compared with students. This assumption can also be extended to effects on social outcomes (*RQ1*). Seniors rated the agent's sociability higher for the cartoonlike humanoid agent Billie and also showed more trust toward and bonding with an agent with this appearance. These findings might be explained by the user groups' expectations. Seniors might expect a virtual agent not to be photorealistic, while students are more used to different forms of virtual characters. Thus, it is possible that the appearance of Billie matched the seniors' expectations the most. As predicted, they showed more positive outcomes for a humanoid agent. Therefore, these results indicate that a virtual agent applied in the context of daily-life assistance for people in need of support should be designed as a humanoid, since seniors showed, inter alia, higher usage intentions, more trust, and more bonding. In the aforementioned application field, these outcomes might be very beneficial, since a steady intense usage of the agent is aimed for. When the appearance is designed in a way that people's usage intention increases and where higher feelings of trust and bonding occur, this can foster a regular usage behavior.

Hypothesis *H4* assumed, based on the findings of Rosenthal-von der Pütten et al. (2018), that seniors evaluate the agent and the interaction therewith more positive than students do. This was partly supported for person perception (*H4a*) and supported for liking (*H4b*) and enjoyment (*H4e*), but not for seniors' usage intention (*H4c*) and perceived usefulness (*H4d*). In line with prior findings (Rosenthal-von der Pütten et al., 2018), seniors liked the agent more and enjoyed the interaction more than students did. As students probably have more experiences with the interaction with virtual agents and therefore have higher and more specific expectations, it is harder to match these expectations and impress them. Seniors by contrast can be assumed to have mostly not interacted with a virtual agent before and probably also have lower general points of contact with such technologies. Therefore, they seem to appreciate the technology and its functions more. Additionally, the application itself can be seen as more beneficial for seniors, who need more support in their daily life. Prior research already emphasized the influence of experience on the perceived usefulness of the technology and the intention to use it (Venkatesh & Davis, 2000). Although the interaction was designed to be suitable for both groups, the general application of daily-life assistance is more adapted to people in need of support. It was therefore assumed that seniors perceive the agent as more useful and show higher usage intentions. However, these assumptions were not confirmed by the data, since no differences between the two groups were found. Perhaps social desirability influenced the seniors' ratings. Qualitative research with elderly participants demonstrated that some elderly are afraid of asking and accepting help and that especially elderly women feel the reception of help as a loss of independence (Roe, Whattam, Young, & Dimond, 2001). Seniors (and especially women, who mainly participated in this study) might want to hide their potential need of support and want to maintain the illusion that they do not need any help (Roe et al., 2001). As a consequence, they might state lower usage intention and usefulness ratings than they actually had. In line with this, Yaghoubzadeh et al. (2013) found a third-person effect (Davison, 1983), and reported that the elderly perceive virtual assistants as useful, but not for themselves but for a third person. Hence, seniors might be frightened to be perceived as vulnerable and therefore state a lower usefulness and usage intention. Moreover, the seniors that participated in the study were required to visit the laboratory autonomously, which means that they needed to be mobile and healthy. Therefore, these seniors might not perfectly match the target group that is in need of support.

Nevertheless, with regard to the aforementioned interaction effects, seniors rated the agent's usefulness and their usage intention higher than students did when a specific appearance was presented. Seniors can be seen to perceive a humanoid agent as more useful

and prefer to use an agent with this appearance. In conclusion, it can be assumed that the appearance does have a specific effect on the target groups' perception of the agent's usefulness and their usage intention. Thus, to enhance the usage intention of seniors a virtual assistant should be designed with a humanoid appearance.

Furthermore, in line with hypothesis *H5,* students perceived the agent as easier to use compared with seniors. This might be explained by differences in technical skills between the two groups. Students are mainly described as digital natives, who are highly familiar with the use of technology, while seniors are not. In addition, seniors' self-efficacy with regard to these entities might be lower (Rosenthal-von der Pütten, Bock, & Brockmann, 2017). Users of the agent only need to talk to the agent and therefore the actual use of the agent should be equally easy for both groups, as no technical skills are needed. Nevertheless, it has been demonstrated that seniors are more skeptical (Scopelliti et al., 2005) and are assumed to have lower self-efficacy (Rosenthal-von der Pütten et al., 2017), and thus seniors seem to perceive the agent as less easier to use than students do.

Besides the effect of appearance, we was aimed to investigate which other cues drive the participants' usage intention. According to the acceptance models of technology (Davis, 1989) and assistive agents (Heerink et al., 2010), perceived usefulness, ease of use, and enjoyment of the interaction were assumed to enhance participants' usage intention (*H9*). The findings of this study can also support the effect of usefulness, where participants' usage intention increases with the perceived usefulness of the agent. Ease of use and enjoyment were not found to affect the usage intention. Moreover, in line with hypothesis *H10,* bonding was found to affect usage intention. The more people feel a bond with the agent, the higher their usage intention. This supports prior research from humans' close relationships, where feelings of intimacy and affiliation correlate with the time spent together (more conversations; McAdams & Constantian, 1983). Supporting the media equation theory (Reeves & Nass, 1996), these findings are transferable to human–agent relationships. Furthermore, bonding mediates the effect of trust and liking on usage intention (*H11*). When people like the agent more or trust it more, then they also state a higher bond between themselves and the agent and this leads to higher usage intention. Therefore, it can be assumed that bonding enhances usage intention. In therapeutic applications, bonding between participants and the therapist is used to enhance the outcome (Horvath & Greenberg, 1986). The same process can be used with regard to virtual assistance. As people in need of support might be skeptical of virtual assistance (Scopelliti et al., 2005), agents who evoke liking and trust and therefore lead to more bonding can enhance the usage.

In Study 2, effects of user characteristics, such as attitude or personality, on usage intention were found, and these effects where further explored in a human–agent interaction (*RQ2*). Contradicting the findings of Study 2, only the general former attitude toward virtual agents can predict the usage intention of the participants. People who had a generally positive attitude toward virtual agents reported higher usage intention after the interaction with the agent. Thus, overall, the attitude toward virtual agents should be enhanced, so as to foster a higher usage intention.

Overall, social processes can be used to enhance the usage intention and the evaluation of the agent. A key piece especially with regard to differences between target groups seems to be the agent's appearance, which can trigger a specific evaluation and social outcomes.

8.4.2 Limitations and Future Work

Although Study 3 offered valuable insights into appearance effects within a human–agent interaction and highlighted especially the impact of the target group, some shortcomings have to be discussed.

First, the small sample size and unbalanced distribution of age groups have to be noted. Too few seniors participated in this study. During the recruitment of participants, a balanced distribution of both groups was aimed at. However, it was challenging to find elderly volunteers who are able to participate in a laboratory study, as they need to be mobile and able to visit the university on their own. Furthermore, this study required the participants to have never participated in a study with a virtual agent before, in order to exclude confounding effects due to prior interactions and the former agent's appearance. Many volunteers stemming from the cooperating health-care foundation had already either participated in prior studies or had seen an interaction of a virtual agent. Thus, the recruitment of new senior participants was needed. Nevertheless, as this study already showed the influence of the target group, studies that investigate this effect more closely with the use of bigger samples are needed. Furthermore, other target groups that match the application field should be taken into account.

Although having never participated in a virtual agent study before was a requirement for the participants and was communicated during the recruitment phase, some of the participants especially the senior participants might have noticed the agent with the appearance of Billie before. Since multiple studies have been conducted in tandem with the health-care foundation where a big part of the senior sample stem from, they might have seen Billie on posters or other modes of presentation. Therefore, they might be more used to and

familiar with Billie's appearance than with the others and this might confound the present findings. The same might be true for students, although the chance that they noticed posters from prior studies is lower than for the senior sample. Nevertheless, only four seniors and 22 students reported that they had interacted with an embodied virtual agent before. Independent of their potential prior experience with the agent appearances, they might have been influenced by other (virtual) figure appearances, for example, seen in movies or computer games. Empirical evidence exists showing the effect of science fiction media on the expectations toward virtual entities (Sandoval, Mubin, & Obaid, 2014). Thus, the user's prior media experiences might trigger specific expectations also in regard to appearance that may influence the participants' evaluations. Based on these effects, it is nearly impossible to recruit an unbiased sample, as virtual figures and agents become increasingly important and are frequently embedded in our daily life.

The present work aimed to investigate the evaluation and perception of agents especially with regard to age-related differences in a human–agent interaction in a systematic manner. Accordingly, the agents' appearances were chosen with regard to species, realism, and embodiment. Yes, the appearances and agents that were used had to match certain technical criteria to be able to run an interaction study. Therefore, the appearances could not be manipulated as systematically as in Study 2 and have more variance than the presented factors (species, realism, and embodiment). Although stimuli with similar hair styles and clothing styles were chosen for the humanoid agents, they differ not only in their degree of realism but also, for example, in perceived age. Thus, other variables than the intended ones might influence the perception and evaluation of the agents such as the agent's feature specifications (age or sex). In addition, no significant difference with regard to perceived realism was found for the realism contrast. Thus, participants seem to perceive both humanoid agents as equally realistic. This means that the manipulation aimed for has to be obtained with caution. Future studies should design appearances that are systematically varied (see stimulus material of Study 2) and where other confounding variables are mostly eliminated. Additionally, only a limited number of appearance variables could be addressed in this experiment. Hence, other appearance variables and forms than the ones presented here are worth investigating. Because animals, for example, in Study 2 were found to have positive effects and differ in some evaluations from humans, these differences should be addressed in prospective research. Nevertheless, valuable insights were gained in this study especially into age-related differences in the evaluation of the agent and the evoked social processes with regard to different appearances.

In this study, not static material but an actual interaction with the agent was used. Within this interaction, the agent presented behavior related to the participants' responses. Despite the fact that the interaction was scripted and the agent's answers were limited to a certain set of possible reactions, this behavior might have influenced the interaction and the participant's evaluation of the agent afterwards. As the interaction was aimed to be enjoyable and relatable, the agent's answers had to rely on the prior statements of the participants. Thus, very small differences in the interaction and the agent's responses occurred. However, one cannot prevent these small differences if an actual relational interaction is required. The interaction topic itself could also be more suitable for some participants than for others. During the interaction design, a close contact with the staff of the health-care foundation was maintained to foster an interaction that is suitable and valuable for seniors as well as students. Nevertheless, some people might enjoy talking about health more than others. When a relational and also comparable interaction is needed, these effects cannot be excluded, as a specific topic needs to be chosen. However, people's attitude to the interaction topic itself has not been measured and possible moderation effects of this should be kept in mind for upcoming research.

Like most prior studies, the present study's results are also only based on a single interaction and one point of measurement. But the application used aims for a steady, longitudinal usage of a virtual assistant, to enhance autonomy and autonomous living (Kopp et al., 2018). Initial studies showed that the perception of the agent might be revised after time (Bergmann et al., 2012). Accordingly, it is questionable how the effects of appearance develop over time. Potentially, they are helpful during the first interactions to create an initial attraction and enhance usage intention. Nevertheless, it might be conceivable that afterwards the interaction and the agents' behavior are more important. Thus, long-term studies are needed to investigate the development of the presented findings in multiple interactions.

8.4.3 Conclusion

The present study investigated the effect of species, realism, and embodiment with regard to age-related differences on a human–agent interaction. Therefore, a between-subjects laboratory study with a Wizard-of-Oz setting was used, where four different appearances and two different age groups were examined. The interaction was embedded in virtual assistance and participants filled in a virtual health diary together with the speech-based agent. Results emphasized once again the importance of the target group, as age-related differences were found in the general evaluation and in the evaluation of the appearances. Seniors showed

higher usage intention, trust toward, and bonding with a humanoid agent (with a cartoon stylization) than students did. By contrast, overall machinelike agents were more likable and liked by the participants. The realism of the agent was not found to affect the evaluation. Moreover, social processes like bonding can enhance participants' usage intention and mediate the effect of liking and trust on usage intention. Thus, when a virtual assistant is designed the target group needs to be determined first. For seniors, a humanoid agent can be more appreciated, since it enhances usage intention and social processes.

9 STUDY 4: THE INFLUENCE OF INTERACTION FREQUENCY AND APPEARANCE FACTORS ON HUMAN–AGENT RELATIONSHIP

Consecutively, the last study tests the effect of appearance in a long-term interaction. In line with prior research (e.g., Bickmore, Gruber, & Picard, 2005) the findings of Study 3 demonstrated that social processes such as bonding between humans and agents occur. It was further shown that appearance has an effect on bonding. As bonding is a construct that is even more important in a longitudinal interaction, the present study aims (1) to investigate the development of appearance effects over time and (2) the effect of appearance and interaction frequency on the interpersonal relationship (such as bonding). Although the previous studies of this dissertation demonstrated that the target group is important, for this study no age-related differences are investigated. Conducting a long-term study involves enormous effort and participants need to visit the laboratory for several weeks, therefore only students were recruited, since it is difficult to find senior volunteers who are able to visit the laboratory over several weeks and it is even more difficult to make them enthusiastic about participating in such a study. However, this study offers relevant insights into the effect of appearance on the human–agent interaction over time.

9.1 BACKGROUND AND HYPOTHESES

As presented in the research background summarized in the theoretical part of this work (c.f. II3.2), the agent's appearance has already been found to impact the human–agent interaction. Appearance was found to have an effect on buying intention (e.g., Terada, Jing, & Seiji Yamada, 2015), motivation (e.g. Baylor, 2009a), learning outcomes (e.g., Baylor & Kim, 2004; Sträfling, Fleischer, Polzer, Leutner, & Krämer, 2010), and the overall impression and evaluation of the agent (e.g., McDonnell, Breidt, & Bülthoff, 2012; Ring, Utami, & Bickmore, 2014; Zibrek & McDonnell, 2014). To conclude, appearance is an important factor in the human–agent interaction. Nevertheless, the effects of appearance have not been investigated over a longer period. Bergmann et al. (2012) stressed that with regard to specific appearances, the impression of the agent can change over time. Therefore, it is highly important to explore the effects of appearance in a long-term human–agent interaction. This is especially important in terms of virtual assistance, since virtual agents are designed to interact longitudinally with the user. Virtual assistants should ubiquitously help the users mostly on a daily basis, for example, help to structure the day, maintain the user's calendar, and schedule appointments and remind the user of important things such as taking medications. Therefore,

it is important to investigate effects that occur between humans and agents during multiple interactions over time. On the one hand, long-term studies are highly informative, but on the other hand they are also time-consuming and often difficult to realize in a controlled way. There are major reasons for the lack of long-term studies in the research field of human–agent interaction.

Nevertheless, there are some studies (mostly by the research group led by Timothy Bickmore) that investigated human–agent interaction in the long term. First of all Bickmore, Caruso, Clough-Corr, and Heeren (2005) tested the acceptance and efficacy of relational designed agents. Senior citizens participated in the study for 2 months and either interacted with a relational designed agent or used print material instead. The authors' goal was to enhance participants' physical activities and for this reason they measured the steps walked with a pedometer. The agent was designed for daily use and asked participants to enter the number of steps walked, showed participants' former results and progress, and provided positive feedback to enhance the physical activity. Overall, participants interacted about 5–10 min daily with the virtual assistant. Results demonstrated that the relational agent could enhance seniors' physical activity, while the paper-based material could not. Moreover, it was shown that participants' in the relational agent condition felt some kind of bond that enhanced their motivation to be physically active (Bickmore, Caruso, Clough-Gorr, et al., 2005). A second study, in which participants either interacted over 30 days with a relational agent, a non-relational agent, or used paper-based material, investigated the effect on participants' bonding with the agent in the context of health behavior more deeply (Bickmore, Gruber, & Picard, 2005). In the relational agent condition, the communication of the agent included social dialogue, empathic feedback, meta-relational communication, humor, and nonverbal immediacy behavior. As assumed, results showed that participants in the relational group reported higher feelings of bonding compared with the non-relational group. The relational agent evoked also more liking and enhanced the intention to continue working with it.

In the context of learning, Doering, Veletsianos, and Yerasimou (2008) investigated the influence of a conversional agent on the learning outcome of students in a 4-week long-term study. The authors reported that participants felt limited support with the material by the agent during their class and that they were disappointed about the learning support achieved by the agent. By contrast, results further demonstrated that participants felt socially supported by the agent. Moreover, participants stated they wished for control over individual characteristics of the agent (Doering et al., 2008). From this study's results it can be derived

that it is important to adopt the users' preferences, such as preferences with regard to appearance factors.

Overall, these long-term studies highlight the impact of an interpersonal relationship between humans and agents. Nevertheless, these studies did not investigate the predictors of bonding and the impact of appearance. Study 3 showed that appearance can have an effect on bonding and trust, and thus this effect will be further examined in the present study. Evidence from close relationships between humans can further substantiate the findings of Study 3. In relationship formation and attraction, key predictors are similarity (Byrne, 1997; Montoya et al., 2008; Urber et al., 1998) and physical attractiveness (Poulsen et al., 2013; Walster et al., 1966). Both factors are taken into account for the current study, where the effect of the agents' humanlikeness and attractiveness are investigated over time.

As physical attractiveness is important in interpersonal relationships (Poulsen et al., 2013; Walster et al., 1966) and several other positive impressions are related to attractiveness (Khan & De Angeli, 2009; Khan & Sutcliffe, 2014; Sobieraj & Krämer, 2014), attractive agents are expected to evoke a more positive evaluation than unattractive agents and foster the interpersonal relationship. Consequently, it can be assumed that:

Hypothesis 1 (H1): Users will evaluate an attractive agent more positively than an unattractive agent.

H1 a): Attractive agents evoke a more positive person perception (more attractive, more likable, more trustworthy, more competent) than unattractive agents do.

H1 b): Users will like attractive agents more than unattractive agents.

H1 c): Attractive agents evoke higher usage intention than unattractive agents do.

H1 d): Attractive agents evoke higher intentions to live healthy than unattractive agents do.

Hypothesis 2 (H2): An attractive agent will influence the user more positively with regard to social outcomes than an unattractive agent will.

H2 a): Participants will have a stronger relationship with attractive agents than with unattractive ones.

H2 b): Participants will trust attractive agents more than unattractive ones.

The perceived similarity between a virtual agent and the user can be evoked by the agent's appearance. As virtual agents can easily take up various forms of species (e.g., humans, animals, robots, or objects) and degrees of realism (e.g., cartoon stylized vs. photorealistic) the appearance can be designed like a real human or in any other way. With regard to perceived similarity, in the current study the agent's humanlikeness is varied. While a realistic humanoid agent is assumed to evoke high feelings of similarity, users will probably see fewer similarities between themselves and a machinelike agent. Moreover, we chose to compare humanlike and machinelike agents in terms of long-term interactions, since in Study 2 and 3 machinelike agents were evaluated as more likable and users liked this appearance more. Thus, there might be a trade-off between humanoid agents unconsciously evoking more social outcomes because of their higher perceived similarity and machinelike agents that are more liked in general.

Therefore, based on the findings of prior research (Bergmann et al., 2012) and the results of Study 2 and 3, it is assumed that machinelike agents will be evaluated more positively than humanoid ones.

Hypothesis 3 (H3): Users will evaluate a machinelike agent as more positive than a humanoid agent.

H3 a): Machinelike agents evoke a more positive person perception (more attractive, more likable, more trustworthy, more competent) than humanlike agents do.

H3 b): Users like machinelike agents more than humanoid agents.

H3 c): Machinelike agents evoke higher usage intention than humanlike agents do.

On the other hand, with regard to social outcomes (such as bonding, closeness, or trust) and the unconscious process of social reactions, humanoid agents are assumed to evoke a more positive resonance based on the higher perceived similarity (Byrne, 1997; Montoya et al., 2008; Urber et al., 1998). In addition, empirical evidence indicates that demographic similarity is related to higher trust (Levin, Whitener, & Cross, 2006); thus, as humanlike agents are assumed to be perceived as more similar, users are expected to trust humanlike agents more than machinelike ones.

Hypothesis 4 (H4): A humanlike agent will influence the user more positively with regard to social outcomes than a machine agent will.

H4 a): Participants will have a stronger interpersonal relationship with humanlike agents than with machinelike ones.

H4 b): Participants will trust humanlike agents more than machinelike ones.

The impact of interaction frequency on the evaluation of the agent and the relationship between humans and agents has not been investigated to date. Thus, no experimental evidence exists. Nevertheless, according to the mere-exposure theory (Zajonc, 1968), participants' will evaluate the agent more positively over time. Moreover, as intimacy and affiliation between humans are correlated with time spent together, such as in joint conversations (McAdams & Constantian, 1983), and against the background of media equation theory (Reeves & Nass, 1996), it can be assumed that with an increasing number of interactions participants will have higher feelings of bonding, closeness, and trust. On this basis, it is expected that:

Hypothesis 5 (H5): Users' interpersonal relationship with the agent will increase with the number of interactions.

Hypothesis 6 (H6): Users' trust will increase with the number of interactions.

Hypothesis 7 (H7): Participants will evaluate the agent more positively over time.

H7 a): Participants will state a more positive person perception of the agent over time.

H7 b): Participants will like the agent more over time.

H7 c): Participants will state a higher intention to use the agent over time.

H7 d): Participants will perceive the agent's usage easier over time.

Bergmann et al. (2012) demonstrated that the impression of an agent could change depending on the agent's appearance. Participants relativized their impression of a robot appearance from the first impression to a second evaluation after a second period of interaction, while this was not the case for a humanoid agent. However, only one interaction with the agent was used and the first impression was measured after a few seconds. Thus, no real inferences for the present long-term study can be made, since the development of an agent's evaluation over multiple interactions is examined. However, the results of Bergmann et al. (2012) emphasize that (1) participants are assumed to revise their evaluation and impression of an agent over time and (2) that this might interact with the presented appearance of the agent. Therefore, the following research question is posed:

Research Question 1 (RQ1): How does the influence of appearance variables develop over time?

Against the background of technology acceptance (Davis, 1989) and acceptance of virtual assistants (Heerink et al., 2010), as in Study 3, it is assumed that:

Hypothesis 8 (H8): High ratings of perceived usefulness, ease of use, and perceived enjoyment increase the user's usage intention.

Moreover, the same hypotheses assumed in Study 3 are used, since social bonding was found to affect usage intention and mediate the effect of liking and trust on usage intention. These processes become more crucial in a long-term interaction, since bonding is assumed to change over time. Therefore, the following hypotheses are worth investigating:

Hypothesis 9 (H9): A user's bonding with the agent affects the user's usage intention positively.

Hypothesis 10 (H10): A user's bonding mediates the effect of the user's liking of the agent (*a*) and the user's trust toward the agent (*b*) on their usage intention of the agent.

Hypothesis 11 (H11): High usage intention will lead to higher usage behavior of the online agent.

Again, to acknowledge interpersonal differences and referring to the results of Study 2 and 3 the effect of user characteristics such as age, sex, attitude and personality are investigated.

Research Question 2 (RQ2): In which way do user characteristics such as attitude and personality traits influence (a) users' liking of the agent and (b) their usage intention?

9.2 METHOD

In the following Section, the experimental study will be explained.

9.2.1 Experimental Design

The main goal of the present study is to investigate the effects of appearance variables in long-term interactions. Prior research demonstrated that attractiveness (Poulsen et al., 2013; Walster et al., 1966) and perceived similarity (Byrne, 1997; Montoya et al., 2008; Urber et al., 1998) are key factors influencing bonding and closeness in human–human interactions. According to on the media equation theory (Reeves & Nass, 1996), the same principles and cognitive rules are applied in interactions with artificial entities and, thus, attractiveness and similarity are assumed to influence bonding and closeness in human–agent interactions. By contrast, prior studies demonstrated that machinelike appearances are evaluated more positively (Bergmann et al., 2012; Study 2), although machinelike agents can be assumed to be less similar to humans, which contradicts the similarity theory described above. Therefore, in terms of similarity, humanlike and machinelike agents are compared to explore these contradicting findings more. Moreover, attractive and unattractive agents were designed along human attractiveness stereotypes (c.f. Sobieraj & Krämer, 2014). To investigate the effects of these appearance variables on social outcomes in a long-term interaction, a 4-week longitudinal laboratory experiment ($N = 75$) was conducted. Appearance variables were manipulated in a 2 (attractiveness: attractive vs. unattractive) × 2 (humanlikeness: humanlike vs. machinelike) between-subjects design with five repeated human–agent interactions. The local ethics committee approved the procedure of the study.

9.2.2 Procedure

The current study is a long-term laboratory experiment where participants interacted over 4 weeks (five interactions) with a virtual agent. Participants visited the laboratory once a week for 4 weeks to interact and evaluate a virtual assistant. A fixed time slot (e.g., every Monday at 10 a.m.) was assigned to each participant, in order to have the same interaction intervals for all participants. Overall, each participant interacted five times on a regular basis with the virtual assistant.

All laboratory sessions had a questionnaire and an interaction part. For detailed information about the interactions and questionnaires, see descriptions in Chapters 9.2.3 and 9.2.4. In the following, the procedure of all laboratory sessions will be explained in detail. Since only the first and last session have a different procedure than the other sessions, Sessions 2–4 will be summarized.

Session 1

In the first laboratory session, participants were welcomed, the study content was explained to them, and informed consent was signed. A cover story was used, since participants should believe that they would interact over 5 weeks (six interactions) with the agent for the behavioral measurement. The experimenter told the participants that the main goal of the study is the evaluation of a virtual assistant system and its stability over multiple weeks. Participants were further told that they needed to interact for 5 weeks with the virtual assistant, that six laboratory sessions are needed, and that participants should use an online platform during these sessions. More details on the necessity of this instruction are given in Chapter 9.2.4. When this first briefing was finished, participants began with a questionnaire part in which control variables such as prior experiences and personality traits were captured. To obtain a baseline of the participant's preferences on the agent's appearance, 10 different pictorial stimuli (the four experimental conditions and six distracting stimuli) were evaluated. Participants were told that one of these 10 agents was randomly assigned to the participants. To prevent participants believing they could manipulate the agent assignment with their evaluation, a random generator was presented and participants were told that this generator randomly picks one of the 10 agents. However, this generator did not randomly choose one of the 10 stimuli, but was manipulated in the sense that the picture of the (chosen beforehand by the experimenter) experimental condition was presented. Although the experimenter already knew in which experimental condition the participant was, participants were asked to tell the experimenter the designation of the chosen agent (e.g., "Agent-08") to enhance the credibility of the cover story. When participants named the agent's designation, the interaction part with the agent began.

Participants had to change their position and sit in front of a big screen on which the agent was presented. Before the interaction began, the experimenter explained the whole interaction setup to the participants. Since participants should believe that they would interact with an autonomous agent, the experimenter showed the camera and explained that the agent will understand the participant via that camera using audio and video data. Furthermore, the experimenter explained that the participant would be left alone in the laboratory while he or she is interacting with the virtual agent to be less restricted by the experimenter's presence. Participants were asked to start the interaction by saying, "Hello Billie" when the system was booted and the agent appeared on the screen. Before the experimenter left the room, she started the camera from a second laptop and told the participant that she is starting the agent and that it will take a couple of minutes until the system is fully booted.

Since a Wizard-of-Oz setting was used for this study, the experimenter controlled the agent's behavior from a laptop in a neighboring room. Therefore, the experimenter switched rooms after the setup was explained and the camera was started. To control the agent, a pre-scripted dialogue system was used, where the experimenter responded to the participants' answers using either fixed (pre-scripted) buttons ("I can directly see that you are taller than me, but how tall are you exactly?") or could also type in some key information ("Ok, so you are about PARTICIPANTS_HEIGHT tall") to refer directly to the participant's statements. When the interaction was finished, the participant could call the experimenter back by ringing a bell.

After that interaction part, a second questionnaire part followed in which participants evaluated the agent and their interaction with it. When participants finished their evaluation, the first session was over, participants were reminded of the next session and the experimenter said goodbye.

Sessions 2–4

In the next session, participants began directly with the interaction part, since no control variables or prior measures were assessed. The interaction procedure was exactly the same as described for Session 1. Participants sat in front of the big screen, started the agent with "Hello Billie" after the experimenter had started the camera, switched rooms, and began to control the agent, so that the agent appeared on the screen. Again, when the interaction was completed, participants called the experimenter back by using a bell.

To evaluate the agent and the interactions, participants went to the questionnaire computer and filled in the questionnaire.

The online version of the agent, which was used as a behavioral measure (see Chapter 9.2.4) and to give more opportunities for interaction with the agent, was introduced in Week 2. Since at the beginning of Week 1 there were technical problems with the online platform, the online interaction started in Week 2. This was done to ensure that all participants had the same chances to use the online version for the same amount of time. Thus, in Week 2, participants received their login data for the online platform and got the instructions accordingly. To check whether the accounts and the platform were working correctly, participants should log in once during the laboratory session. Later, participants were asked to login as often as possible and the experimenter explained that the platform is important for the evaluation of the whole system. But since participants' usage frequency of the online agent functioned as a measurement, no fixed requirements were communicated. At the end of each

session, participants were reminded to use the online platform to enhance usage frequency in general.

Session 5

In the last session, participants again started with the interaction part and then made their evaluation by filling in the questionnaire. Besides the regular evaluations, the behavioral measurement was assessed in the sense that participants believed that after this session one additional session would follow (which was not the case) and that they could chose an agent for this session with their evaluation in the questionnaire (for details, see Chapter 9.2.4). Certain control variables such as the manipulation check and sociodemographics were captured after the evaluation. In the end, participants were debriefed, the experimenter explained the cover story, and participants received their incentives.

9.2.3 Independent Variables

The appearance of the virtual agent, with which participants interacted for 4 weeks, was varied with regard to attractiveness and humanlikeness. During the 4 weeks, participants had five laboratory interactions and had, additionally between laboratory sessions, the chance to have small online interactions from their homes.

Stimulus Material and Pretest

The current study focuses on social effects of appearance variables. Therefore, the appearance of the virtual agent was manipulated concerning attractiveness and humanlikeness. To design different appearances, the Character Generator by Autodesk (https://charactergenerator.autodesk.com) was used. In this software, different variables (e.g., skin color, eye size and color, body proportions and clothing) can be manipulated, but in a restricted way. Therefore, based on the possibilities of the software, different appearances varying in attractiveness were designed. Since the same attractiveness principles are applied to virtual characters as to humans (Sobieraj & Krämer, 2014), these principles were used to design attractive and unattractive humanlike agents (e.g., for attractive agents big eyes, full lips, high cheeks, and small chin).

The machinelike agents were designed according to the humanlike agents (attractive and unattractive). As mentioned earlier, the design possibilities of the tool were limited, thus, the machinelike agents still had humanlike characteristics, but instead of skin they had a more mechanical body (see *Figure 22*). Although these agents still had humanlike characteristics,

this was the most systematic way to produce experimentally controlled stimulus material that would be perceived as machinelike.

A pretest guaranteed that the stimuli used were designed and manipulated in the intended way. In an online survey with a 2 (attractiveness) × 2 (humanlikeness) × 2 (physiological version) × 2 (clothing) within-subjects design, 25 participants (21 women; M_{age}= 29.76, SD_{age}=9.70) evaluated 16 different stimuli (see *Figure 22*).

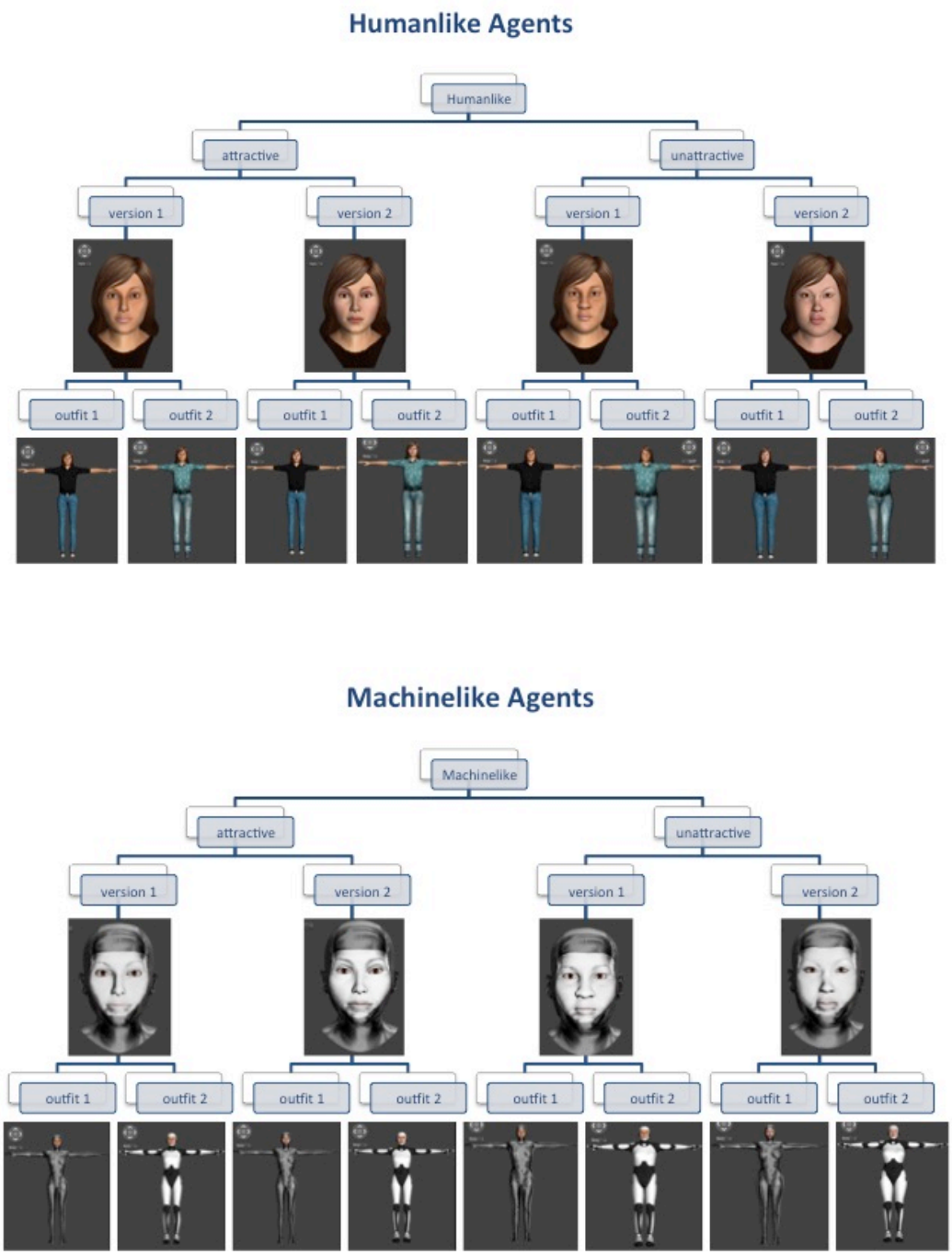


Figure 22. All stimuli variations used in the pre-test

Each participant evaluated all 16 stimuli on four single items regarding attractiveness, realism, humanlikeness, and willingness to interact: “The agent is attractive”; “The agent is realistic”; “The agent is humanlike”; and “I could imagine to interact with this agent.” A 10-point Likert scale was used ranging from 1 = *totally disagree* to 10 = *totally agree*.

Multiple repeated-measure ANOVAs revealed significant difference concerning attractiveness and humanlikeness for all four dependent variables (see *Figure 28*). Regarding the descriptive values, on the one hand it can be stated that attractive appearance was indeed evaluated as more attractive and participants reported a higher willingness to interact with these agents compared with unattractive agents. On the other hand, humanlike agents were perceived as more realistic and humanlike than machinelike agents. Therefore, the main manipulation of the stimuli could be described as successful.

Two different versions of each experimental condition were tested in order to find appearances that strongly differ with regard to attractiveness. However, analyses demonstrated that no significant difference between these two versions exist for any of the dependent variables. Thus, the versions with most extreme evaluations of attractiveness were chosen as stimuli for the main study.

In the beginning of the main study, participants are asked to evaluate all stimuli, and therefore we aimed to find different clothing for the agents to increase the variety and make the variation less salient to the participants. The repeated-measure ANOVAs showed that there was a significant difference between both clothing versions for participants’ willingness to interact with the agent. Participants were more willing to interact with an agent with Version 2 clothes. Further analyses showed that there was an interaction effect between clothes and humanlikeness, $F(1,24) = 17.09$, $p < .001$, $\eta_{part.}^2 = .416$, in the sense that the clothing of humanlike agents did not differ with regard to interaction willingness, but that for machinelike agents the Version 1 clothing ($M = 4.02$, $SE = .45$) evoked lower interaction willingness compared with Version 2 ($M = 5.63$, $SE = .49$). To avoid variance based on the clothing, for machinelike agents only the Version 2 clothing was used. But as mentioned above, little variety in the clothing was desired. Therefore, Version 2 clothing was slightly modified. The white parts of the outfit were colored in gray for the unattractive agent. *Figure 23* presents the final stimuli that were selected and used for the main study

.

Table 28
Results of the ANOVAs calculated in the pre-test

	The agent is attractive.			*The agent is realistic.*			*The agent is humanlike.*			*I could imagine to interact with this agent.*		
	M	*SE*	ANOVA	*M*	*SE*	ANOVA	*M*	*SE*	ANOVA	*M*	*SE*	ANOVA
Attractiveness			$F(1,25) = 142.07$, $p < .001$, $\eta_{part.}^2 = .850$			$F(1,25) = 60.79$, $p < .001$, $\eta_{part.}^2 = .717$			$F(1,25) = 35.13$, $p < .001$, $\eta_{part.}^2 = .584$			$F(1,25) = 67.79$, $p < .001$, $\eta_{part.}^2 = .739$
attractive	5.70	.32		5.60	.32		5.84	.32		6.59	.29	
unattractive	3.00	.29		4.49	.32		5.03	.32		4.75	.38	
Humanlikeness			$F(1,25) = 12.01$, $p = .002$, $\eta_{part.}^2 = .325$			$F(1,25) = 31.75$, $p < .001$, $\eta_{part.}^2 = .570$			$F(1,25) = 76.41$, $p < .001$, $\eta_{part.}^2 = .753$			$F(1,25) = 19.33$, $p < .001$, $\eta_{part.}^2 = .446$
humanlike	4.88	.32		6.22	.35		7.09	.34		6.51	.32	
machinelike	3.83	.33		3.87	.40		3.78	.40		4.83	.42	
Version			$F(1,25) = 1.04$, $p = .318$, $\eta_{part.}^2 = .040$			$F(1,25) = .75$, $p = .395$, $\eta_{part.}^2 = .030$			$F(1,25) = .77$, $p = .389$, $\eta_{part.}^2 = .030$			$F(1,25) = 1.54$, $p = .227$, $\eta_{part.}^2 = .060$
Version 1	4.41	.28		4.98	.30		5.34	.32		5.77	.33	
Version 2	4.29	.30		5.11	.34		5.48	.32		5.57	.32	
Clothing			$F(1,25) = .29$, $p = .593$, $\eta_{part.}^2 = .012$			$F(1,25) = 2.46$, $p = .130$, $\eta_{part.}^2 = .093$			$F(1,25) = .63$, $p = .435$, $\eta_{part.}^2 = .025$			$F(1,25) = 9.28$, $p = .006$, $\eta_{part.}^2 = .279$
Outfit 1	4.32	.30		4.95	.32		5.37	.33		5.33	.32	
Outfit 2	4.39	.29		5.14	.31		5.50	.33		6.01	.36	

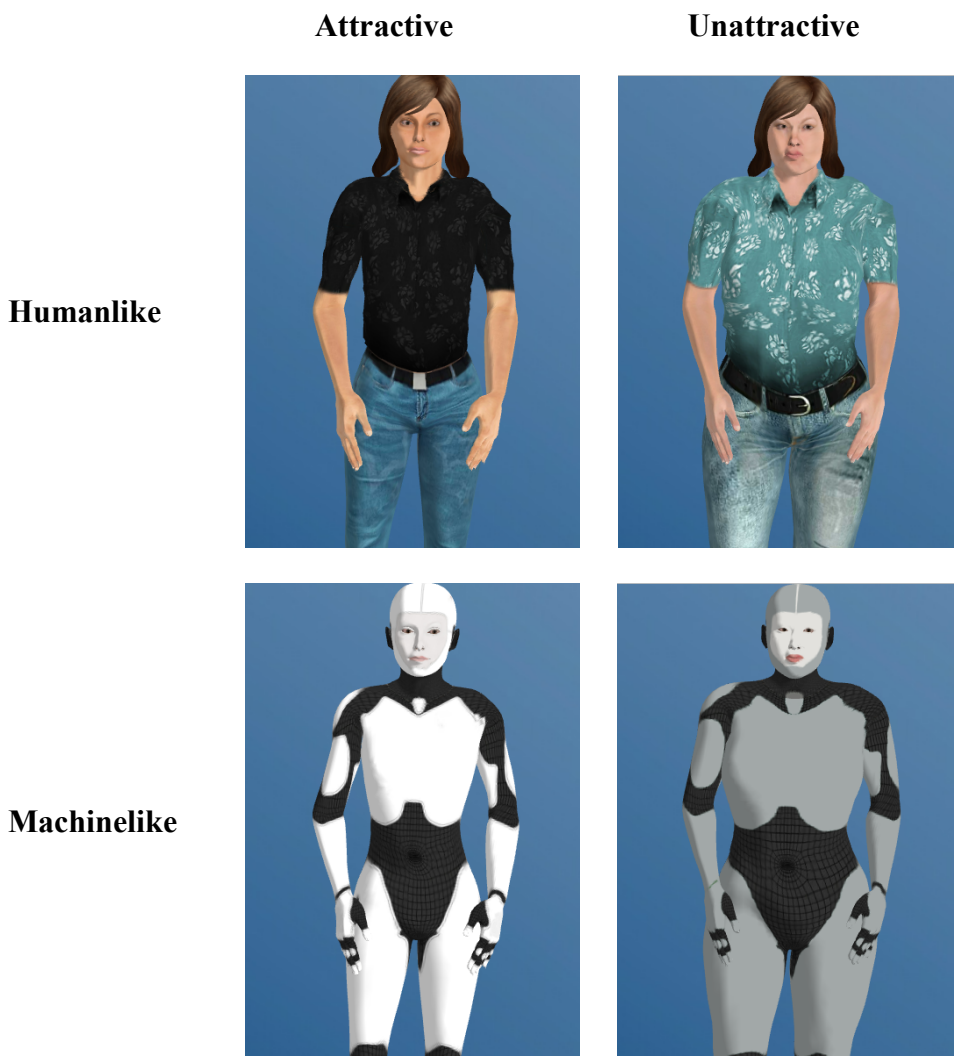


Figure 23. Final stimuli of the main study

Interactions

During the 4 weeks, participants used a virtual assistant in the context of healthy living. The context of healthy living was chosen because on the one hand it is highly relevant for all kinds of users and often discussed as a beneficial application field of virtual agents. On the other hand, it is very reasonable to interact over a longer period with an assistant in order to track progress. Participants had five laboratory interactions, where they talked with the virtual assistant about different topics influencing a healthy life. In addition, between the laboratory sessions participants had the chance to interact with their virtual assistant via an online platform. These interactions were short click-based interactions that are consistent with the laboratory interaction topics to enhance interaction frequency and to track participants' willingness to interact with the assistant as well as their health behavior.

Overall, five different topics regarding a healthy life were discussed: sport and fitness, diet, sleep behavior, drinking behavior, stressors, and negative factors. To avoid sequence effects, these topics were presented in two different orders:

1) Diet, sport and fitness, sleep behavior, drinking behavior, stressors, and negative factors

2) Sport and fitness, stressors and negative factors, diet, drinking behavior, sleep behavior

Analyses showed that the order in which the topics were presented had no influence on the main dependent variables (see Section Manipulation Check).

Laboratory Interactions

The laboratory interactions were speech-based. Although participants thought they would talk to an autonomous virtual agent who is able to understand and respond to their verbal inputs automatically, the experimenter in fact controlled the agent's reactions from a neighboring room. In this Wizard-of-Oz setting, participants talked with the virtual assistant about different factors of a healthy life. Each interaction contained six different parts: a welcome, a short recap of the last session (except for the first interaction), the user's preferences and habits regarding the present topic, a diary entry about the habits of the last week, a game, and a goodbye. These parts are described as follows:

- *Welcome*: In the beginning of each interaction, the agent welcomed the participant and started a quick social dialogue (e.g., "Hi, nice to see you again! How are you today?").
- *Recap last session*: To simulate a memory of shared conversations and actions, in the beginning the assistant referred to the last interaction (e.g. "In the last week, we talked about a healthy diet and I noticed that you are already experienced with it. Do you remember our guessing game? That was fun, wasn't it?"). After that, the agent shortly introduced the topic of the recent interaction.
- *Preferences*: Depending on the interaction topic, the agent queried the participant's overall preferences on the topic. The participant had to answer different questions where they could decide between two different options (e.g., "Do you like to exercise rather in the morning or in the evening?"). The agent responded appropriately and sometimes added useful information (e.g., "Some say that peanuts are healthier, since they have a lot of protein. However, this has nothing to do with calories").

- *Diary*: The diary part tracked the participant's actual habits in the past week regarding the different topics (e.g., "How often did you exercise last week?" or "How many liters of water did you drink daily in the last week?").
- *Game*: To enhance the entertainment of the interaction, various short games like quizzes or a guessing game were embedded in each interaction. These games were further used to convey more information about the recent topic (e.g., different sports).
- *Goodbye*: When the interaction came to an end, the agent referred to the online interaction (e.g., "I'm already curious to see how you'll implement physical activities into your daily life"), said goodbye, and asked the participant to call the experimenter back in.

On average, the interactions lasted 13 min (M = 12.93, SD = 2.17).

Previous studies demonstrated that a relational agent is evaluated more positively and users show more bonding with a relational agent, and therefore the assistant's dialogue was conducted in a relational way. As suggested by Bickmore and colleagues (2005), empathic feedback (e.g., "Good to hear! Yes, it is a beautiful day today!" or "I'm sorry about that! Yes, it is an odd day!"), social dialogue (e.g., "Have you talked to a virtual agent before our meeting?"), meta-communication (e.g., "Are you ready to start the game?"), and humor (e.g. "All right, I promise to handle your data carefully and that I'll not sell it to some third party. I'll leave this to Alexa") were embedded in the dialogue. Beside the relational design of the dialogue, the variability of the dialogue was found to be important for repeated interactions over a longer period (Bickmore & Schulman, 2009). We aimed to minimize repetitiveness by using different topics for each interaction (see above) and by randomizing the order of the interaction parts (diary, preferences, and game). The order of interaction parts within one topic was held constant, so that the interaction for one topic was the same for all participants. Therefore, the dialogues did not differ between the experimental conditions and all participants went through the same interaction (as mentioned earlier, only the order of topics was varied to avoid sequence effects).

To enhance the perceived liveliness of the agent, small nonverbal behaviors were embedded. Using audiovisual prosody, the agent moved his head according to his speech acts. Additionally, small beat gestures and pointing gestures (toward the user and the interaction content, e.g., diary table) were added. However, since the appearance of the agent was the focus of the current work, the nonverbal behavior should not be to salient and was held constant for all interactions and conditions.

In addition to the agent, tables depicted the user inputs or were used during the games when necessary. The agent was placed on the right side and if tables were used, they appeared on the left side. During the interactions, the agents did not move through the virtual room and stuck to his fixed position (see *Figure 24*).

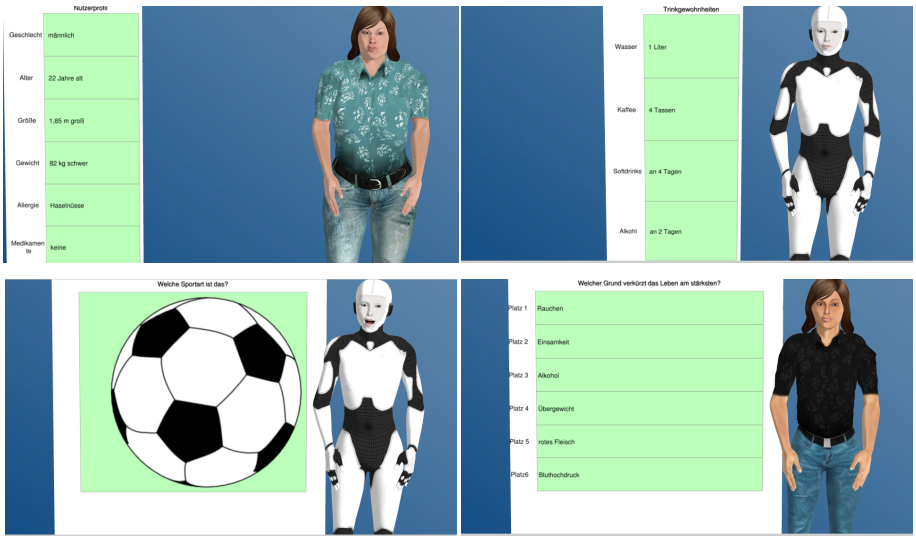


Figure 24. Examples of used tables and pictures during the interactions

Online Interactions

Each participant had an exclusive account for logging in to the platform. This account guarantees that each participant interacts with the assigned agent (depending on the experimental condition) and tracks the participant's inputs individually, which is used for the behavior measures. When participants were logged in, they had the chance to track their behavior regarding the week's interaction topic (e.g., after the laboratory interaction with the topic diet, the participant's eating habits were tracked). Since a speech-based interaction was not feasible for an online interaction, the agent communicated via speech (presented in video format, while users were told it was the same agent as in the laboratory interactions) and participants typed their answers using a keyboard. After participants responded via text and pressed enter, the agent automatically asked the next question. For each interaction topic, four different questions were used to track the user's habits (see Table 29). Although the questions for 1 week were constantly the same for each topic, to enhance variability two different versions of each question were constructed and used in randomized order and combinations. In the welcome and goodbye session, more than two different wordings were used and randomized over the 4 weeks, since there was no topic-based variety over the weeks.

Table 29
Scripted Dialogue used by the agent in the online interactions

	Version 1	Version 2
Welcome	Hello! On this page we can record some of your habits during the week. I will ask you some questions every day, which matches to the topic of the week. Hey! Great to have you back. Hello! Nice to see you back here. Hi, nice, that you logged in again. Welcome back to the health platform! Hey you, great that you're answering some questions again.	
Diet	Did you eat junkfood today? Did you eat candy today? How much fruit did you eat today? How many vegetables did you eat today?	Did you have fast food today? Did you have candy today? How much fruit did you have today? How many vegetables did you have today?
Sports	Have you been physically active today? How long have you been physically active today? Which sports did you do today? On a scale from 1 not exhausting at all to 5 very exhausting, how exhausting was that activity?	Did you strengthen yourself physically today? How long did you spend time on your physical activity? What kind of sports did you do today? Please rate this activity on a scale from 1 not exhausting at all to 5 very exhausting. How exhausting was it?
Drinking Behavior	How many liters water did you drink today? How many cups of coffee did you drink today? Did you drink alcohol today? How many glasses of other drinks did you drink today?	How many liters of water did you have today? How many cups of coffee did you have today? Did you have alcohol today? How many glasses of others drinks did you have today?

Sleep Behavior	When did you go to bed last night? When did you stand up this morning? Did you have a relaxing sleep? Did you wake up often?	At what time did you went to bed last night? At what time did you stand up this morning? Would you describe your sleep as relaxing? Have you been awake often during the night?
Stressors	Have you been in a stress situation today? Did you get worked up about anything today? Did you have a lot to do today? Were you in a hurry today?	Did you have a mad rush today? Have you angered by anything today? Did you have a long agenda today? Were you in a hectic pace today?
Goodbye	Bye Ciao Take care! Till then! See You later! See You tomorrow! Till the next time! Have a nice day! See ya! See you! Goodbye!	

Owing to technical problems in the beginning, the online interactions started in Week 2 in order to give each participant the same interaction possibilities.

9.2.4 Dependent Variables

As dependent measurements, a combination of self-reported data queried by questionnaires and additionally objective measurements were used to explore participants' actual behavior toward the assistant. Table 30 presents an overview of all the measures used and the frequency in which they were queried.

Table 30
Overview of all measures and its point of measurement

Scales	**Subscales**	**Points of Measurement**					
		T 0 Before Int1	*T1* After Int1	*T2* After Int 2	*T3* After Int 3	*T4* After Int 4	*T5* After Int 5
Person Perception	Realism		X	X	X	X	X
	Likability		X	X	X	X	X
	Attractiveness		X	X	X	X	X
	Trust-worthiness		X	X	X	X	X
	Competence		X	X	X	X	X
Bonding			X	X	X	X	X
Closeness			X	X	X	X	X
Interaction Evaluation	Enjoyment		X	X	X	X	X
	Ease of use		X	X	X	X	X
	Sociability		X	X	X	X	X
	Qualitative Statements						X
	Comparison to previous interactions			X	X	X	X
Usage intention			X	X	X	X	X
Perceived usefulness			X	X	X	X	X
Trust			X	X	X	X	X
Liking			X	X	X	X	X
Similarity				X	X	X	X
Intention	Questionaire		X				X
Health Behavior	Likert Scale Percent		X				X
User Charac-teristics	Socio-Demografics	X					
	Personality Traits (Big5)	X					
	Anthrophomor-phism Tendency	X					
	Anxiety	X					
	General Attitude	X					
	Health Behavior (Trait)	X					
Mani-pulation Check	Humanlikeness						X
	Machinelikeness						X
	Condition guessing						X
Loyality							X
Login Behavior online				Log-Files for all online interactions			

Questionnaires

The pivotal variables were measured at each laboratory session and thus five points of measurement exist. Those variables will be presented here. Please consult Table 31 for values of internal consistency.

Person Perception

To measure the evaluation of the agent, different subscales of person perception were used. Perceived *realism, likability, trustworthiness, competence,* and *attractiveness* were measured with 5-point semantic differentials (two opposite adjectives). These subscales were constructed based on different person perception measurements, for example, the Godspeed Questionnaire (Bartneck et al., 2009), attractiveness (McDonnell et al., 2012), and acceptance (Heerink et al., 2010) measurements. Seven adjective pairs were used to measure realism (e.g., *fake–natural*), seven for likability (e.g., *unfriendly–friendly*), four paired adjectives to measure attractiveness (e.g., *unattractive–attractive*), trustworthiness was measured with five pairs of adjectives (e.g., *not trustworthy–trustworthy*), and for competence again five paired adjectives (e.g., *incompetent–competent*) were used. Participants were asked to evaluate the agent using these five subscales after each interaction; thus, five points of measurements exist. The reliability of all scales to all points of measurement is acceptable (Cronbach's $\alpha > .781$).

Additionally, the *liking of the agent* and *perceived similarity* with the agent were assessed. Both constructs were measured with self-constructed scales. To assess participants' liking of the agent, five items were used (e.g., "I think, I would like this agent" or "I do not want to change something about the agent's appearance"), while perceived similarity was measured with three items (e.g., "Seems like the agent and I have a lot in common" or "Seems like the agent and I are pretty similar"). The reliability of both scales was acceptable for all points of measurement (Cronbach's $\alpha > .675$). Both were measured with a 5-point Likert scale ranging from 1 = *totally disagree* to 5 = *totally agree*.

To analyze the relationship between participants and agent based on the participants' prior preferences, participants were asked to evaluate all four experimentally manipulated appearances before the first interaction. Since the manipulation should not be too obvious for the participants, six distracting stimuli (pictures used in Study 2) were also evaluated, although this evaluation is not of interest for the present study. In order to keep the questionnaire brief, a short version of the person perception scales was used. All 10 stimuli picture were evaluated with regard to realism, likability, and liking of the agent. Realism was

measured with two items ("abstract" and "realistic"), while likability was captured with eight items (e.g., "likable"). Person perception for all stimuli was assessed before the first interaction and after the last interaction. The same liking scale (five items, e.g., "I think I would like this agent") used after the interactions was used to query participants' prior preferences with regard to the pictorial stimuli before the first interaction took place.

Interpersonal Relationship

The relationship between user and agent is of pivotal interest, thus, participants' *bonding* and *closeness* to the agent were measured. Therefore, the *bonding* sub-score of the Working Alliance Inventory (WAI; Horvath & Greenberg, 1986) with 12 items (e.g., "I feel uncomfortable with the agent" or "My relationship with the agent is very important to me") was used to assess the bonding between participants and the agent. A 5-pointed Likert scale ranging from 1 = *totally disagree* to 5 = *totally agree* was used for evaluation of these items. The internal consistency of this scale was acceptable for all points of measurement (Cronbach's α >.776).

Additionally, to measure *closeness,* the Inclusion of Other in the Self Scale (IOS; Aron, Aron, & Smollan, 1992) was used. Participants were asked to state their relationship to the agent ("How would you describe your relationship with the virtual agent? Please choose the figure that fits best to your relationship with the agent"). For this evaluation a 7-point pictorial scale was used, where circles that vary in their amount of shared space (starting with no overlap and getting closer until they are nearly totally overlapping) indicate the closeness between two interactants (see *Figure 25*). One circle symbolizes the participant and the other represents the agent. The original scale was slightly adapted in the sense that usually the pronouns *I* and *other* were used, but for the current study the wording was *I* and *agent*.

Figure 25. Pictorial scale used to measure participants' closeness to the agent

The perception of the assistant's social skills (*perceived sociability;* Heerink et al., 2010) was measured with four items (e.g. "I consider the agent a pleasant conversational partner" or "I think the agent is nice") rated on a 5-point Likert scale (1 = *totally disagree* to 5 = *totally agree*). Cronbach's α values demonstrate a good internal consistency (Cronbach's α >.814).

Interaction evaluation

In addition to the evaluation of the perception of the agent itself, participants were asked to evaluate the interaction quality. Therefore, two subscales from Heerink, Kröse, Evers, and Wielinga (2010) were used: *enjoyment* and *ease of use*.

To measure participants perceived *enjoyment* during the interactions, five items (e.g., "I enjoyed the agent talking to me" or "I find the agent enjoyable") were rated on a 5-pointed Likert scale ranging from 1 = *totally disagree* to 5 = *totally agree*. The reliability of this subscale was good for all points of measurements (Cronbach's $\alpha > .878$).

Participants' beliefs that the use of the agent is not associated with a high effort (*ease of use*) was measured with five items (e.g., "I think I will know quickly how to use the agent" or "I find the agent easy to use"), which were also evaluated with a 5-point Likert scale ranging from 1 = *totally disagree* to 5 = *totally agree*. The reliability of the first two interactions was very low (Cronbach's $\alpha = .512, .545$), but since for most of the points of measurement the reliability values are acceptable this subscale was kept for further analyses. However, related results needs to be interpreted with caution.

Participants were asked to evaluate each *interaction in relation to the previous ones* on a single item: "How would you describe this interaction in comparison with the others? If you imagine the previous interactions, how would you evaluate today's interaction?" (5-point Likert scale ranging from 1 = *much worse than the ones before* to 5 = *way better than the ones before*). Additionally, participants' *opinion* about the assistant and the interaction was queried with some open questions: "Which factors of the assistant should be improved?"; "Which factors appeared to be particularly negative?"; "Which factors appeared to be particularly positive?"; "Which factors were worse than the previous interactions?"; and "Which factors were better than the previous interactions?" Owing to the extensive number of data, these qualitative answers referring to the interaction evaluation were not analyzed in this work.

Usage evaluation

Again, two subscales of the acceptance scale for assistive social agents (Heerink et al., 2010) were used in the questionnaire to measure *perceived usefulness* and participants' *usage intention*. Each score comprised three items and was rated on a 5-point Likert scale ranging from 1 = *totally disagree* to 5 = *totally agree*. For *perceived usefulness,* items like "I think the agent is useful to me" were used and participants' *usage intention* was captured with items such as, "I think I'll use the agent during the next few days". Five points of measurement

exist for both constructs, since participants evaluated these items after each interaction. Good internal consistency was found (Cronbach's $\alpha > .902$).

Trust. Participants' trust in the virtual assistant was queried with two items (e.g., "I would trust the robot if it gave me advice"; 5-point Likert scale, 1 = *totally disagree* to 5 = *totally agree*) based on Heerink et al. (2010) after each interaction (five points of measurement). The reliability of this scale was evaluated as good for all points of measurement (Cronbach's $\alpha > .851$).

Because the aim was for a persuasive long-term effect of the interaction with the agent, participants' health behavior was only assessed at two points of measurement (at the beginning and in the end).

Health behavior

Since the virtual assistant aims to bring participants to a healthier lifestyle and teach them important factors for a healthy life, participants' health behaviors were measured with five items (e.g., "Eat a well-balanced diet" or "Eat fresh fruits and vegetables") from Moorman and Matulich (1993). Because this scale mostly contains behaviors regarding diet behavior, three items concerning physical exercises from Cunningham and Kwon (2003) were added (e.g., "I'm planning to be physical active on a regular basis next week"). All items were on a 5-point Likert scale ranging from 1 = *totally disagree* to 5 = *totally agree*. For the pretest–posttest comparison of participants' health behavior, we assessed once after the first interaction and once in the end after the last interaction to see how the behavior changed after regular use of the assistant. The internal consistency of all eight items was good for both points of measurement (Cronbach's $\alpha = .780, .784$).

Additionally, participants' behavior with regard to each interaction topic (diet, physical activity, sleep, and drinking) was queried with five items where participants stated the likelihood in percent (ranging from 0% to 100%) of whether they will adopt certain behavior (e.g., "get enough sleep" or "drink on a regular basis"). Again, a pretest–posttest comparison was desired; therefore, two points of measurements exist. However, since the internal consistency of both measurement points is rather low (Cronbach's $\alpha = .551; .523$), this scale was not included in further analyses.

Behavioral Measures

Loyalty

A special focus of the current study was the relationship between humans and agents. After each interaction, the participants' relationship was measured using self-reported data. However, since especially within human–computer interaction participants might not be able or willing to state their actual thoughts or feelings (c.f. Nass & Moon, 2000), it is important to combine subjective measures with objective ones. Therefore, a behavioral measurement was developed to check participants' loyalty toward the agents. Using a cover story, the experimenter told participants that the whole study lasts 5 weeks and six interactions are needed to participate, while there were in fact only five interactions. After interacting for 4 weeks with the same agent, in the last session of the experiment participants had the chance to choose their preferred agent for the next interaction (which was only part of the cover story). Therefore, participants built a ranking of all the different agents that they already evaluated in the beginning. Using this ranking, we could analyze whether participants show loyalty and ranked their assistant (the one they interacted with for 4 weeks) high for the following interaction. For the analyses, the ranking was coded in a binary variable, with 0 = *participants choose a new agent for the next interaction* and 1 = *participants stick to the former agent*.

Online interaction behavior

As described earlier, during the laboratory session participants had the chance to use an online version of the agent and fill in the health diary daily. The number of logins for each participant was tracked in order to see how often participants actually used the agent between the laboratory session. This number is employed as a behavioral measurement of usage. Moreover, participants' answers to the weekly questions were also tracked, but the qualitative data are not investigated in this work.

Table 31
Number of items and reliability values of all measured variables

Scales	**Subscales**		**Cronbach's Alpha's**				
			T1	*T2*	*T3*	*T4*	*T5*
Person Perception							
	Realism	7 items	.863	.851	.901	.923	.931
	Likability	7 items	.916	.920	.929	.935	.923
	Attractiveness	4 items	.781	.810	.860	.874	.880
	Trustworthiness	5 items	.821	.829	.858	.878	.822
	Competence	5 items	.888	.877	.861	.893	.874
Bonding		12 items	.796	.776	.837	.861	.854
Interaction Evaluation							
	Enjoyment	5 items	.893	.878	.929	.937	.888
	Ease of use	5 items	.512	.545	.765	.709	.814
	Sociability	4 items	.862	.822	.814	.863	.871
Usage intention		3 items	.909	.917	.932	.952	.961
Perceived usefulness		3 items	.902	.914	.930	.921	.938
Trust		2 items	.851	.928	.933	.908	.898
Liking		5 items	.797	.875	.827	.868	.862
Similarity		3 items	-	.921	.941	.957	.946
Intention Health Behavior							
	Questionaire	8 items	.780	-	-	-	.784
	Likert Scale						
	Percent	5 items	.551	-	-	-	.523
User Characteristics							
	Anthrophomorphism Tendency	20 items (10 each subscale)	Child: .864 General: .808 All: .871				
	Anxiety	4 items	.718				
	General Attitude	3 items	.787				
	NARS	6 items 5 items 3 items	S1: .566 S2: .529 S3: .547				

Manipulation Check and User Characteristic Variables

A manipulation check was used in the end to test a successful appearance variation. Therefore, participants stated humanlikeness and machinelikeness on single items using a Likert scale ranging from 1 = *totally disagree* to 10 = *totally agree*. After the cover story was partly revealed, participants should guess their experimental group, in order to check whether the appearance was perceived in the intended way.

Furthermore, participants were asked to write down their presumption about the study's hypotheses and research questions.

Beside these manipulation checks, several user characteristics were assessed, since they could be potential confounding variables (see online study). All control variables were measured only once (either in the beginning or after the last session).

Anthropomorphic tendency

The Anthropomorphism Questionnaire by Neave, Jackson, Saxtion, and Hönekopp (2015) was used to evaluate the users' tendency to anthropomorphize. With two subscales that contain 10 items each, the present behavior (e.g., "I sometimes wonder if my computer deliberately runs more slowly after I have shouted at it") and the behavior as a child (e.g., "When I was a child, I held birthday parties for my favorite toys") were assessed. Participants rated all 20 items on a 5-point Likert scale from 1 = *totally disagree* to 5 = *totally agree*. The internal consistency was good for all sub-scores (Cronbach's $\alpha > .808$).

Prior experiences

Participants' were asked about their knowledge about and prior experiences with virtual agents. Therefore, all participants had to state whether they know what a virtual agent is and were asked to define their understanding of virtual agents. Additionally, participants should answer whether they already interacted with a virtual agent, in which context the interaction was placed, and how the virtual assistant's appearance was shaped.

Attitude and anxiety

Three items were used to measure users' attitude toward agents (e.g., "I think it's a good idea to use a virtual agent") and four items assessed participants' anxiety toward agents (e.g., "If I should use a virtual agent, I would be afraid to make mistakes with it"). Both Scales originate from Heerink et al. (2010) and were evaluated with a 5-point Likert scale ranging from 1 = *totally disagree* to 5 = *totally agree*. Again, the internal consistency is acceptable for both anxiety (Cronbach's $\alpha = .718$) and attitude (Cronbach's $\alpha = .787$). In addition, the Negative Attitude Towards Robots Scale (NARS; Nomura, Kanda, & Suzuki, 2006) with its three subscales was adapted to the context of virtual agents. Three items measured participants' negative attitude toward situations of interaction with agents (e.g., "I would feel uneasy if I was given a job where I had to use virtual agents"), five items assed the negative attitude toward the social influence of agents (e.g., "Something bad might happen if virtual agents developed into living beings"), and three items measured participants' attitude toward emotions in interactions with agents (e.g., "I would feel relaxed talking with virtual

agents"). Again, participants rated all items with a 5-point Likert scale ranging from 1 = *totally disagree* to 5 = *totally agree*. Since the internal consistency of the three subscales is not satisfactory (Cronbach's $\alpha < .556$), results with regard to this measure have to be interpreted with caution. Nevertheless, owing to their relevance for the research topic, they were kept for the analyses.

Sociodemographics

Participants' sex, age, educational background, current profession, and field of study were queried after the last interaction.

Personality traits

Moreover, the short version of the Big Five Inventory (BFI-10) was used to assess participants' personality traits (Rammstedt & John, 2007). But since it is not the focus of the present research, this measure is not included in the findings presented here.

9.2.5 Participants

Overall, 75 students participated in the study for 4 weeks and completed all five interactions with the agent. Two runs of data collection were used (16.10.2017 – 17.11.2017 and 20.11.2017 – 22.12.2017), since each participant received a fixed time slot of 1 hr and, thus, not all 75 participants could participate within one run. In total, 42 participants signed in for the first run and 41 for the second one, but eight participants did not participate in all five laboratory interactions and were therefore excluded from the sample. In total, 375 hr and 10 weeks were spent in the laboratory to run the experiment.

All participants in the finale sample were German students of the University of Duisburg-Essen. Sex was nearly equally balanced, with 44 women and 30 men taking part (one participant made no statement about sex). The average age was 21 years ($M = 20.93$, $SD = 2.67$).

All participants were students and either had an A-level school diploma ($n = 73$) or a university degree ($n = 2$). Most people studied applied cognitive and media science ($n = 68$), five participants were psychology students, while two studied mechanical engineering and one participant was a math student. A larger part of 64 participants were at an early stage of their study program ($n_{\text{1st semester}} = 39$, $n_{\text{2nd semester}} = 1$, $n_{\text{3rd semester}} = 24$), while 11 participants had been studying for more than three semesters. This means that most of the participants were

not experienced with the participation in and performance of experimental studies, which is important in terms of the cover story and Wizard-of-Oz design.

Table 32 presents an overview of the sex and age distribution in the four experimental conditions. A simple ANOVA with age as dependent and experimental condition as independent variable showed no difference in age between the groups, $F(3,71) = 1.49$, $p = .225$, $\eta_{part.}^2 = .059$. To check whether sex was equally distributed in the experimental groups, a chi-square test was calculated and results show no significant differences of sex in the experimental groups, $\chi^2(3) = 1.43$, $p = .700$.

Table 32
Distribution of Sex and descriptive values of Age for the four experimental conditions

	Sex				**Age**	
	overall	women	men	n/a	*M*	*SD*
Humanlike Attractive	19	12	6	1	21.05	2.82
Humanlike Unattractive	18	9	9	0	20.78	3.10
Machinelike Attractive	19	11	8	0	20.05	1.51
Machinelike Unattractive	19	12	7	0	21.84	2.87
Overall	75	45	30	1	20.93	2.67

Furthermore, a one-factorial MANOVA was used to check whether the experimental groups differ in the control variables (Anthropomorphism Tendency, Personality Traits, Anxieties, and Attitudes and Health Behavior as a trait). Sine the results of the MANOVA (see Table 33) showed no significant differences in any of these variables for the experimental groups, Pillai's trace = .44, $F(33,189) = .99$, $p = .497$, $\eta_{part.}^2 = .147$, the randomization of the participants in the experimental groups was successful.

The descriptive values are presented in Table 34. Overall, participants showed a rather low tendency to anthropomorphize. With regard to personality traits, participants showed a moderate level of extraversion, agreeableness, conscientiousness, and neuroticism, while the level of participants' openness was rather high.

Participants stated rather low anxieties toward virtual agents and moderately negative as well as positive attitudes toward the use of virtual agents (see Table 34).

Table 33
Results of the MANOVA for the user characteristics and the experimental conditions as independent variables

Dependent Variable	Source	*df*	*F*	*p*	$\eta_{part.}^2$
Antropomorphism Tendnecy_Child	Condition	3	,18	,910	,008
	Error	71			
Antropomorphism Tendnecy_General	Condition	3	,37	,775	,015
	Error	71			
Antropomorphism Tendnecy_Overall	Condition	3	,33	,806	,014
	Error	71			
Extraversion	Condition	3	,33	,806	,014
	Error	71			
Agreeableness	Condition	3	,39	,758	,016
	Error	71			
Conscuentiousness	Condition	3	,84	,475	,034
	Error	71			
Neuroticism	Condition	3	1,26	,295	,050
	Error	71			
Openness	Condition	3	,41	,747	,017
	Error	71			
NARS	Condition	3	1,74	,168	,068
	Error	71			
Anxiety	Condition	3	1,68	,178	,066
	Error	71			
Attitude	Condition	3	,44	,727	,018
	Error	71			
Health Behavior	Condition	3	1,60	,197	,063
	Error	71			

Only nearly half of the participants reported knowing what a virtual agent is (*n* = 38). However, after a definition of a virtual agent was presented, 50 people stated they had already interacted with a virtual agent before. Most of the individuals (*n* = 36) who had already used a virtual agent interacted with an agent without any embodiment (e.g., Siri). When participants interacted with an embodied agent, they mostly described the agent as humanoid (*n* = 8). One person said they had used Ikea's Anna, and Microsoft's Clippy was used by two participants. In addition, one participant had interacted with the virtual Nao robot. Thus, the overall experiences with virtual agents were given and only 25 participants had no prior experiences with a virtual agent before they participated in the current study. Nevertheless, participants mostly used non-embodied agents and only a few of them (*n* = 14) had experiences with embodied agents.

Table 34
Descriptive Values of user characteristics among the experimental conditions

	Humanlike - attractive		Humanlike - unattractive		Machinelike - attractive		Machinelike - unattractive		Overall	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Anthropomorphism Tendency Child	2.14	0.79	2.14	0.75	2.27	0.89	2.08	0.82	2.16	0.80
Anthropomorphism Tendency General	1.72	0.71	1.67	0.62	1.74	0.61	1.54	0.52	1.67	0.61
Anthropomorphism Tendency	1.93	0.63	1.91	0.55	2.01	0.68	1.81	0.54	1.91	0.59
Extraversion	3.16	0.93	3.47	1.10	3.26	1.17	3.42	1.15	3.33	1.08
Agreeableness	2.97	0.92	3.08	0.79	3.29	0.96	3.13	0.96	3.12	0.90
Conscuen-tiousness	3.29	1.03	2.97	0.88	3.34	0.91	3.37	0.50	3.25	0.85
Neuroticism	3.34	0.83	3.39	1.01	2.84	0.91	3.05	1.18	3.15	1.00
Openness	3.97	1.06	3.75	0.99	3.66	0.96	3.92	0.99	3.83	0.99
NARS	2.80	0.37	2.92	0.55	3.11	0.44	3.03	0.39	2.96	0.45
Anxiety	1.82	0.61	2.06	0.74	2.37	0.86	2.05	0.81	2.07	0.77
Attitude	3.33	0.97	3.07	0.95	3.23	0.84	3.39	0.81	3.26	0.88
Health Behavior	3.32	0.82	2.90	0.74	3.40	0.73	3.31	0.67	3.24	0.75

9.2.6 Pilot Study

Since the method of this study is associated with enormous effort and because a reworking of the study or a second recruitment phase is not easily feasible, technical and

dialog-related bugs needs to be excluded from the beginning. Therefore, a small pilot study (*N* = 6) was run first to test the whole study setup.

The pilot study comprised six female participants with an average age of 21 years (*M* = 21.33, *SD* = 2.73). Instead of one interaction per week, the pilot study lasted only 1 week and participants interacted 5 days in a row (Monday to Friday) with the agent. Beside this adaptation, the same interactions and procedure as used in the main study were tested. Since the evaluation of the functions and study setup was the focus, only one version of the agent was used. Thus, all six participants interacted with the humanlike attractive appearance.

Participants ran through the same procedure every day and had a fixed time slot for their participation (e.g., each day at 10 o'clock). After being welcomed to the study and informed about the study content, they interacted with the virtual agent. To evaluate the quality of the interactions and the designed dialogue, a short questionnaire was presented later. An extract of the measures used in the main study was used for this questionnaire. Therefore, participants' prior experiences with and acceptance of virtual agents, interaction evaluation (enjoyment, ease of use, and sociability), evaluation of the agent, liking of the agent, and usage intention were queried using the same methods described earlier (see Chapter 9.2.4). Additionally, participants had the chance to fill in some open questions about faults within the dialogue and they had the chance to compare each interaction with the previous ones in a quantitative (using a single item: "How would you describe this interaction in comparison to the others? If you imagine the previous interactions, how would you evaluate todays interaction?" on a 5-point Likert scale ranging from 1 = *much worse than the ones before* to 5 = *way better than the ones before*) and qualitative manner (two open questions: "Which factors were worse than the previous interactions?" "Which factors were better than the previous interactions?"). Besides the self-reported data of the participants, the experimenter was obliged to detect faults and bugs within the interaction and wrote them down as well.

Based on these data, the interactions and dialogues were optimized for the main study. In general, the pilot study showed that the setup and method that were used worked very well and thus only small adaptations were made.

To check whether there were differences in the five interaction dialogues, multiple repeated-measure ANOVAs were calculated with the interactions as repeated measures and enjoyment, ease of use, sociability, and comparison with previous interactions as dependent variables. No significant differences between the interactions were found for any of the

dependent variables (see Table 35). Thus, although the interactions deal with different topics, they were perceived as equally enjoyable, easy to handle, and social.

Table 35
Results of the repeated measures ANOVAs for the evaluation of interactions

Dependent Variable	**Source**	***df***	***F***	***p***	$\eta_{part.}^2$
Enjoyment	Interaction	1.35	.55	.699	.100
	Error	6.75			
Ease of Use	Interaction	4	1.92	.156	.324
	Error	16			
Sociability	Interaction	4	.76	.564	.160
	Error	16			
Comparison with previous interactions	Interaction	2	.44	.656	.100
	Error	8			

Reviewing the mean values (see Table 36), the data indicate that the interaction was enjoyable, the agent was easy to use and sociable, and the interactions were evaluated as equal (not better or worse than the previous ones).

Table 36
Descriptive Values for the evaluation of the interactions of the pilot study

	T1		**T2**		**T3**		**T4**		**T5**	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Enjoyment	4.12	.69	4.04	.52	3.92	.54	4.24	.55	4.16	.67
Ease of Use	4.20	.45	4.28	.41	4.48	.52	4.60	.57	4.56	.54
Sociability	4.08	.52	4.13	.44	4.29	.37	4.25	.47	4.33	.56
Comparison with previous interactions	-	-	-	-	3.20	.84	3.60	.55	3.20	.84

As mentioned, all qualitative comments and the experimenter's notes were used to optimize the dialogue when possible.

Since the same person functioned as experimenter in the pilot study and the main study, the pilot study could additionally be seen as a training run for the Wizard-of-Oz part. Thus, in the main study the experimenter was trained, which led to a shorter reaction time for the agent's answers and, in addition, the chosen answers were more standardized, as the wizard was more experienced.

9.3 RESULTS

To investigate the influence of appearance and number of interactions on usage behavior, usage intention, ease of use and perceived usefulness, multiple mixed ANOVAs with 2 between factors (attractiveness and humanlikeness) and one within factor (number of interactions) as repeated measures was calculated. When the assumption of sphericity is violated, Greenhouse-Geisser correction was used. The effects of interaction frequency were further investigated with polynomial contrasts. As longitudinal data was used, these effects got further investigated using multilevel modeling with growth models (Field, 2018). Since these analyses yield the same effects than the mixed ANOVAs those results are not further presented in the following description.

9.3.1 Manipulation Check

To check the success of the manipulation, in the end of the study after the debriefing participants were asked to guess to which experimental condition they had been assigned. This guess was analyzed using a chi-square test. As depicted in Table 37 the actual experimental conditions differ with regard to the perceived experimental conditions significantly, $\chi^2(9, N=75) = 84.79, p < .001$.

Table 37
Results of the manipulation check for assumed experimental condition

Perceived Experimental Condition	Actual Experimental Condition				
	Humanlike attractive	Humanlike unattractive	Machinelike attractive	Machinelike unattractive	Overall
Humanlike attractive	13_a	2_b	0_b	1_b	16
Humanlike unattractive	2_a	15_b	2_a	5_a	24
Machinelike attractive	3_a	0_a	14_b	3_a	20
Machinelike unattractive	0_a	2_a	$3_{a, b}$	10_b	15
Overall	18	19	19	19	75

Overall, participants were able to guess their experimental condition correctly. But still 23 participants assumed another experimental condition than they have been actually assigned to.

Moreover, a MANOVA with humanlikeness as independent variable and perceived humanlikeness and perceived machinelikeness as dependent variables was calculated, to check whether participants perceived the agents in the intended way. Results show overall a significant difference between both experimental conditions, $V = .144$, $F(2, 70) = 5.89$, $p = .004$. In line with this the univariate analyses show a significant difference in perceived humanlikeness ($F(1, 71) = 7.54$, $p = .008$, $\eta_{part.}^2 = .10$) and perceived machinelikeness ($F(1, 71) = 10.95$, $p = .001$, $\eta_{part.}^2 = .134$) between humanlike and machinelike agents. As presented in Table 38, humanlike agents have been indeed perceived as more humanlike than machinelike agents and in return machinelike agents have been perceived as more machinelike than humanoid agents.

Table 38
Descriptive Values for perceived humanlikeness and machinelikeness

Experimental Condition	**Perceived Humanlikeness**		**Perceived Machinelikeness**	
	M	*SD*	*M*	*SD*
Humanlike agents	5.89	2.17	6.86	2.02
Machinelike agents	4.45	2.30	8.18	1.37
Overall	5.14	2.34	7.55	1.83

Nevertheless, as some participants had problems to guess their experimental condition correctly, it might be true, that those perceive the agent in a different way. To check whether people, who correctly assumed their experimental condition, perceive the agent's humanlikeness or machinelikeness differently than those who failed the manipulation check, a second MANOVA with the same dependent variables and the result of the prior manipulation check (1 = assumed right experimental condition, 2= assumed wrong experimental condition) was done. Here, no significant differences were found (multivariate test: $V = .002$, $F(2, 70) = 0.08$, $p = .926$; humanlikeness: $F(1, 71) = 0.01$, $p = .928$, $\eta_{part.}^2 = .000$; machinelikeness: $F(1, 71) = 0.13$, $p = .722$, $\eta_{part.}^2 = .002$) and it can be assumed that people did not perceive the agent differently when they stated a wrong experimental condition.

Thus, although some participants had troubles to state their experimental condition correctly, the manipulation was perceived in the intended way and the material can therefore be seen as successfully manipulated.

As described in the Chapter 9.2.3 to avoid sequence effects of the interaction topics, two different sequences have been used. To check whether the order of interaction topics has an impact on the main dependent variables a MANOVA with usage intention, usefulness,

trust, liking, realism, likability, attractiveness, trustworthiness, competence and bonding (all measured at the last interaction) as dependent and order of interactions topics (sequence 1 or 2) as independent variable was used. Results are presented in Table 39. According to Pillai's trace no significant multivariate effect of the interaction sequence occurred, $V = .165$, $F(10, 64) = 1.26$, $p = .271$.

Table 39
Results of the MANOVA with pivotal dependent variables and the order of interaction topics as independent variables

Dependent Variable	Source	*df*	*F*	*p*	$\eta_{part.}^2$
Usage Intention	Order of Interaction topics	1	2.45	.122	.032
Usefulness	Order of Interaction topics	1	1.72	.194	.023
Trust	Order of Interaction topics	1	0.12	.726	.002
Liking	Order of Interaction topics	1	0.44	.507	.006
Realism	Order of Interaction topics	1	0.14	.705	.002
Likability	Order of Interaction topics	1	1.04	.311	.014
Attractiveness	Order of Interaction topics	1	0.01	.933	.000
Trustworthiness	Order of Interaction topics	1	0.31	.582	.004
Competence	Order of Interaction topic	1	0.84	.364	.011
Bonding	Order of Interaction topics	1	0.08	.777	.001
	Error	73			

As a result in the following analyses the order of interaction topics is not further considered and both sequences were collapsed.

Humanlike agents are assumed to enhance the perceived similarity. To check whether this assumption is applicable to the used stimuli multiple ANOVAs with humanlikeness and

attractiveness as independent variables and similarity measured at the 4 points of measurement as dependent variables. Results revealed for no point of measurement a significant difference neither between humanlike and machinelike agents nor for attractive and unattractive agents. Additionally, no interaction effects of both variables occurred (refer to Table 40 for statistical terms). Hence, the assumption that humanlike agents evoke higher perceived similarity than machinelike agents is not confirmed. These results are discussed with regard to the overall implications of the study.

Table 40
Results of ANOVAs with attractiveness and humanlikeness as independent variables and similarity as dependent variables

	Source of variation	***df***	***F***	***p***	$\eta_{part.}^2$
Similarity Interaction2	Attractiveness	1	3.58	.063	.048
	Humanlikeness	1	0.12	.733	.002
	Attractiveness X Humanlikeness	1	0.27	.609	.004
	Error	71			
Similarity Interaction3	Attractiveness	1	1.34	.251	.019
	Humanlikeness	1	0.49	.487	.007
	Attractiveness X Humanlikeness	1	0.15	.702	.002
	Error	71			
Similarity Interaction4	Attractiveness	1	2.51	.117	.034
	Humanlikeness	1	0.39	.535	.005
	Attractiveness X Humanlikeness	1	0.01	.914	.000
	Error	71			
Similarity Interaction5	Attractiveness	1	1.13	.291	.209
	Humanlikeness	1	0.26	.611	.006
	Attractiveness X Humanlikeness	1	0.01	.935	.001
	Error	71			

9.3.2 Person Perception

With regard to hypotheses *H1a* (effect of attractiveness), *H3a* (effect of humanlikeness) and *H7a* (effect of number of interactions) the influence of appearance and number of interactions on the agent's person perception (realism, likability, trustworthiness, attractiveness and competence) has been analyzed. The statistical values of the multiple mixed design ANOVAs are presented in Table 41. Consecutively the results are presented for the different person perception variables.

Table 41
Results of mixed methods ANOVAs for person perception variables with attractiveness, humanlikeness and number of interactions as independent variables

	Source of variation	***df***	***F***	***p***	$\eta_{part.}^2$
Realism	*Between -subjects factors*				
	Attractiveness	1	3.34	.072	.045
	Humanlikeness	1	5.57	.021	.073
	Attractiveness X Humanlikeness	1	0.45	.833	.001
	Error	71			
	Within -subjects factors				
	Number of Interactions	2.76	1.03	.375	.014
	Number of Interactions X Attractiveness	2.76	0.31	.805	.004
	Number of Interactions X Humanlikeness	2.76	0.44	.711	.006
	Number of Interactions X Attractiveness X Humanlikeness	2.76	1.06	.363	.015
	Error	195.87			
Likability	*Between -subjects factors*				
	Attractiveness	1	3.57	.063	.048
	Humanlikeness	1	0.13	.724	.002
	Attractiveness X Humanlikeness	1	0.00	.967	.000
	Error	71			
	Within -subjects factors				
	Number of Interactions	3.36	2.43	.059	.033
	Number of Interactions X Attractiveness	3.36	0.73	.547	.010
	Number of Interactions X Humanlikeness	3.36	0.27	.869	.004
	Number of Interactions X Attractiveness X Humanlikeness	3.36	0.44	.750	.006
	Error	238.28			
Trustworthiness	*Between -subjects factors*				
	Attractiveness	1	2.70	.105	.037
	Humanlikeness	1	3.20	.078	.043
	Attractiveness X Humanlikeness	1	0.03	.856	.000
	Error	71			
	Within -subjects factors				
	Number of Interactions	2.79	30.19	<.001	.298
	Number of Interactions X Attractiveness	2.79	0.39	.746	.005
	Number of Interactions X Humanlikeness	2.79	1.26	.291	.017
	Number of Interactions X Attractiveness X Humanlikeness	2.79	0.53	.651	.007
	Error	197.87			
Attractiveness	*Between -subjects factors*				
	Attractiveness	1	18.72	<.001	.209
	Humanlikeness	1	0.40	.531	.006
	Attractiveness X Humanlikeness	1	0.09	.771	.001

	Error	71			
	Within -subjects factors				
	Number of Interactions	3.01	3.12	.027	.042
	Number of Interactions X Attractiveness	3.01	1.01	.389	.014
	Number of Interactions X Humanlikeness	3.01	1.92	.127	.026
	Number of Interactions X Attractiveness X Humanlikeness	3.01	1.71	.166	.024
	Error	213.83			
Competence	*Between -subjects factors*				
	Attractiveness	1	2.80	.099	.038
	Humanlikeness	1	1.61	.209	.022
	Attractiveness X Humanlikeness	1	0.56	.459	.008
	Error	71			
	Within -subjects factors				
	Number of Interactions	2.71	0.73	.520	.010
	Number of Interactions X Attractiveness	2.71	0.65	.567	.009
	Number of Interactions X Humanlikeness	2.71	1.68	.179	.023
	Number of Interactions X Attractiveness X Humanlikeness	2.71	0.37	.756	.005
	Error	192.35			

A significant difference between humanlike and machinelike agents was found for perceived realism, while no effect of attractiveness and no significant interaction effect of both appearance variables have been found. Independent of the number of interactions or the agent's attractiveness humanoid agents (*M*= 2.90, *SE*= 0.12) were perceived as more realistic compared to machinelike agents (*M*= 2.50, *SE*= 0.12). Thus, once again the manipulation was found to be successful. Perceived realism did not differ between the interactions and no interaction effects between number of interaction and the appearance variables were found.

Also for varied attractiveness of the agent's appearances the manipulation can be seen as successful, since both attractiveness levels differed in participants' evaluated attractiveness of the agent. Attractive agents (*M*= 3.04, *SE*= 0.11) were indeed perceived as more attractive than unattractive agents (*M*= 2.35, *SE*= 0.11), while humanlike agents and machinelike agents did not differ in evaluated attractiveness. Additionally, no significant interaction effect of both variables was found. Participants rated the agents' attractiveness differently after the various interactions. A quadratic trend was found for attractiveness ratings referring to the evaluations after the five interactions, $F(1,71) = 7.64$, $p = .007$, $\eta_{part.}^2 = .097$. The agent's attractiveness decreased from the first interaction until the fourth interaction and increased from interaction four to the last interaction (see Table 42 and *Figure 26*).

Table 42
Descriptive Values for perceived attractiveness of the agent of all Point of measurements

	M	*SD*
Interaction 1	2.80	0.82
Interaction 2	2.73	0.79
Interaction 3	2.64	0.84
Interaction 4	2.60	0.88
Interaction 5	2.72	0.91

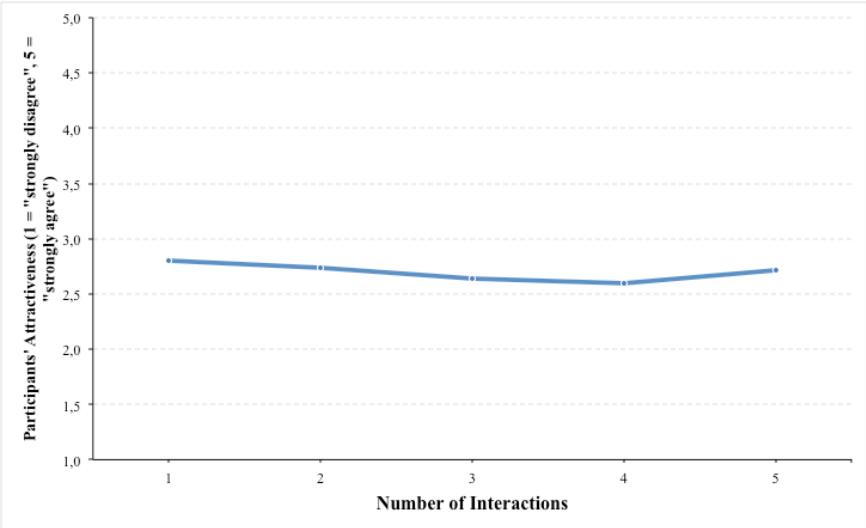


Figure 26. Development of perceived attractiveness over time

The agent's appearance (with regard to humanlikeness and attractiveness) did not differ in participants' perceived trustworthiness of the agent, while a significant main effect of the number of interactions was detected. Here a linear trend of trustworthiness was found with trustworthiness increasing together with the number of interactions, $F(1,71) = 63.26$, $p < .001$, $\eta_{part.}^2 = .471$. Consult Table 43 and *Figure 27* for means and standard deviations.

Table 43
Descriptive values of the agent's trustworthiness for all points of measurement

	M	*SD*
Interaction 1	3.16	0.78
Interaction 2	3.44	0.77
Interaction 3	3.63	0.77
Interaction 4	3.74	0.79
Interaction 5	3.91	0.72

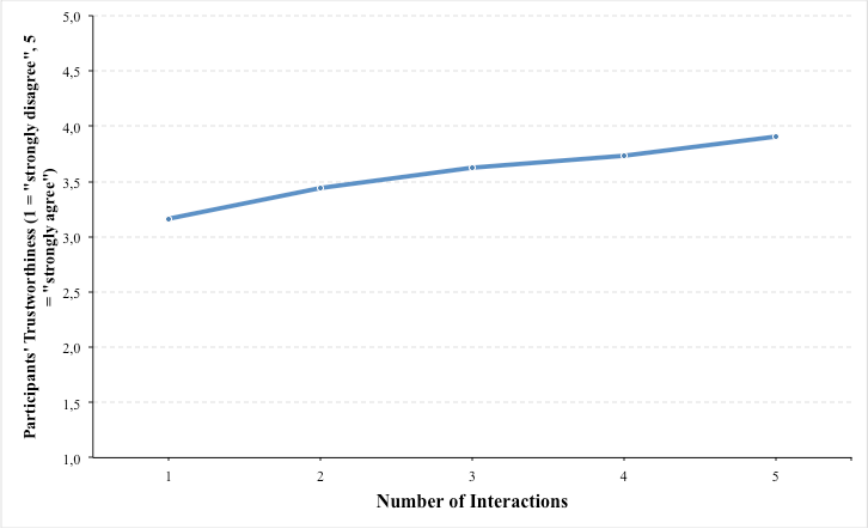


Figure 27. Development of perceived trustworthiness over time

Nevertheless, neither for likability nor for competence significant differences between the appearances or number of interactions have been found.

As only a significant difference between attractive and unattractive agents was detectable for perceived attractiveness, but not for the other person perception variables, hypothesis *H1a* has to be rejected. Also hypothesis *H3a* has to be rejected, since both experimental conditions only differ with regard to perceived realism and not with regard to the other variables such as likability, trustworthiness, attractiveness and competence. Nevertheless, the reported significant differences with regard to attractiveness and perceived realism support once again, a successful manipulation.

Referring to hypothesis *H7a* the number of interactions differ with regard to the agent's person perception and it was assumed that participants would evaluate the agent more positive over the time. While this was true for trustworthiness, where participants perceived the agent more trustworthy with increasing number of interactions, the opposite was found for attractiveness. The agents were rated as less attractive the more participants interacted with them. Additionally, no effects for realism, likability and competence with regard to the interaction number were found. Thus, overall hypothesis *H7a* is not supported by the data.

9.3.3 Liking

Again a mixed design ANOVA was calculated to analyze the effect of appearance (humanlikeness (*H1b*) and attractiveness (*H3b*)) and number of interactions (*H7b*) on participants' liking of the agent. While no significant interaction effects of appearance and number of interaction occurred, significant main effects of interaction number and attractiveness were found (see Table 44).

Table 44
Results of the mixed design ANOVAs for participants' liking of the agent

Source of variation	***df***	***F***	***p***	$\eta_{part.}^2$
Between -subjects factors				
Attractiveness	1	13.47	<.001	.159
Humanlikeness	1	2.91	.093	.039
Attractiveness X Humanlikeness	1	0.12	.732	.002
Error	71			
Within -subjects factors				
Number of Interactions	2.45	13.85	<.001	.163
Number of Interactions X Attractiveness	2.45	0.48	.661	.007
Number of Interactions X Humanlikeness	2.45	0.77	.488	.011
Number of Interactions X Attractiveness X Humanlikeness	2.45	0.92	.418	.013
Error	173.92			

Participants liking of the agent differs significantly between participants who interacted with an attractive agent to those who interacted with an unattractive agent. Overall, attractive agents (M = 2.30, SD = 0.88) were more liked than unattractive ones (M = 1.76, SD = 0.59). As participants like attractive agents more than unattractive, hypothesis *H1b* was supported. Participants who interacted with a humanlike agent did not differ significantly in their liking of the agent compared with participants' interaction with a machinelike agent. Therefore, hypothesis *H3b* that assumed participants to like machinelike agents more than humanlike ones has to be rejected based on the presented findings. Additionally, no significant interaction effect of humanlikeness and attractiveness was found.

To explore the effect of number of interactions, polynomial contrasts were used. Here, a significant linear ($F(1,71) = 12.44$, $p = .001$, $\eta_{part.}^2 = .149$), quadratic ($F(1,71) = 10.64$, $p = .002$, $\eta_{part.}^2 = .130$), cubic ($F(1,71) = 23.39$, $p < .001$, $\eta_{part.}^2 = .248$) and quartic ($F(1,71) = 12.97$, $p = .001$, $\eta_{part.}^2 = .154$) trend was found. See Table 45 for descriptive values..

Table 45
Means and standard deviations for liking from all points of measurement

	M	*SD*
Interaction 1	1.71	0.78
Interaction 2	2.18	0.79
Interaction 3	2.03	0.71
Interaction 4	2.09	0.82
Interaction 5	2.17	0.87

Referring to the effect sizes the cubic trend describes the data the best. Participants' liking raises from the first interaction to the second interaction, then drops to interaction three and increases till the last interaction (see *Figure 28*). Overall, from interaction one to the last interaction participants tend to like the agent more. Thus, hypothesis *H7b* can be seen as supported, since the liking of the agent increases over time.

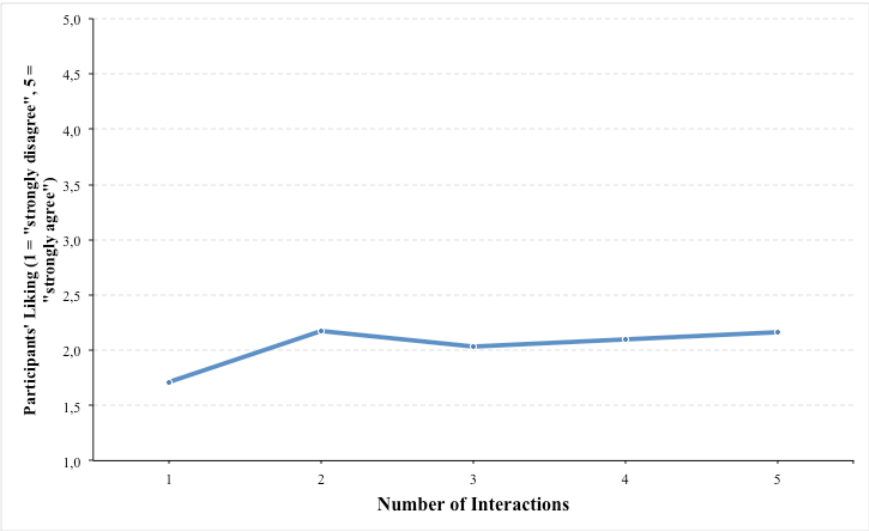


Figure 28. Development of participants' liking over time

Moreover, the influence of user characteristics on liking of the agent was investigated (*RQ2a*) running a hierarchical regression analysis with four steps. Here the same variables used in Study 3 have been chosen: Step 1: age and sex of the participants, Step 2: participants' enjoyment, ease of use and usefulness, Step 3 users' tendency to anthropomorphize (both subscales: Child and Adult) and Step 4: the influence of anxiety, attitude and negative attitude towards virtual agents on the liking behavior. Results of the regression analysis are depicted in Table X. While user characteristics were found to not affect participants' liking of the agent, the enjoyment of the interaction and perceived ease of use can significantly predict the liking of the agent. High enjoyment lead to higher ratings of liking and a negative relationship between ease of use and usage intention was found. A high ease of use seems to lead to a lower liking of the agent.

Table 46
Hierarchical regression analysis with liking as dependent variable

	Liking				
	b (*SE*$_b$)	*ß*	*t*	*p*	ΔR^2
Step 1					.002
Sex	-.08 (.22)	-.04	-0.35	.730	
Age	.01 (.04)	.04	0.29	.774	
Step 2					**.318**
Enjoyment	.27 (.13)	**.30**	2.07	.042	
Ease of Use	-.28 (.13)	**.30**	-2.21	.031	
Usefulness	.18 (.10)	.25	1.74	.087	
Step 3					.011
Anthropomorphism Tendency (Child)	.04 (.14)	.04	0.29	.774	
Anthropomorphism Tendency (Adult)	.13 (.16)	.09	0.80	.427	
Step 4					.033
Anxiety toward agents	.00 (.17)	.00	-0.00	.998	
Attitude toward agents	-.01 (.13)	-.01	-0.07	.941	
Negative attitude toward situations of interaction with agents	.20 (.21)	.15	0.96	.343	
Negative attitude toward social influence of agents	-.20 (.15)	-.16	-1.31	.698	
Negative attitude toward emotions in interaction with agents	-.11 (.16)	-.09	-0.70	.488	
Total R^2					.873
Step1	$F(2,72) = 0.80$, $p = .923$				
Step 2	$F(3,69) = 6.50$, $p < .001$				
Step 3	$F(2,67) = 4.73$, $p < .001$				
Step 4	$F(5,62) = 2.06$, $p = .003$				

9.3.4 Trust, interpersonal Relationship & Sociability

Again no main effects of the agent's attractiveness and humanlikeness on participants' trust toward the agent were found, while participants' trust varied significantly between the different interactions (see Table 47).

In hypothesis *H2b* it was assumed that attractive agents evoke more trust than unattractive ones, while hypothesis *H4b* assumed humanoid agents to lead to higher feelings of trust than machinelike ones. Unlike these assumptions, no main effects for attractiveness and humanlikeness were found and the hypothesis *H2b* and *H4b* have to be rejected. In line with the assumption of hypothesis *H6* that expected trust to increase over time, the mixed ANOVA showed a significant main effect for number of interactions. However, polynomial

contrast indicates a quadratic trend of participants' trust, $F(1,71) = 7.69$, $p = .007$, $\eta_{part.}^2 = .098$.

Table 47
Results of the mixed design ANOVA for trust with attractiveness, humanlikeness and number of interactions as independent variables

Source of variation	***df***	***F***	***p***	$\eta_{part.}^2$
Between -subjects factors				
Attractiveness	1	0.02	.888	.000
Humanlikeness	1	1.12	.293	.016
Attractiveness X Humanlikeness	1	0.06	.801	.001
Error	71			
Within -subjects factors				
Number of Interactions	3.26	3.11	.024	.042
Number of Interactions X Attractiveness	3.26	0.74	.539	.010
Number of Interactions X Humanlikeness	3.26	2.01	.108	.027
Number of Interactions X Attractiveness X Humanlikeness	3.26	0.65	.597	.009
Error	231.77			

Participants' trust decreased with increasing number of interactions till interaction 4 and then increased from interaction 4 to the last interaction (see Table 48 and *Figure 29*). So overall in contrast to the assumed increase (*H6*), participants' trust seems to decrease with rising number of interactions. As a consequence, hypothesis H6 has to be rejected.

As presented in Table 47 no interaction effect of the agent's appearance and the number of interactions was revealed by the analysis.

Table 48
Means and standard deviations for participants' trust for all points of measurement

	M	*SD*
Interaction 1	3.10	1.01
Interaction 2	2.89	1.10
Interaction 3	2.85	1.16
Interaction 4	2.79	1.15
Interaction 5	2.93	1.20

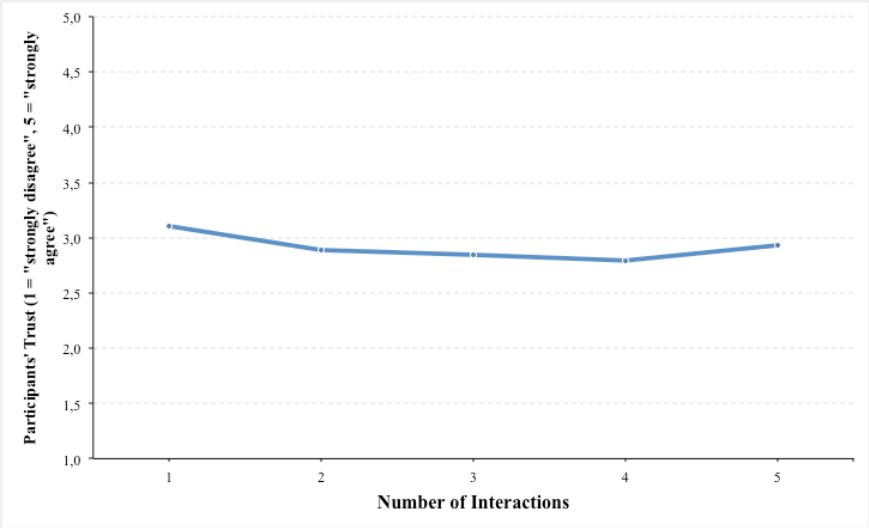


Figure 29. Development of participants' trust over time

Participant's stated bonding toward the agent neither varied between the different appearances nor between the various interactions (see Table 49). In addition, the analysis revealed no significant interaction effects of the independent variables. Thus, bonding seems to be neither affected by the agent's appearance (*H2a* & *H4a*) nor by the number of interactions (*H5*).

Table 49
Results of the mixed design ANOVAs for participants' bonding and attractiveness, humanlikeness and number of interactions as independent variables

Source of variation	***df***	***F***	***p***	$\eta_{part.}^2$
Between -subjects factors				
Attractiveness	1	1.02	.316	.014
Humanlikeness	1	0.22	.644	.003
Attractiveness X Humanlikeness	1	0.11	.747	.001
Error	71			
Within -subjects factors				
Number of Interactions	2.54	2.62	.062	.036
Number of Interactions X Attractiveness	2.54	0.99	.388	.014
Number of Interactions X Humanlikeness	2.54	0.21	.862	.003
Number of Interactions X Attractiveness X Humanlikeness	2.54	0.05	.975	.001
Error	180.53			

As a second measure of participants' interpersonal relationship towards the agent, the perceived closeness was measured and analyzed. Consult Table 50 for statistical values of the mixed design ANOVA with regard to closeness.

Table 50
Results of mixed design ANOVAs for closeness with attractiveness, humanlikeness and number of interactions as independent variables

Source of variation	***df***	***F***	***p***	$\eta_{part.}^2$
Between -subjects factors				
Attractiveness	1	0.10	.751	.001
Humanlikeness	1	0.12	.731	.002
Attractiveness X Humanlikeness	1	0.04	.851	.001
Error	71			
Within -subjects factors				
Number of Interactions	2.32	3.50	.026	.047
Number of Interactions X Attractiveness	2.32	0.63	.558	.009
Number of Interactions X Humanlikeness	2.32	0.57	.592	.008
Number of Interactions X Attractiveness X Humanlikeness	2.32	0.68	.529	.009
Error	164.72			

Participants stated neither different values of closeness for attractive or unattractive nor for humanlike or machinelike agents. Thus, again no impact of attractiveness (*H2a*) and humanlikeness (*H4a*) on participants' closeness was found and the corresponding hypotheses have to be rejected. The interaction of attractiveness and humanlikeness turned also not significant. But participants indicated different levels of closeness after the five interactions. With increasing number of interactions participants stated higher level of closeness to the agent and a significant linear trend was found, $F(1,71) = 4.65$, $p = .034$, $\eta_{part.}^2 = .061$ (see Table 51and *Figure 30*). In hypothesis *H5* it was assumed that participants would feel a stronger interpersonal relationship with increasing number of interactions. While this was not found for participants' bonding, the findings referring to closeness can support this hypothesis.

Table 51
Means and Standard Deviations for participants' closeness from all points of measurement

	M	*SD*
Interaction 1	2.17	1.10
Interaction 2	2.27	1.11
Interaction 3	2.20	1.16
Interaction 4	2.40	1.43
Interaction 5	2.48	1.48

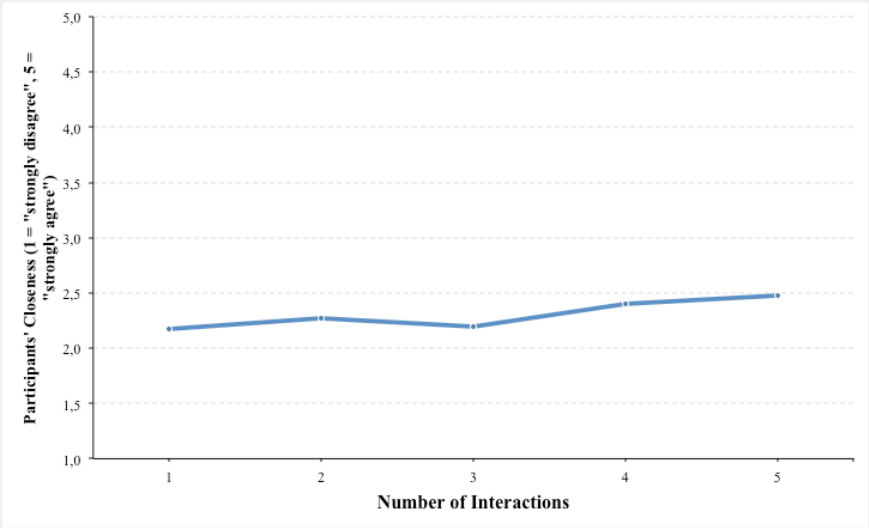


Figure 30. Development of user's closeness over time

As exploratory variables the agent's perceived sociability was analyzed. Results show that participants do not perceive attractive, unattractive, humanlike and machinelike agents as significantly different social, since no main effect for attractiveness or humanlikeness on the agent's sociability was found (see Table 52).

But the analysis revealed a significant main effect of the number of interactions on sociability. Referring to polynomial contrasts, a linear ($F(1,71) = 4.65$, $p = .034$, $\eta_{part.}^2 = .061$) and quadratic ($F(1,71) = 4.65$, $p = .034$, $\eta_{part.}^2 = .061$) trend of sociability was found. Means indicate that the perceived sociability of the agent decreases from the first to the fourth interaction and increases again from interaction four to five (see Table 53 and *Figure 31*). The number of interactions did not interact significant with the appearance variables.

Table 52
Results of the mixed design ANOVA for perceived sociability with attractiveness, humanlikeness and number of interactions as independent variables

Source of variation	***df***	***F***	***p***	$\eta_{part.}^2$
Between -subjects factors				
Attractiveness	1	1.53	.220	.021
Humanlikeness	1	0.44	.511	.006
Attractiveness X Humanlikeness	1	0.04	.842	.001
Error	71			
Within -subjects factors				
Number of Interactions	2.91	3.99	.009	.053
Number of Interactions X Attractiveness	2.91	0.46	.703	.006
Number of Interactions X Humanlikeness	2.91	0.73	.532	.010
Number of Interactions X Attractiveness X Humanlikeness	2.91	0.75	.519	.010
Error	206.63			

Table 53
Descriptive values for perceived sociability for all points of measurement

	M	*SD*
Interaction 1	3.55	0.91
Interaction 2	3.32	0.83
Interaction 3	3.31	0.84
Interaction 4	3.28	0.93
Interaction 5	3.31	0.97

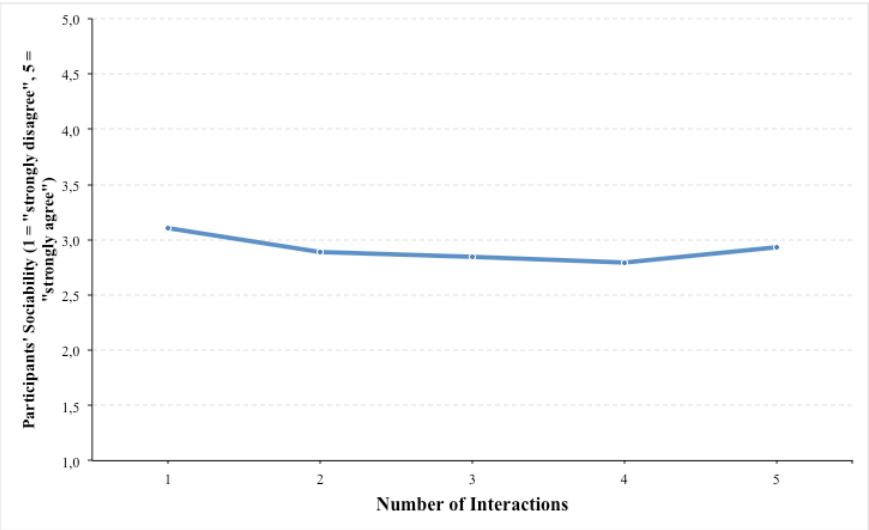


Figure 31. Development of perceived sociability over time

9.3.5 Usage Behavior, Usage Intention, Ease of Use, Perceived Usefulness

Effects of appearance (*H1c*, *H3c*) and number of interactions (*H7*) on usage intention were assumed. However, no significant effects of appearance neither for attractiveness nor for humanlikeness were found. Additionally, the number of interactions did also not affect participants' usage intention and no interaction effect of the independent variables was found (see Table 54). Thus, hypothesis *H1c* (where effects of the attractiveness of the agent on usage intention were assumed) has to be rejected and in line with this *H3c* (that assumed machinelike agents to evoke a higher usage intention) is also not supported by the data. Additionally, hypothesis *H7c* is also rejected, since no differences between the various interactions were found.

Table 54
Results of the mixed design ANOVA for usage intention with attractiveness, humanlikeness and number of interactions as independent variables

Source of variation	***df***	***F***	***p***	$\eta_{part.}^2$
Between -subjects factors				
Attractiveness	1	0.49	.486	.007
Humanlikeness	1	0.01	.918	.000
Attractiveness X Humanlikeness	1	0.44	.510	.006
Error	71			
Within -subjects factors				
Number of Interactions	3.14	0.99	.399	.014
Number of Interactions X Attractiveness	3.14	0.38	.774	.005
Number of Interactions X Humanlikeness	3.14	0.34	.805	.005
Number of Interactions X Attractiveness X Humanlikeness	3.14	0.14	.942	.002
Error	222.64			

Perceived usefulness was also not impacted by the experimental manipulations, as no significant main effects of attractiveness, humanlikeness and number of interactions were found and no interaction effects occurred. Consult Table 55 for exact values and outcomes of the mixed ANOVA.

The stated ease of use did not differ between the various appearances, since no significant effects of attractiveness, humanlikeness or interaction between both were found (see Table 56). In contrast participants' ratings of ease of use were found to differ significantly between the different numbers of interactions.

Table 55
Results of the mixed design ANOVA for perceived usefulness with attractiveness, humanlikeness and number of interactions as independent variables

Source of variation	*df*	*F*	*p*	$\eta_{part.}^2$
Between -subjects factors				
Attractiveness	1	1.10	.299	.015
Humanlikeness	1	0.06	.804	.001
Attractiveness X Humanlikeness	1	0.61	.436	.009
Error	71			
Within -subjects factors				
Number of Interactions	3.15	2.27	.078	.031
Number of Interactions X Attractiveness	3.15	0.58	.640	.008
Number of Interactions X Humanlikeness	3.15	0.24	.876	.003
Number of Interactions X Attractiveness X Humanlikeness	3.15	0.76	.524	.011
Error	223.88			

A significant linear ($F(1,71) = 41.53$, $p < .001$, $\eta_{part.}^2 = .369$) and quadratic ($F(1,71) = 15.41$, $p < .001$, $\eta_{part.}^2 = .178$) trend of number of interactions was found. In accordance with the effect sizes, the linear trend fits the best. With increasing number of interaction participants perceive the agent easier to use (see Table 57 and *Figure 32*). Thus, hypothesis *H7d* was confirmed.

Table 56
Results of mixed design ANOVA for ease of use with attractiveness, humanlikeness and number of interactions as independent variables

Source of variation	*df*	*F*	*p*	$\eta_{part.}^2$
Between -subjects factors				
Attractiveness	1	0.39	.536	.005
Humanlikeness	1	0.06	.802	.001
Attractiveness X Humanlikeness	1	0.00	1.000	.000
Error	71			
Within -subjects factors				
Number of Interactions	3.51	18.57	<.001	.207
Number of Interactions X Attractiveness	3.51	1.71	.156	.024
Number of Interactions X Humanlikeness	3.51	1.86	.126	.026
Number of Interactions X Attractiveness X Humanlikeness	3.51	0.36	.810	.005
Error	248.91			

Table 57
Descriptive Values for ease of use from all points of measurement

	M	*SD*
Interaction 1	3.97	0.65
Interaction 2	4.23	0.64
Interaction 3	4.35	0.66
Interaction 4	4.49	0.60
Interaction 5	4.44	0.71

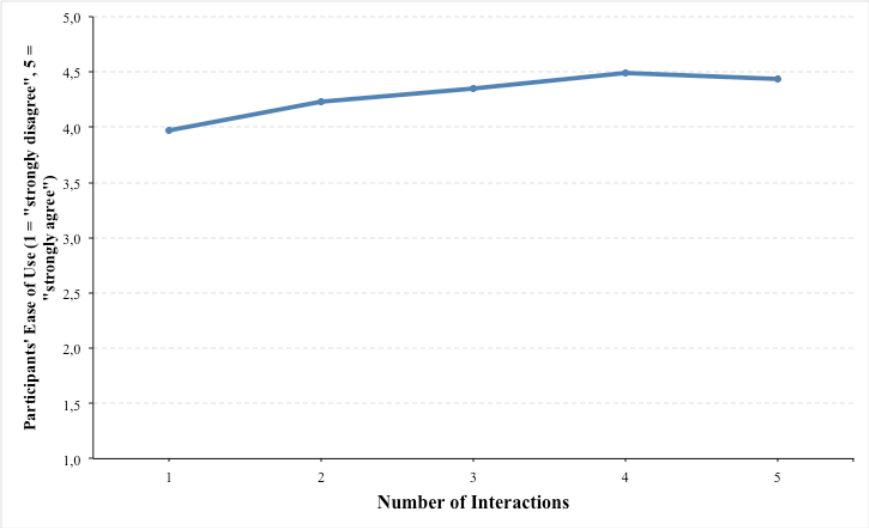


Figure 32. Development of ease of use over time

However, no significant interaction effects of number of interactions and both appearance variables were found (see Table 56).

To answer research question RQ2b a second multiple regression was calculated. First participants sex and age was entered in the model (Step 1), than enjoyment, ease of use and usefulness are added (Step 2). Additionally, both subscales of anthropomorphism tendency were includes in the third step and in the final step participants' general attitude toward virtual agents, their anxieties towards agents and the three dimensions of negative attitudes towards virtual agents are entered (Step 4). Results of the analysis are depicted in Table 58. Only two parameters of the users' characteristics were found to predict participants' usage intention significantly: perceived usefulness of the agent and a negative attitude toward social influence of virtual agents. The more useful the agent was perceived, the higher participants' usage intention at the end of the study was. Furthermore, when participants had high negative attitudes with regard to the social influence of virtual agents, than the usage intention at the

end of the study is lower, since a negative effect of negative attitude towards social influence on usage intention occurred.

Table 58
Hierarchical regression analysis with usage intention as dependent variable

	Usage intention				
	b (SE_b)	*β*	*t*	*p*	ΔR^2
Step 1					.010
Sex	.10 (.33)	.04	0.30	.763	
Age	-.05 (.06)	-.10	-0.84	.401	
Step 2					**.838**
Enjoyment	.17 (.09)	.12	1.80	.076	
Ease of Use	-.12 (.09)	-.07	-1.36	.180	
Usefulness	.88 (.07)	**.82**	11.97	<.001	
Step 3					.006
Anthropomorphism Tendency (Child)	.14 (.10)	.09	1.44	.154	
Anthropomorphism Tendency (Adult)	.01 (.11)	.00	0.08	.937	
Step 4					.020
Anxiety toward agents	-.11 (.11)	-.06	-0.96	.341	
Attitude toward agents	.17 (.09)	.11	1.88	.064	
Negative attitude toward situations of interaction with agents	.23 (.14)	.12	1.69	.096	
Negative attitude toward social influence of agents	-.22 (.10)	**-.12**	-2.21	.031	
Negative attitude toward emotions in interaction with agents	.02 (.11)	.01	0.16	.875	
Total R^2					.873
Step1	$F(2,72) = 0.36$, $p = .700$				
Step 2	$F(3,69) = 76.64$, $p < .001$				
Step 3	$F(2,67) = 55.55$, $p < .001$				
Step 4	$F(5,62) = 35.50$, $p < .001$				

Note. Values in bold indicate significant relationships.

Hypothesis *H11* assumes that usage intention affects the actual usage of the agent's online version (The frequency how often participants logged in to use the online version of the agent was used as a behavioral measure.). To investigate the effect of appearance on participants' actual usage behavior, a 2-way ANOVA with humanlikeness and attractiveness as independent and the number of online interactions as dependent variable was calculated. Results revealed no impact of appearance on the usage of the agent's online version neither for attractiveness ($F(1, 71) = 2.02$, $p = .160$, $\eta_{part.}^2 = .028$) nor for humanlikeness ($F(1, 71) = 2.02$, $p = .160$, $\eta_{part.}^2 = .028$).

Further the impact of participant's usage intention stated in the lab sessions on the usage of the online agent was analyzed running a linear regression with all five points of measurement of usage intention as predictors and number of online interactions as dependent

variable. Results demonstrate no significant impact of usage intention on the number of online interactions ($F(5, 69) = 0.77$, $p = .574$; see Table 59).

Table 59
Results of linear regression with usage intention of all measuring points as predictors and the number of online interactions as dependent variable

	Usage Online Interaction				
	b (SE_b)	*ß*	*t*	*p*	R^2
					.053
Usage Intention Int 1	-.48 (.72)	-.12	-0.67	.508	
Usage Intention Int 2	-.63 (.89)	-.17	-0.70	.484	
Usage Intention Int 3	.08 (.88)	.02	0.09	.933	
Usage Intention Int 4	-.38 (1.24)	-.11	-0.31	.758	
Usage Intention Int 5	1.51 (1.01)	.44	1.49	.141	

To examine effect of bonding on usage intention (*H9*) and its mediating effect of the influence of liking and trust on usage intention (*H10*), two mediation analyses were conducted. Again the INDIRECT macro from Preacher and Hayes was used with a 5,000 bootstrap resample (percentile-based 95% confidence interval). As all three variables were measured at all measurement points (after all five interactions), the outcomes of each variable were aggregated over time. According to Preacher (2015), this is a possible way, when no differences in the process between points of measures are assumed. Thus, the aggregated means were entered in the mediation analysis.

At first it was investigated whether bonding mediates the effect of liking on usage intention. No significant effect of liking on bonding was found ($b = 0.16$, $SE_b = 0.09$; $t = 1.72$, $p = .091$), while bonding influenced the participants usage intention positively, $b = 1.12$, $SE_b = 0.17$; $t = 6.77$, $p < .001$. Moreover, liking does also not affect usage intention, $b = -0.16$, $SE_b = 0.13$; $t = -1.18$, $p = .242$. Therefore, bonding cannot mediate the effect of liking on usage intention, as a significant effect of liking on bonding und usage intention is missing.

Running a second mediation analyses, the effect of trust on usage intention mediated by bonding was examined. Trust has a positive effect on bonding, $b = 0.44$, $SE_b = 0.05$; $t = 8.53$, $p < .001$. Additionally, trust ($b = 0.47$, $SE_b = 0.13$; $t = 3.50$, $p = .001$) and bonding ($b = 0.55$, $SE_b = 0.21$; $t = 2.58$, $p = .012$) have both a direct effect on usage intention. Consequently, bonding was found to mediate the effect of trust on usage intention significantly $b = 0.22$, $SE_b = 0.08$; 95% CI [0.06, 0.38].

With respect to the hypotheses, bonding was found to enhance usage intention, so that hypothesis *H9* was supported. While hypothesis *H10a* has to be rejected, *H10b* was supported, since bonding mediates the effect of trust on usage intention.

9.3.6 Enjoyment

Another mixed design ANOVA did neither reveal significant main effects of attractiveness or humanlikeness nor a significant interaction effect of both appearance variables on participants' enjoyment with the interaction (see Table 60).

Table 60
Results of mixed design ANOVA for enjoyment of the interaction with attractiveness, humanlikeness and number of interactions as independent variables

Source of variation	***df***	***F***	***p***	$\eta_{part.}^2$
Between -subjects factors				
Attractiveness	1	0.29	.592	.004
Humanlikeness	1	0.74	.390	.010
Attractiveness X Humanlikeness	1	0.20	.656	.003
Error	71			
Within -subjects factors				
Number of Interactions	2.78	7.54	<.001	.096
Number of Interactions X Attractiveness	2.78	1.19	.314	.016
Number of Interactions X Humanlikeness	2.78	0.59	.608	.008
Number of Interactions X Attractiveness X Humanlikeness	2.78	0.69	.549	.010
Error	197.45			

While no interaction between appearance and the number of interaction was found, participants stated different levels of enjoyments after the five interactions, as a significant main effect of number of interactions occurred. To investigate this effect further polynomial contrasts were carried out and indicated a significant linear ($F(1,71) = 6.13$, $p = .016$, $\eta_{part.}^2 = .079$), quadratic ($F(1,71) = 16.52$, $p < .001$, $\eta_{part.}^2 = .189$) and quartic ($F(1,71) = 6.14$, $p = .016$, $\eta_{part.}^2 = .080$) trend of enjoyment with regard to the number of interactions (see Table 61 and *Figure 33*). Participants' enjoyment decreased from the first interaction to the third and then increased from interaction 3 to the last interaction.

Table 61
Descriptive Values for enjoyment for all points of measurement

	M	*SD*
Interaction 1	3.70	0.90
Interaction 2	3.58	0.87
Interaction 3	3.25	1.00
Interaction 4	3.38	1.07
Interaction 5	3.48	0.96

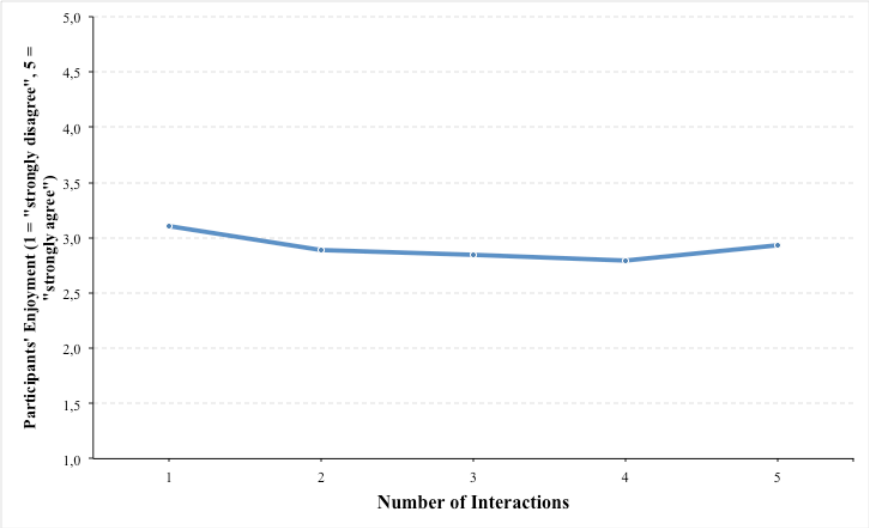


Figure 33. Development of enjoyment over time

9.3.7 Health Intention

Participants' intention the live a healthy lifestyle (e.g. eat a healthy diet or drink water) was measured as a pre-post comparison (after interaction 1 and 5). To investigate the assumption of *H1d*, that attractive agents lead to a higher intention to live healthy, a mixed design ANOVA was calculated. The mixed design ANOVA revealed a significant main effect for attractiveness, but not for humanlikeness and no significant interaction effect (see Table 62). Participants showed higher intention to live healthy after interaction with an attractive agent ($M = 3.93$, $SE = 0.09$) compared to interactions with an unattractive agent ($M = 3.53$, $SE = 0.10$). Nevertheless, no significant main effect of the interactions occurred and the interaction effect between appearance and number of interaction did also not turn significant.

Table 62
Results of the mixed design ANOVA for health intention with attractiveness, humanlikeness and number of interactions as independent variables

Source of variation	***df***	***F***	***p***	$\eta_{part.}^2$
Between -subjects factors				
Attractiveness	1	9.24	.003	.115
Humanlikeness	1	1.18	.281	.016
Attractiveness X Humanlikeness	1	0.00	.963	.000
Error	71			
Within -subjects factors				
Number of Interactions	1	0.60	.443	.008
Number of Interactions X Attractiveness	1	1.68	.199	.023
Number of Interactions X Humanlikeness	1	0.14	.707	.002
Number of Interactions X Attractiveness X Humanlikeness	1	1.57	.214	.022
Error	71			

9.3.8 Summary of the results and decisions for hypotheses

Please obtain the assumed hypotheses and research question together with the decisions based on the reported findings in Table 63.

In hypothesis *H1* was assumed that attractive agents receive a more positive evaluation, were more liked and evoke higher usage intention. Results of the analyses can support this assumption for perceived attractiveness, liking and evoked health intention, while no differences between attractive and unattractive agents were found. Overall, no differences between the appearances neither for attractiveness nor for humanlikeness in their evoked bonding, closeness and trust were obtainable. Thus, the hypotheses *H2* and *H4* have to be rejected. The agent's humanlikeness was expected to influence the evaluation of the agent, since in the prior studies machinelike agents were evaluated more positively compared to humanlike agents. The results of this study cannot replicate those findings and thus, hypothesis *H3* was also rejected by the data.

While the interaction frequency did not affect bonding, closeness increases over time. Thus, hypothesis *H5* was partly supported for closeness. In contrast, participants' trust did not change over time and hypothesis *H6* was also rejected. Additionally, time influenced the evaluation of the agent in some parts (*H7*). While attractiveness, decreases in a quadratic trend, trustworthiness increases linear from the first to the last interaction. In line with this, overall, liking was enhanced over time, but in a cubic trend. Participants stated the same

values of usage intention after all five interactions, but they perceived the usage of the agent easier over time. No interaction effects between both appearance factors and number of interactions occurred. As a consequence, the effect of time seems to be the same regardless of the agent's appearance.

Like in Study 3, the perceived usefulness of the agent was found to predict participants' usage intention, while enjoyment and ease of use did not. In addition, trust was found to have a positive effect on usage intention mediated by participants feeling of bonding. In contrast to the findings of Study 3, no such effect was found for participants' liking of the agent. Referring to the online login behavior of the participants, usage intention was not found to affect actual use (*H11*).

While no user characteristic was found to predict liking of the agent, former negative attitude toward the social influence of social agents affect participants' usage intention negatively.

Table 63
Summary over all results and decisions for hypotheses and research questions

Hypotheses & Research Questions	**Independent Variables**	**Dependent Variables**	**Results**	**Decision**
Hypothesis *H1*				
H1a	Attractiveness	Person Perception	Attractive agents were evaluated as more attractive (successful manipulation)	Partly supported
H1b	Attractiveness	Liking	Attractive agents are more liked than unattractive	Supported
H1c	Attractiveness	Usage Intention	No differences	Not supported
H1 d	Attractiveness	Heath Intention	Attractive agents evoke higher health intention	Supported
Hypothesis *H2*				
H2a	Attractiveness	Bonding & Closeness	No differences	Not supported
H2b	Attractiveness	Trust	No differences	Not supported
Hypothesis *H3*				
H3a	Humanlikeness	Person Perception	Humanlike agents are perceives as more realistic than machinelike agents	Not supported
H3b	Humanlikeness	Liking	No differences	Not supported
H3c	Humanlikeness	Usage Intention	No differences	Not supported
Hypothesis *H4*				
H4a	Humanlikeness	Bonding & Closeness	No differences	Not supported
H4b	Humanlikeness	Trust	No differences	Not supported
Hypothesis *H5*	Number of Interactions (Time)	Bonding & Closeness	No differences for bonding Linear increase of closeness over time	Partly supported
Hypothesis *H6*	Number of Interactions (Time)	Trust	Quadratic trend, with trust decreasing over time	Not supported
Hypothesis *H7*				
H7a	Number of Interactions (Time)	Person Perception	Quadratic trend of attractiveness (overall decrease) Linear increase of trustworthiness	Partly supported
H7b	Number of Interactions (Time)	Liking	Cubic trend with an overall enhancement of liking	Supported

H7c	Number of Interactions (Time)	Usage Intention	No differences	Not supported
H7d	Number of Interactions (Time)	Ease of Use	Linear enhancement of ease of use	Supported
Hypothesis *H8*	Usefulness, Ease of Use Enjoyment	Usage Intention	Usefulness enhances user intention	Partly supported
Hypothesis *H9*	Bonding	Usage Intention	Direct effect of bonding on usage intention found	Supported
Hypothesis *H10*				
H11a	Liking (Mediated by Bonding)	Usage Intention	No effect of liking on bonding nor on usage intention→ no mediation	Not supported
H11b	Trust (Mediated by Bonding)	Usage Intention	Bonding mediates the effect of trust on usage intention	Supported
Hypothesis *H11*	Usage Intention	Usage Online Version	Usage Intention did not predict the usage of the online version	Not supported
Research Question *RQ1*	Number of Interactions X Appearance	Evaluation & Social Outcomes	Overall no interaction effects	
Research Question *RQ2*				
RQ2a	Attitude Anxiety Anthropomorphism Tendency	Liking	With increasing enjoyment liking increases With decreasing ease of uses liking increases No effects of user characteristics	
RQ2b	Attitude Anxiety Anthropomorphism Tendency	Usage intention	Negative attitudes towards social agents predict usage intention	

9.4 DISCUSSION

The last study examined the effect of attractiveness and humanlikeness on the evaluation and perception of the agent and the social outcomes evoked by it in a longitudinal human–agent interaction. For this, a 2 × 2 between-subjects design of a long-term experiment with a Wizard-of-Oz setting for 4 weeks was conducted with 75 students. The results presented earlier will be summarized and discussed against the background of theoretical and empirical implications in the following Chapter. Afterwards, limitations of the study are mentioned and a summarizing conclusion emphasizing the main findings of the study is presented.

9.4.1 Results Summary and Interpretation

First, attractiveness was assumed to enhance the perception and evaluation of the agent and to evoke higher social outcomes, since such inferences can be made from the attractiveness stereotype found in human–human interaction (Walster et al., 1966) and for virtual figures (Khan & De Angeli, 2009; Sobieraj & Krämer, 2014).

The results emphasized that the manipulation of attractiveness was successful, with attractive agents being perceived more positively than unattractive agents. Moreover, attractiveness affected participants' liking and their reported health intention after the five interactions. Thus, people who interacted with an attractive agent were persuaded to live healthier after the interaction, while this was not the case for those interacting with an unattractive agent. This finding is in line with prior research that highlights a stronger persuasive effect of attractive agents (Khan & Sutcliffe, 2014). Based on a halo effect, attractive people were also assumed to have other positive capabilities (c.f. the "What is beautiful is good" stereotype; Dion, Berscheid, & Walster, 1972). Against the background of this stereotype, attractive agents were found to evoke a better evaluation, for example, with regard to social competence, social adjustment, and intellectual competence (c.f. Khan & De Angeli, 2009).

In addition, the body type of both kinds of agents might mostly cause this result. Attractive agents had an idealized body shape, while unattractive agents were designed heavier than ideal. People with heavier body types are usually seen as unhealthier, since overweight is associated with health problems. Thus, with regard to the health application, the attractive (thin) agent might be more persuasive, since it might be more credible.

However, because, in contrast to the assumed halo effect of attractiveness (Dion et al., 1972; Khan & De Angeli, 2009), no differences with regard to other person perception variables (except for attractiveness) were found, attractive agents cannot be assumed to be

more trustworthy or more competent in this regard. But the attractive agent may function as a role model. People might associate the attractive agent thanks to its thin body type with health and might wish to achieve the same status, thereby leading to higher health intentions. What was not investigated in this study is the match between the body type of the agent and the one of the participant. There is evidence that people perceive an agent with a heavier body as more credible and that they would like to interact with a heavier health assistant more than with an idealized one, especially when it matches the user's own body type (van Vugt et al., 2006). Although body weight was not measured in the present study, most participants seemed to have normal weight and no kind of the agent was perceived to be more similar to the participants. Therefore, the findings of van Vugt et al. (2006) cannot be applied to the present results, although they would be more crucial when heavy-weighted people are the target group.

Hence, from the findings of the present study it can be concluded that an attractive assistant in terms of health is more effective, since people showed higher health intentions after the interaction. Attractive agents were additionally more liked by the participants than unattractive agents. Thus, although with regard to other outcome variables no differences between attractive and unattractive agents were found, attractive agents seem to be more beneficial, since they evoked higher health intentions and were more liked. This is especially helpful when the agent is used as a personal assistant, since it might – through its attractive appearance – enhance the users' health behavior.

As a second factor of appearance, the agent's humanlikeness was investigated. With regard to this appearance factor, contradictory findings were considered. While previous studies demonstrated that machinelike agents were evaluated more positively (derived from findings of Study 2 and 3 and from Bergmann et al., 2012), based on the similarity attraction theory (Byrne, 1997) it has to be assumed that humanlike agents are perceived as more similar and will therefore evoke more positive outcomes. Referring to this contradiction, it was expected that the perception of the agent would be more positive for machinelike agents, while unconsciously processed social variables such as trust, bonding, and closeness might underlie the similarity attraction more. Thus, humanlike agents were perceived to evoke greater trust, bonding, and closeness.

None of these assumptions are supported by the findings of this study, since the humanlikeness of the appearance was only found to affect perceived realism (which was seen as support for a successful manipulation). Humanlike agents did not differ in other person

perception variables, liking, usage intention, trust, bonding, closeness, or any other variable. Therefore, the findings from the prior studies (Study 2, Study 3, Bergmann et al., 2012) could not be replicated in the long-term interaction. One possible explanation might be the stimulus material itself. While in the former studies real robot appearances were used, the appearances in this study were more humanoid with a machinelike look. Although the manipulation was found to be successful, since machinelike agents were perceived as more machinelike and unrealistic, they might still be more humanoid than the stimuli used in previous studies. This assumption is supported by the results regarding perceived similarity. In contrast to the aforementioned assumption that machinelike agents will be perceived as less similar, no differences in perceived similarity between the experimental conditions occurred. Thus, the machinelike agents might either be perceived as more similar to the user than expected or, by contrast, the humanlike agents are perceived as less similar owing to the artificial nature of virtual agents.

Moreover, as perceived similarity was measured after the interaction with the agent, participants might have relied more on the interaction with the agent and the agents' behavior than to its appearance. Therefore, the behavior might have overwritten the perceived similarity based on the mere appearance. Thus, the differences in both appearances might have been too small to affect the perception and interaction with the agent. It can be concluded that machinelike agents have to be clearly distinguishable from humanlike agents to evoke a positive perception. Since both kinds of agents might be perceived as artificial, a machinelike skin type alone does not seem to be enough. Referring to the results of Study 2b – in which realism was not found to affect likability, liking, or usage intention, while the species did – it can be assumed that the manipulation used in this study is more connected to the perceived realism. The body shape was the same for both versions of humanlikeness. In previous studies, the body shape differed clearly between the species (robots have different head shape and less humanlike features), while realism was inter alia manipulated via skin details and shading (except for stylized proportions). Hence, the perception of species might be derived more from the body shape presented and a machinelike agent should clearly have a different body shape to humanoid ones. Linking these results to the statements from Study 1, it can be assumed that the manipulation does not lead to a clear difference between humans and agents. Students reported wishing a clear border in the appearance to differentiate between virtual and real worlds, but the machinelike agent was not able to match this expectation. A machinelike agent was therefore not evaluated more positively than a humanoid one as was the case for the robots in Study 2 and 3.

With regard to the long-term usage of virtual assistance aimed for, we mainly investigated how the appearance affects social outcomes. However, social outcomes such as bonding, closeness, and trust remained unaffected by the agent's appearance. Based on the results of this study, the evidence from human–human relationships (McAdams & Constantian, 1983; Walster et al., 1966) cannot be applied in a human–agent interaction. Although the appearances were perceived differently in terms of attractiveness and attractive agents were liked more, they were not found to evoke a stronger interpersonal relationship.

It has been explained before that humanoid and machinelike agents did not differ in perceived similarity. Thus, the lack of differences between humanoid and machinelike agents with regard to usage intention, bonding, closeness, and trust might be caused by this failed manipulation. Since similarity is seen as an important influence variable in interpersonal relationships (Byrne, 1997; Montoya et al., 2008; Urber et al., 1998; which will cause higher usage intention) and participants were not found to rely on the agent's appearance to assume the similarity between themselves and the agent, an agent's humanlikeness cannot affect the social outcomes.

Overall, the agent's appearance was not found to be a crucial variable in affecting the interpersonal relationship in a human–agent long-term interaction. In line with this, since differences in perceived similarity were not found, trust was also not affected by the manipulation of humanlikeness as was assumed based on prior evidence (Levin et al., 2006). Thus, the user cannot be manipulated simply by the mere appearance so as to bond stronger with the agent, trust it more, or use it more often. People are not naïve and they do not rely on superficial cues like the appearance of their interlocutor for social relations and usage intention. These variables seem to be explained more by other variables such as functions. This concurs with the finding that the perceived usefulness of the agent could explain usage intention. Thus, the interaction with the agent has to be perceived as useful and this can enhance the usage intention.

Overall, bonding and closeness were assumed to affect usage intention and, thus, enhancing the interpersonal relationship through a virtual agent's appearance could lead to a higher usage of virtual agents. This may be promising, since for some applications, especially with regard to virtual assistance, a steady usage is required and this may have been guaranteed through appearance. Nevertheless, in contrast to Study 3, the findings of this study cannot support these assumptions. For a student sample and in a health application, the virtual agent's appearance did not enhance usage intention and social outcomes. As students are highly experienced in the usage of technology and most participants already had prior

experiences with virtual assistants, it could be that they might rely more on function (such as perceived usefulness) than on appearance.

Referring to the results of Study 3, where differences were seen in the bonding between seniors and students, the sample used may have led to these finding. In Study 3, students showed weaker bonding with the agent than seniors and in addition students stated in Study 1 that they would want to avoid treating virtual agents like realistic humans. Statements of the interviews emphasized that they wished for a clear border between virtual and real worlds. Therefore, they might be more aware of these processes. Most of the measures used are self-report measures, which is in this regard problematic. Previous studies already stated that users deny social outcomes with virtual entities, while they produce them mindlessly (Kim & Sundar, 2012; Nass & Moon, 2000). To avoid these effects and to supplement the self-reported measures, a behavioral measure of commitment was used. Nevertheless, no differences between appearances were found for the behavioral measures too. While physical attractiveness is often seen as a predictor of interpersonal relationships (Poulsen et al., 2013; Walster et al., 1966), the literature also mentions that other variables like proximity (Berscheid & Resi, 1998) or attitude similarity are important. Therefore, the appearance might help to build a first impression and to enhance the liking of the agent, but cannot enhance usage intention and cannot evoke feelings of bonding or closeness. However, the appearance of the agent can at least increase the liking of the agent and can foster users' intentions to live healthy, if the agent is designed in an attractive way. This is an important finding with the application (see discussion above).

A major benefit of this study is the long-term investigation and examination of the effect of time on the aforementioned processes and outcomes. According to the mere-exposure effect (Zajonc, 1968), the number of interactions was assumed to improve the evaluation and to foster social processes. The results support this assumption partly. While trustworthiness increases linearly over time, attractiveness decreases in a quadratic manner. Users perceive the agent as more trustworthy over time, and this effect can be highly beneficial, since trustworthiness in terms of virtual assistance is very important. If participants should, for example, follow the agent's advice (e.g., with regard to health behavior) or rely on the agent's recommendations (e.g., with regard to social activities), it is helpful if the agent is perceived as trustworthy. But by contrast, it has to be noted that the trust toward the agent decreases over time. Thus, while the agent becomes more trustworthy during multiple interactions, the opposite was found for users' trust toward the agent. Usually, it is assumed that one trusts someone who is trustworthy and, thus, that trustworthiness is connected with

trust outcomes (Colquitt, Scott, & LePine, 2007). This relation was not found in the current study, since trustworthiness increased, while trust decreased. The evoked trust was measured in regard to following the advice of the agent, while trustworthiness was measured more generally. It may be possible that since people get to know the agent better over time and feel a stronger closeness, the agent's trustworthiness increases in general; while the trust toward the agent might be more context-sensitive and depend more on the topic of the interaction with the agent. Perhaps something in the interaction with the agent decreases the specific topic-related trust. Hence, participants perceived the agent as trustworthy and this perception becomes even stronger over time (which is most likely connected to bonding and closeness), but with regard to the health application, they seem not to trust the agent and this decreases over time.

The attractiveness of the agent also decreases over time. Attractiveness might be more important in the beginning of an interaction to attract attention and an overall willingness to interact with the agent (although in this study attractiveness did not increase usage intention). This might fit the novelty effect often found for technology (e.g., Wells, Campbell, Valacich, & Featherman, 2010). People might perceive the agent as attractive, since they are new to its appearance and functions; however, over time the perception flattens as they get used to it. This tendency was found for technologies where the innovation was not great enough (Carbon & Leder, 2005). While designs with lower innovation were rated as attractive in the beginning, this perception flattens over repeated measures (Carbon & Leder, 2005).

By contrast, liking and closeness were found to increase over time. Thus, the novelty effect (Carbon & Leder, 2005; Wells et al., 2010) is not applicable to liking of the agent and closeness between the user and the agent. It can be assumed that social processes overcome the novelty effect. While more superficial variables like attractiveness seem to be affected by the novelty effect, the underlying social processes evoked by the virtual agent as social actor (Nass & Moon, 2000; Nass et al., 1994) are affected by the mere-exposure effect (Zajonc, 1968). Thus, people feel closer over time to the agent and like the agent more, the more they interact with the agent. This is somehow in line with findings by prior studies that demonstrate that people bond with a virtual health coach (e.g., Bickmore, Caruso, et al., 2005). However, they did not investigate the development of this bonding over time. While for bonding no effect of the number of interactions was found, the closeness increased with increasing number of interactions. Thus, just as in human–human interaction (McAdams & Constantian, 1983), users feels closer to the agent the more they interact with the agent. Nevertheless, the inconsistency between both measures has to be noted critically. This might

be caused by the measures themselves. While bonding was measured using self-report items that might be more salient (e.g., "My relationship with the agent is very important to me"), closeness was measured using a pictorial scale where it might be less likely to guess the construct aimed for. As described earlier, people tend to deny their actual social behavior toward virtual entities, although they show it unconsciously (Nass & Moon, 2000). Therefore, two kinds of measures were used and an additional behavioral measure was applied, but only at the end of the study. To examine the somewhat contradictory findings of bonding and closeness, future studies should include behavioral measures at all points of measurement. Nevertheless, this study provides evidence that people build a stronger interpersonal relationship with the agent over time.

Against the assumptions, liking neither predicted usage intention nor bonding and the assumed mediation of bonding on the effect of liking on usage intention was found. Thus, while liking increases over time, it cannot lead to higher usage intention. But trust was found to be mediated by bonding and affected usage intention positively. In combination with the aforementioned effect that usefulness also affects usage intention, it can be deduced that the functions of the agent such as usefulness and trust to follow the agent's advice are more important.

Moreover, with increasing number of interactions, the agent's ease of use increases, meaning participants perceived it as easier to use over time. People become more experienced in the interaction with the agent and therefore the agent is perceived as easier to use. Although the agent is easy to use right from the start thanks to the humanlike communication pattern of virtual agents (Cassell, 2000), this can be beneficial especially for users with lower technical experiences and capacities.

The present study further examined a possible interaction effect of appearance and number of interactions, but no such interaction effects were found. Bergmann et al. (2012) was the first to suggest that for specific appearances (specifically machinelike agents) the evaluation will be revised after the first impression, while for others not. One could assume, based on the findings of Bergmann et al. (2012), that the appearance is important in the beginning and for the first impression, while users adjust it after time and then, for example, the functions and interaction contents are more important. However, no such findings were obtainable in the present study. Therefore, the interaction frequency or time might affect the evaluation and perception of the agent regardless of the appearance.

9.4.2 Limitations

Besides the beneficial insights from the present study, like all studies it has limitations and methodological shortcomings.

First of all, although the duration of the study and the multiple interactions are the main advancement of this study, one can critically question whether five laboratory interactions can be described as a long-term study. With regard to the application, virtual agents should be used over months or years and thus the 4 weeks in this study are rather short. Nevertheless, when an experimentally controlled study is aimed for, it needs to be performed in the laboratory to exclude confounding variables. This means that on the one hand participants need to be committed enough to participate in multiple laboratory sessions and on the other hand the study is extremely time-consuming. For the current study, the experimenter spent nearly 400 hr in the lab, which means that human resources as well as the laboratory are bound over a long period. Referring to this effort, it is difficult to conduct a laboratory study over various months. Nevertheless, when these cost and time restrictions are ruled out, a laboratory study in which participants interact with the agent on a daily basis would be worth conducting. In particular, virtual assistants become more beneficial when they help and surround the user daily. Moreover, due to the extreme effort, only a limited number of participants were able to take part in the present study. Therefore, the number of people in the four experimental groups was rather small. This might have resulted in mostly nonsignificant results between the appearance groups, as the power was limited. By contrast, the power of the within-factor analysis (number of interactions) was large enough and here promising results were found. Thus, a bigger sample is desirable for future studies.

While a laboratory study was chosen to ensure a maximally controlled environment and exclude confounding variables, participants were not in their regular environment and might therefore have behaved differently. Laboratory studies always induce a testing situation and are therefore less generalizable. Especially with regard to long-term effects, a field study is needed.

The sample of the current study is a limitation. Since the previous studies highlighted the impact of the target group and especially participants' age, different effects were assumed to occur when, for example, seniors are examined. Additionally, the mentioned application of virtual assistance can be seen as specifically beneficial for elderly people and other people with need of support (Kopp et al., 2018; Yaghoubzadeh et al., 2013). Thus, there are only limited implications for this target group with regard to long-term effects based on the present study. While the underlying social processes based on the media equation (Reeves & Nass, 1996) are assumed to be equal for all kinds of users, the effects of appearance variables in

particular might look different when the study is repeated with seniors. Therefore, a replication with senior participants could be highly beneficial with regard to implications for the design of virtual assistants.

9.4.3 Conclusion

The present study investigates the effect of appearance for the first time in a long-term human–agent interaction. With regard to effect based on evidence from human interpersonal relationships, the agent's attractiveness and humanlikeness were examined. Attractive agents were more liked and could enhance participants' intention to live healthy. The agent's humanlikeness had mostly no effect on the evaluation of a virtual agent and also did not influence the social outcomes produced by the agent (bonding, closeness, and trust). Results demonstrated that users felt an increasing closeness over time, but this closeness was not affected by the agent's appearance. Moreover, the agent was perceived as more trustworthy and participants liked it more with an increasing number of interactions, although the trustworthiness did not lead to an increase of trust (which decreased over time). Thus, like the relationship between humans, the relationship between agents and humans became stronger over time, since people liked it more and felt more closeness. However, this relationship was not found to impact usage intention, which would have been helpful to guarantee a steady usage of virtual assistance.

.

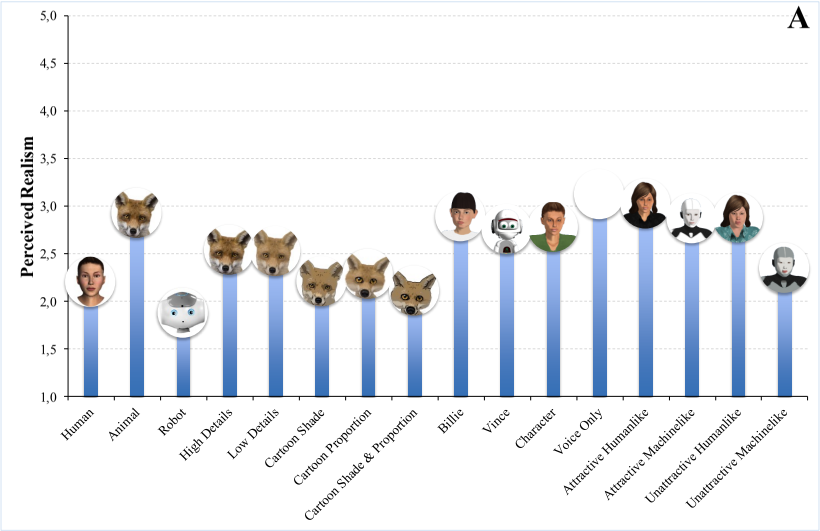
10 COMPARISON OF APPEARANCE EFFECTS CONSULTING ALL STUDIES

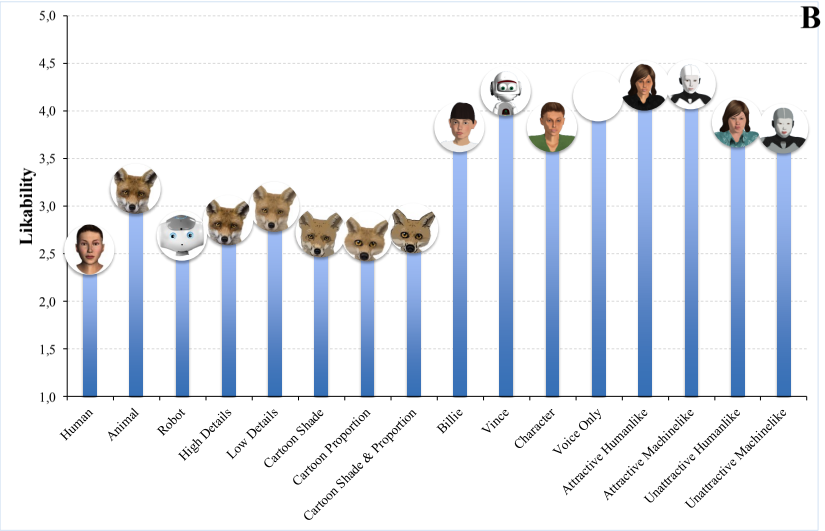
From the used multiple studies with different methodological approaches and different stimulus material, promising implications can be derived. To derive overarching patterns from all studies, the stimuli of all studies will be discussed with respect to each other in the following Section. The qualitative results of Study 1 are not included since there has been no experimental investigation of the used appearance variables. As described, the results are caused by different studies, and the following assumptions are rather a descriptive overview and no concrete statistically proven results. Hence, conclusions can only be derived with caution. Nevertheless, it helps to detangle the effects of appearance variables that have been found in the present dissertation.

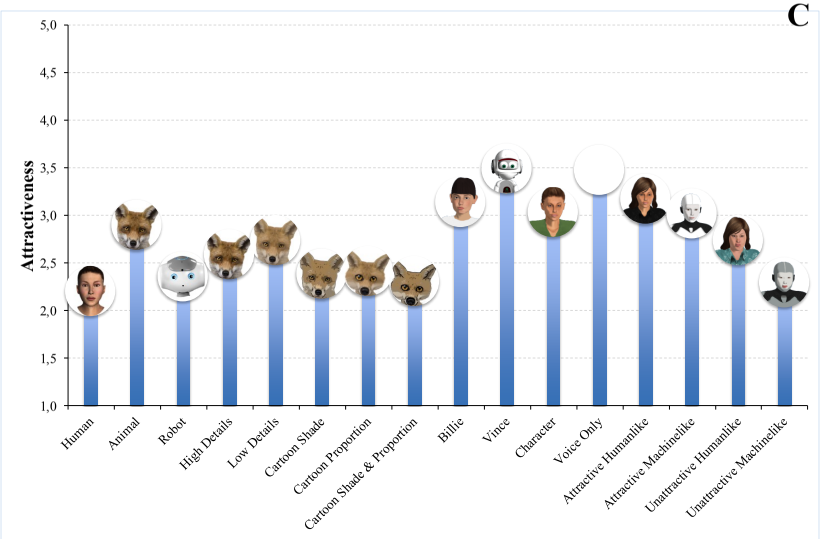
Figure 34 presents an overview of the results of all stimuli about person perception (perceived realism (A), likability (B), attractiveness (C), trustworthiness (D), and competence (E)), liking (F) and usage intention (G). The presented values in die Figure are independent of the age group. Therefore, only the results from Study 2B and not from Study 2A are represented. Thus, all stimuli have been tested in a between-subjects design. Since nearly no interaction effects of species and realism occurred in Study 2B, not all combinations of species and realism are shown, but the collapsed values. In the comparative observation of all experimental studies' results, differences between stimuli that have been tested only by pictures and results from interaction studies are observable.

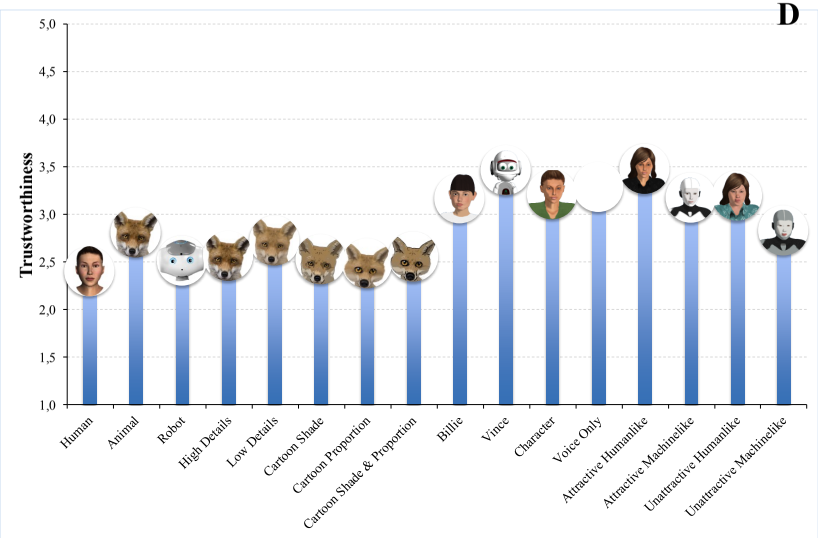
From the *Figure 34*, it is obtainable that especially for likability (B), trustworthiness (D), competence (E), and usage intention (G), the appearances that have been tested in an interaction were evaluated more positively than the appearances tested by pictures only. Thus, the interaction itself can be seen to enhance the evaluation of the agent.

Although the overall patterns of the evaluation of likability and attractiveness are mainly similar (see *Figure 34*), the difference between the effects of the mere pictures compared to the appearances tested in an interaction is greater for likability than for attractiveness. Hence, likability is more related to interaction cues than the perception of attractiveness. In contrast, people can judge other people's attractiveness more accurate even just by the mere appearance.


A
Perceived Realism
5,0
4,5
4,0
3,5
3,0
2,5
2,0
1,5
1,0
Human
Animal
Robot
High Details
Low Details
Cartoon Shade
Cartoon Proportion
Cartoon Shade & Proportion
Billie
Vince
Character
Voice Only
Attractive Humanlike
Attractive Machinelike
Unattractive Humanlike
Unattractive Machinelike


B
Likability
5,0
4,5
4,0
3,5
3,0
2,5
2,0
1,5
1,0
Human
Animal
Robot
High Details
Low Details
Cartoon Shade
Cartoon Proportion
Cartoon Shade & Proportion
Billie
Vince
Character
Voice Only
Attractive Humanlike
Attractive Machinelike
Unattractive Humanlike
Unattractive Machinelike


C
Attractiveness
5,0
4,5
4,0
3,5
3,0
2,5
2,0
1,5
1,0
Human
Animal
Robot
High Details
Low Details
Cartoon Shade
Cartoon Proportion
Cartoon Shade & Proportion
Billie
Vince
Character
Voice Only
Attractive Humanlike
Attractive Machinelike
Unattractive Humanlike
Unattractive Machinelike


D
Trustworthiness
5,0
4,5
4,0
3,5
3,0
2,5
2,0
1,5
1,0
Human
Animal
Robot
High Details
Low Details
Cartoon Shade
Cartoon Proportion
Cartoon Shade & Proportion
Billie
Vince
Character
Voice Only
Attractive Humanlike
Attractive Machinelike
Unattractive Humanlike
Unattractive Machinelike

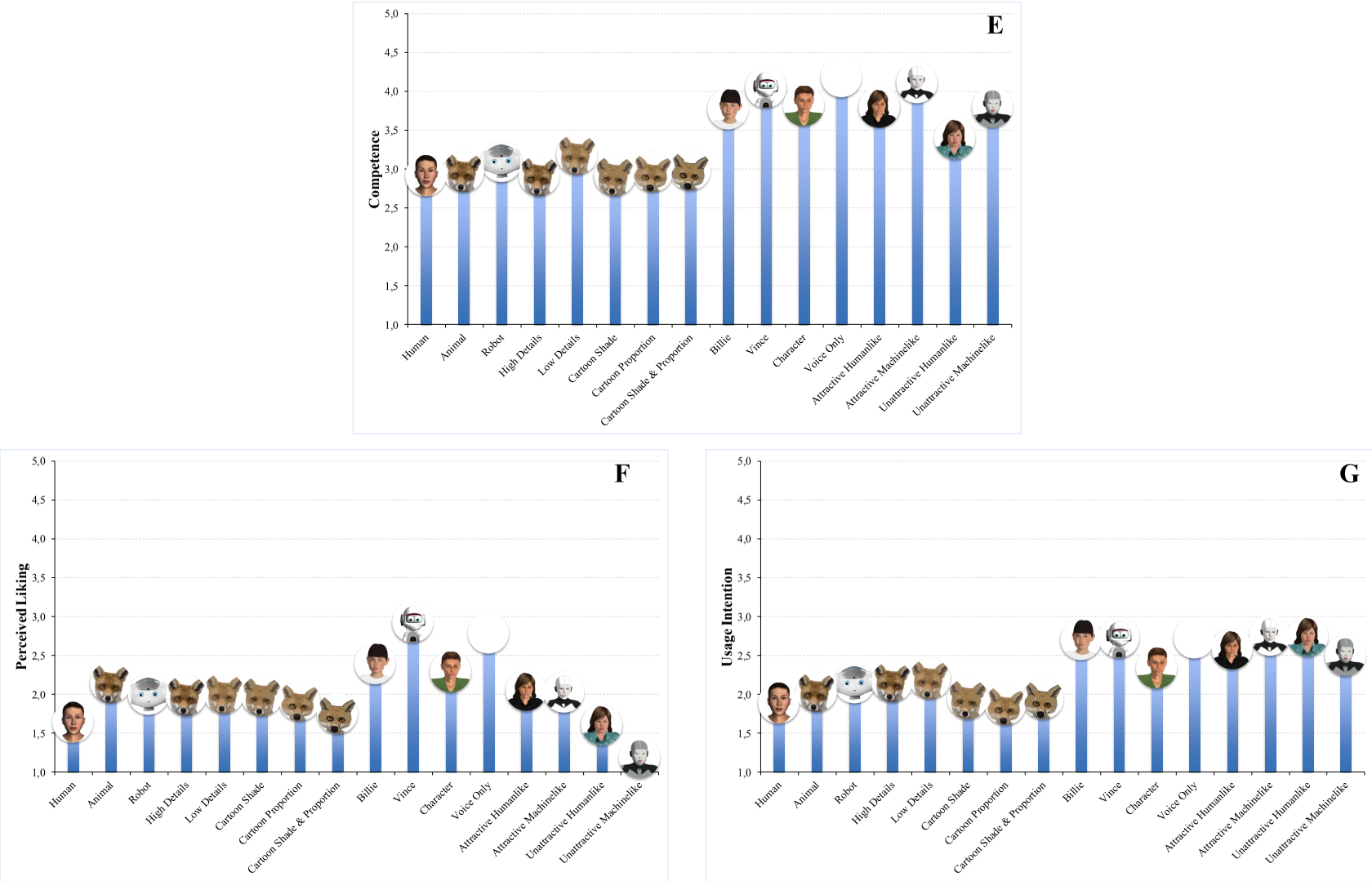


Figure 34. Means for the main dependent variables evoked by the various stimuli of the different studies

The presented functions of the agent in the interaction seem to foster the likability, trustworthiness, competence and usage intention. It has already been found, that nonverbal behavior, for instance, can affect the agent's person perception (Straßmann, Rosenthal-von der Pütten, Yaghoubzadeh, Kaminski, & Krämer, 2016). Moreover, due to the wizard-of-Oz settings, the agent responded in an effective way and only little failures in the interaction occurred, which might enhance the perception of competence and trustworthiness (especially when people are used to existing speech-based agents, where misunderstandings are more frequent). The studies' results demonstrated that perceived usefulness affects the usage intention positive. It could be assumed that participants perceived the agent based on the interaction as more useful than people, who had not the chance to interact with a virtual agent. Therefore, the usage intention might be higher for appearances tested in actual interaction. However, as interaction cues were not in focus of this work, it is not clear which interaction cues especially drove these effects.

In contrast, it is remarkable, that the aforementioned tendency seems not to be true for the perceived realism of the agent since stimuli of the online study were evaluated more or less similar realistic as appearances from the interaction studies. These findings contradict the assumption that behavioral cues like speech-output or nonverbal behavior can enhance the overall perceived realism (Bailenson et al., 2005; Groom et al., 2009). Especially animals from the online study were perceived as similar realistic than the stimuli in the lab studies. Results from Study 2 indicated that animals had been perceived as more realistic than humans and robots. While robots by its nature are more artificial than humans and animals, it is not astonishing that their perceived realism is lower. However, the difference between animals and humans in their perceived realism has to be examined critically. It could be argued that people are more sensitive to judge the realism of humanoid appearances because it is their own species and they want to distingue themselves from virtual humans. Since lab studies mostly use humanoid characters, the sensitiveness against virtual humans and therewith-related lower realism perception might diminish the positive effect of the interaction cues in comparison to virtual animals.

Overall, the found differences between the agent's appearances are smaller in the interaction studies compared to the online study. This finding emphasized that although the appearance is important for the evaluation of the agent; multiple other variables influence the perception and evaluation and might overwrite the appearance effects. Nevertheless, differences based on the appearance were also found and discussed in the subsequent two studies, in which users interacted with the agent.

IV GENERAL DISCUSSION

The following Section outlines the empirical evidence and advancements regarding theoretical and practical outcomes that the current dissertation gives. Accordingly, first the main results get summarized and discussed against the theoretical and empirical background, and implications for the proposed general research questions and the theoretical framework are made. Afterwards, limitations of the whole dissertation and open questions for future research are discussed. The dissertation ends with a conclusion that highlights the main outcomes.

11 SYNOPSIS OF EMPIRICAL RESULTS AND CRITICAL DISCUSSION RESPECTIVE THE THEORETICAL FRAMEWORK

Current developments of virtual agent call for long-term usage (e.g., in terms of virtual assistants for people in need of support; Kopp et al., 2018; Yaghoubzadeh, Kramer, Pitsch, & Kopp, 2013). Thus, the users interact over a steady time with the agent. As it can be assumed that humans show similar social reactions towards virtual entities than towards other humans (Reeves & Nass, 1996), it is likely that within these long-term interactions social relations will occur. Pioneer long-term studies did already demonstrate that people have feelings of bonding towards the agent when they interact over a longer period (Bickmore, Caruso, & Clough-Gorr, 2005; Bickmore, Gruber, & Picard, 2005). What has so far missed out in the prior research is the effect of the agent's appearance. Again against the background of human-human interaction, it can be derived that the appearance in terms of attractiveness (Berscheid & Walster, 1974; Dion et al., 1972; Walster et al., 1966), and similarity (Byrne, 1997; Montoya et al., 2008) have an influence on the development of interpersonal relationships. Therefore, the oval goal of this dissertation was to investigate the effect of appearance variables on social relations between humans and agents in a long-term interaction.

Based on the described research gap, this dissertation examined four main research goals, which will be discussed successively in the following Chapters:

(1) Systematic categorization and examination of appearance variables

(2) Investigation of the effects of user characteristics in terms of appearance effects

(3) Examination of appearance effects' development appearance effects over multiple interactions and the effect on social relations within a long-term interaction

11.1 EFFECTS OF APPEARANCE CATEGORIES

There is a body of research that presents the effects of a virtual agent's appearance regarding different outcome variables such as the overall evaluation of the agent (e.g., McDonnell, Breidt, & Bülthoff, 2012; Ring, Utami, & Bickmore, 2014; Zell et al., 2015), users' motivation (Baylor, 2009b), learning outcomes (Baylor & Kim, 2004; Sträfling et al., 2010), or buying intention (Lunardo et al., 2016; Terada et al., 2015). While this highlights that the appearance has an important role in the human-agent interaction, prior research also comes along with shortcomings that have been addressed in the present dissertation. First, the experimental research has been conducted in a rather unsystematically manner since appearances that differ in multiple variables have been compared, and it is difficult to detangle the effects of specific variables (see Chapter II3.4 for detailed descriptions). To be able to use more systematic approaches, the multiple variables, that a virtual agent's appearance contains, have to be collected and categorized. To this end, a categorization has been constructed based on prior research. This categorization consists of five main categories, which also have multiple sub-categories: (0) embodiment, (1) species, (2) realism, and (4) features specification (please consult *Figure 1* in Chapter II3.1 for an overview over all categories).

This categorization was used to examine the effects of the appearance of a virtual agent (more specific of a virtual assistant) in a systematic manner by different empirical studies. The first two studies investigated the mere effect of appearance by the imagination and observation of the agent's appearance only. While these studies highlighted promising insights into the perception of appearance variables, it was further needed to assess whether these findings are transferable to interactions between the human and the agent. Since a lack of examinations in a reciprocal speech-based interaction between the user and the agent about appearance effects has been identified, in this work, appearance effects have further been investigated in an actual speech-based interaction. This investigation leads to valuable insights regarding the role of appearance in the interaction between humans and agents. In the following Section, the results of the conducted studies become outlined with regard to the general research questions.

11.1.1 Which Appearance Variables are of Pivotal Interest for the Evaluation of Virtual Agents in an Assistive Application? (GRQ 1)

As the categorization was used in this work for the first time in experimental studies, not all variables can be included and specific variables have been extracted. With regard to this selection, Study 1 gathered a holistic view and presented that species and realism are the

ones of most interest concerning virtual assistance. Therefore, these variables have been included in the subsequent studies to experimentally test the effects of species and realism on the evaluation of the agent and in addition to that associated social outcomes. When all results of the studies are taken together, it can be assumed that the species is even more pivotal than the degree of realism. During the studies, more differences on the investigated outcome variables have been found for the species. In the discussion parts of Study 2 and Study 3, it has already been claimed that the used manipulations of realism might be to close together. When differences for the degree of realism have been found, then most of the conditions with extreme forms of realism (such as appearances without any stylization and a cartoon-stylized shade and proportions) evoked different evaluations. Building on the findings from Study 4, where it was intended to manipulate humanlike and machinelike appearances, which would belong to the species of the agent, it also became apparent that the body forms might be more pivotal for the evaluation. Both conditions used the same facial and body features and solely the skin and hair pattern was varied (humanlike normal skin, hair, and clothes vs. mechanical white faces, no human skin but mechanical skin pattern; see *Figure 23*). Since no differences (except for realism) in the investigated outcome variables have been found, it can be assumed that the form of the body might be more important. Species are distinguishable by its body types (humans have other proportions, features, etc. than for example animals). Especially, as students stated in Study 1 that they wish for non-humanoid agents to distingue virtual from the real world, the humanlike facial and body features might be to close to a real human. On the opposite, robots as virtual assistants, as it was tested in Study 2 and 3, have been found to evoke positive outcomes. These findings match the assumption derived above: species has a main influence on the evaluation of a virtual agent because of the different body types that make users compare the virtual agent to the real world. Depending on the target group (see Section 11.2) a close imitation of a human can either enhance the outcomes or diminish it.

While the effect of species has been investigated in prior research (e.g., Bergmann, Eyssel, & Kopp, 2012; Sträfling et al., 2010; Terada et al., 2015), no such systematic approach has been used in combination with the agent's realism. So far no prior study investigated species and realism combined. Therefore, the present work can extend the body of research with assumptions about the pivotal effect of species.

11.1.2 How Should a Virtual Assistant's Appearance be designed? (GRQ2)

When all stimuli are compared regarding their perceived realism, the tendency becomes apparent that robot appearances evoke lower realism perception than humans, but animals have been evaluated as most realistic. The differences with regard to the possible

effect of the interaction have been discussed in Section 0. Nevertheless, in sum humanlike appearances were perceived as more realistic than machinelike ones (referring to the results of Study 2 and 4). These results are most likely caused by the artificial nature of robots or machinelike appearances; in contrast, humans (and animals) are living beings, which leads to a higher realism of the species. While the stylization was also found to lower the perceived realism in Study 2, those effects have not been replicated in the interaction study (Study 3). As results of Study 2 further demonstrated that overall the species is more decisive for the overall evaluation, the differences in stylization between both stimuli in Study 2 might have been too small. Therefore, it could be derived that small differences in the degree of realism seem not to have a big influence as expected, while the species (assessed by the body type) is pivotal for the evaluation of a virtual agent.

What is further remarkable is that the voice only condition evoked a high perception of realism (but with no significant difference in the study), although less social cues were obtainable in this condition. As the voice and the performance of the agent were relatively flawless due to the wizard of Oz setting, it evoked a realistic perception, and no further cues are needed.

Animals have been evaluated as more likable and attractive in Study 2 compared to humans. Moreover, the robot appearance in Study 3 was perceived as more likable and more liked than the humanoid appearances. As assumed, agents with attractive facial and body features have been perceived as more attractive in Study 4.This is in line with prior findings (Khan & De Angeli, 2009; Sobieraj & Krämer, 2014) and demonstrates again that humans apply social rule also to virtual entities (Reeves & Nass, 1996). Since no interaction effect with the humanlikeness occurred, this attractiveness evaluation is also true for unrealistic and machinelike agents. This finding extends prior research, in which only the attractiveness of facial features of a humanlike appearance was tested. Nevertheless, no concrete facial features have been tested, but an overall attractive agent compared to an unattractive agent. Hence, further studies could examine this finding in more detail. With regard to attractiveness, an effect of realism has been found in Study 2B: agents with a realistic stylization and low details are perceived as more attractive than agents with cartoon stylized shade and proportions. To conclude, the perceived attractiveness of an agent is highly related to the agent's appearance.

While Animals have been perceived as more trustworthy than humans and robots in Study 2B, in both lab studies the agent's trustworthiness has not been affected by the agent's appearance. When the means of the different studies are compared descriptively (see Section

0), it was remarkable that the trustworthiness ratings were higher for the appearances tested in the interaction study. It has been concluded, that the interaction itself fosters the trustworthiness and therefore might overwrite the rather small effects of the agent's appearance on trustworthiness.

While the results of Study 2B and Study 4 did not present differences between the various appearances regarding competence, the difference about the embodiment was found in Study 3. A non-embodied agent was perceived as more competent than embodied agents were. Thus, it can be summarized that the embodiment is more effective with regard to the perceived competence.

Study 2B indicated further that non-humanoid agents are more liked than humanoid ones. This assumption is supported by the findings of Study 3 since the robot appearance was more liked than the humanoid agents were. However, the machinelike appearances in Study 4 did not evoke higher liking by the participants than the humanlike agents. As discussed in Chapter III9.4, this might be due to the manipulation, which is not directly comparable to the others, where species have been varied. Since realism was found to have a lower impact than the species, it could still be argued that virtual robots are more likable than virtual humans are. Nevertheless, it depends on the aimed target group.

Robots evoked a higher intention to use than humans and animals in Study 2B. No main effects of the agents' appearance on usage intention have been found in both lab studies. However, as an interaction effect with the age group has been found, the effect of the appearance on usage intention depends on the age group. This is also true for trust and bonding. Seniors were found to be more sensitive for appearance effects with regard to social outcomes like trust and bonding, while overall no effects have been found from appearance variables on trust and bonding (c.f. Study 3 and Study 4).

While animals are evaluated as more realistic, likable attractive and trustworthy, users do not want to use an agent with this appearance. This is against the theoretical assumption of the attractiveness stereotype (Dion et al., 1972), which assumes attractiveness to trigger other positive perceptions. However, as mainly no differences in usage intention referring to the different appearances occurred, people might rely more on functions or interaction cues for their usage intention. This can be supported by findings from Study 3 and 4 since here the perceived usefulness was found to affect usage intention positively.

However, an especially interesting finding occurred in the long-term study with regard to user intention to live healthily. Users' intention to live healthy is greater after an interaction with an attractive agent than for users of an unattractive agent. Thus, although overall only

little effects for person perception or social outcomes have been found, the agent's appearances can have very promising effects, if the appearance is designed attractively.

Taking all variables into account, it has first to be mentioned that for seniors specific outcomes have been found and therefore other derivations have to be made (these become discussed in the following Section). However, when a younger target group is considered, the agent's appearance can evoke positive effects, when it has a robot appearance and when attractive facial cues are presented. While animals evoke high ratings in person perception, those effects are not transferable to the usage intention. Nevertheless, there are still inconsistencies in the findings, which should be further investigated by future research.

11.2 THE INFLUENCE OF USER CHARACTERISTICS ON APPEARANCE EFFECTS

In the theoretical background, different applications fields and its target-groups have been discussed. It was remarkable that especially specific target-groups such as seniors are seen to be promising users of virtual agents (Beer et al., 2015; Kopp et al., 2018; Yaghoubzadeh et al., 2013). It has for other user characteristics already been shown, that they affect the evaluation of appearance variables (Qiu & Benbasat, 2010). While there first studies that explored appearance preferences of seniors (Chattaraman et al., 2011; Tsiourti et al., 2014), no experimental investigations exist. Therefore the present dissertation looked at the effect of user characteristics on appearance effects.

11.2.1 How Do User Characteristics and Target-Groups Influence the Evaluation and Perception of a Virtual Agent's Appearance? (GRQ3)

There was a lack of research that investigates the preferences with regard to appearance variables of the very promising target group of virtual assistants: people in need of support like seniors. Giving first evidence with regard to this research gap, the present dissertation investigated age-related differences and took other user characteristics such as personality traits or former attitude towards virtual agents into account. In the following, first overarching results about age-related differences become discussed and afterward the influence of other user characteristics is summarized. In Study 2A, the perception and evaluation of cognitively impaired people have also been taken into account. The study gives first hints that they like animals more and show higher usage intention for animals than for a virtual agent designed as robot or human. However, as the other studies did not take this target-group into account, no further implications than the aforementioned ones in the related discussion part (Chapter III6.4) can be made.

Age-Related differences

In the course of this dissertation, promising age-related differences became apparent and constantly occurred throughout the different studies.

In Study 1, a qualitative approach was used to explore age-related differences referring to the constructed categorization (Chapter II3.1) for the first time. With regard to embodiment, seniors mostly preferred an embodied character, since they wish for an interlocutor, who can be addressed during the interaction. In contrast, students preferred the non-embodied agent, since they would like to be able to use the agent independent of a certain screen.

Statements of the seniors further indicate that they would prefer to interact with a realistic humanoid appearance since this triggers more social cues. As communication and interaction with humans are familiar, seniors prefer to transfer this familiarity to virtual agents. In line with this, seniors stated that they would have stronger trust in a humanoid agent and would expect themselves to bond towards realistic humanoid agents. Students stated to prefer non-humanoid agents and stylized agents since they assume a realistic humanoid appearance to evoke expectations that the agent might not be capable of (e.g., chitchat). Moreover, they would like to prevent the processes that the seniors appreciate since they wished for a clear border between a virtual agent and real persons.

In addition, differences with regard to the process of customization occurred. Results indicated that students liked to customize the appearance of a virtual agent, but for seniors, a choice out of option can be seen to be more promising. While prior research, on the one hand, presented findings that customization is very promising (e.g. Marathe & Sundar, 2011; Norton, Mochon, & Ariely, 2012), there is also research that the selection out of a wider range of option leads to a more positive evaluation of the agent (while this was in contrast not true for a design process with increasing options; Diehl et al., 2017). However, no research investigated the differences between age groups with regard to customization and design choices. This lack of knowledge is a potential examination for further work.

Overall, seniors highlighted to like the social processes that realistic humanoid agents evoke since it can be assumed to be beneficial for their usage of virtual agents. In contrast, students looked at the agent and its appearance in a more technical way. They saw technical advantages of non-embodied agents and wanted to use a rather unrealistic appearance to diminish social outcomes of the agent to differentiate virtual and real worlds. It can be assumed the same for the customization process: while students see the agent more as a

technical tool, which they can manipulate regarding their preferences, seniors are not used to designing their interaction partner, but to chose with whom they interact.

The first evidence from Study 1 about the agent's species have mainly been replicated in the subsequent experimental studies. In Study 2A it was noticeable that seniors evaluated robots less positive than humans and animals. This is in line with the assumption from Study 1 that they would prefer with agents that trigger social interactions. As humans and animals are living beings, seniors are used to the interaction with it. Therefore, it becomes obtainable that seniors disliked robots whose appearance is more artificial. Additionally, in Study 3 seniors liked a humanoid agent more and showed higher usage intention, bonding, and trust towards a humanoid agent, while this was not true for students. Therefore a humanoid agent can be seen as most appropriate for seniors since it triggers more positive outcomes that can enhance the interaction with the agent (see the discussion of social processes in Chapter 11.3.1).

But in contrast to the statements from Study 1, the results of Study 2A indicated no effect of the degree of realism. The overall pattern was found that seniors rely more on the agent’s species, while students rate the agent based on the degree of realism. With regard to the student's evaluation, the aforementioned assumptions from Study 1 were also not replicated since students evaluate realistic appearances overall more positive than unrealistic appearances.

As more differences in the evaluation of seniors concerning different appearances occurred, and Study 4 (in which only students participated) also demonstrated only little significant effects of the agent’s appearance, the appearance of the agent seems to be more pivotal for seniors.

To conclude, the results highlight the importance of the aimed target group (especially age group), for which the virtual agent will be designed. When seniors are the aimed target group, a humanoid agent can enhance the interaction positively.

Effect of other user characteristics

Building on statements from Study 1, it was assumed that further user characteristics such as the attitude toward virtual agents of the user might effect the evaluation of appearance variables and the overall evaluation of the agent. While the moderating effect of other user characteristics than age has not been found in Study 2, several user characteristics were found to affect the liking of the agent and participants usage intention. Nevertheless, these results were sometimes inconsistent throughout the studies.

Participant's tendency to anthropomorphize was found to influence liking in Study 2B: people, who have a higher tendency to anthropomorphize, liked the agent more in general. However, this personality trait had only an influence when pictures have been evaluated, but in an actual interaction study (Study 3 and 4); no effect of anthropomorphism tendency was observable. This might be caused by the higher social cues that occur in the interaction studies. While in Study 2B the users have only seen the appearance, they seem to need a higher tendency to anthropomorphize to like and to intend to use this agent. But in contrast, when more social cues like speech-output, responses or nonverbal behavior are presented, this personality trait is less relevant, since the cues enhance the perceived anthropomorphism overall.

With regard to usage intention, mainly the user's attitude was found to be an influencing user characteristic. As it has already been claimed that a negative attitude towards artificial entities leads people to avoid the interaction with such entities (Nomura et al., 2006), it is not surprising that people who stated higher negative attitude toward the social influence of virtual agents lead to lower usage intention. Additionally, users' general attitude was found to influence usage intention positively. This finding is also in line with prior research (Heerink et al., 2010). Therefore, to enhance the usage intention of people, their general attitude has to be changed into the positive. This is an interesting finding of the application of virtual agents as assistive technology. As the promising target-group of seniors is seen to be more skeptical toward artificial entities (Scopelliti et al., 2005), it is especially crucial to evoke a positive attitude toward virtual agents. Here, media, family members or health care stuff could help to spark seniors' joy for virtual agents, which would consequently help to increase the usage intention, and might result in higher effectiveness of virtual assistants.

11.2.2 What is the most Effective Appearance for People in Need of Support? (GRQ4)

As described above, differences in appearance effects with regard to the user's age group have been found. Seniors were found to evaluate agents with a robot appearance less positive and showed stronger usage intention for a humanoid agent. Moreover, they felt stronger bonding and trust toward a humanoid agent.

Thus, taking all empirical results into account, the virtual agent of a senior should be designed as a virtual human. Study 2 indicated that seniors do not rely on realism and in addition, the findings regarding bonding and trust were found to be true for an agent with rather cartoon-stylized shades. Thus, unlike the statements of Study 1, no highly realistic

humanoid agent is necessary. As long as the agent is designed as human, seniors show higher usage intention, trust, and bonding.

In sum, the familiarity of a human seems to trigger more trust and bonding, and therefore seniors are more likely to use the agent when it is designed as a human since it triggers more social cues.

11.3 APPEARANCE IN TERMS OF LONG-TERM INTERACTIONS

As especially social relations in long-term interactions have been focused, those results become summarized first along the theoretical framework that has been claimed in Chapter II4.2. Afterwards the impact of the number of interactions on those effects will be discussed.

11.3.1 Which Effect Has a Virtual Agent's Appearance on the Interpersonal Relationship in a Human-Agent Interaction? (GRQ5)

As in applications where the users were assumed to interact with the agent over a long period, the steady and regular use of the agent has to be ensured to foster the effectiveness of the agent. In the present work, it was assumed, that social outcome like an interpersonal relationship between the users and the agent can strengthen users' intention to use and actual usage behavior. Since in human-human interaction the visual appearance is important for the development of interpersonal relationships (Walster et al., 1966), a theoretical framework has been conducted, in which the effect of appearance on social variables have been summarized (see Chapter II4.2). In the following, this framework becomes discussed based on the findings from this dissertation.

Against the background of prior research, it was assumed that the agent's appearance affects the person perception (McDonnell et al., 2012; Ring et al., 2014; van Wissen et al., 2016), liking, and bonding (Berscheid & Walster, 1974; Walster et al., 1966), trust towards the agent (Stirrat & Perrett, 2010), and the intention to use or interact with the agent (Walster et al., 1966). All of these assumed relations have been found, even though they mostly occur with age as moderating variable. The detailed effects of specific appearance variables have already been discussed in the previous Section. While seniors were mainly found to perceive a humanoid agent more positive, like it more, show higher usage intention, no such effects were observable for students. Therefore, seniors can be seen as more sensitive for appearances effects.

Moreover, it has been found that trust and liking affect the relationship towards the agent (users' feelings of bonding) and this leads to higher usage intention. Thus, users bonding towards the agent mediates the effect of trust and liking towards the virtual agent. At

least for seniors, appearance was found to influence liking and trust. It can be assumed that social relations with the agent produce seniors' usage intention since the from the appearance evoked trust and liking lead to higher bonding towards the agent, which results in higher usage intention. Nevertheless, with regard to the mediation of bonding, inconsistent findings have been found. While in Study 3, bonding was found to mediate both trust and liking's effect on usage intention, in Study 4 only a mediation of bonding for trust and usage intention was found.

Based on the findings of Study 4, it cannot be assumed, that usage intention leads to actual usage behavior. This is against the theoretical assumptions from the theory of planned behavior (Ajzen, 1985), where an intention is seen as the best indicator for actual behavior. However, it has to be marked that the agent, for which actual usage was measured, was not identical to the agent, where usage intention was queried. Since the interactions with the agent in the lab were mandatory, participants' usage of an online version was measured, in which the same agent's appearances and interaction topics have been presented. During the online interaction, the functions of the agent were limited, and participants could only interact via pre-scripted keys. Thus, the missing connection between usage intention and usage might be caused by the lower functions of the online version.

As virtual agents are technical entities, also variables from technology acceptance models have been included in the framework. Based on prior research (Heerink et al., 2010; Venkatesh & Davis, 2000) the perceived usefulness, ease of use and perceived enjoyment have been included into the framework and were expected to influence users usage intention of the virtual agent. Results indicate only an impact of perceived usefulness on usage intention. High-perceived usefulness leads to higher usage intention. This finding is constant for both interaction studies. While in contrast to the assumptions ease of use and enjoyment did not influence usage intention, a significant relation to users' liking of the agent has been found. People with higher enjoyment during the interaction liked the agent also more. Again, the finding regarding liking has been found in Study 3 and Study 4. Additionally, during the long-term study (Study 4) a significant negative effect of ease of use was reported. When the agent was perceived as easier to use, people liked the agent less. While this finding is contra-intuitive, it might indicate feelings of boredom or ease of use might function as a mediator of expertise. When people are more used to the usage of virtual entities, they might probably perceive the agent as easier to use. Moreover, people who are more experienced might be harder to impress and therefore more critical in their evaluations. However, this finding was not found in Study 3. Therefore, it has to be handled with caution.

The age group was found to influence the perception and evaluation of appearance variables and users' social outcomes related to this. Moreover, other user characteristics have been investigated. Details have already been summarized in Section 11.2. Mainly the attitude toward virtual agents (general and negative) was found to affect peoples' intention to use the agent. This is also in line with prior research (Heerink et al., 2010). Nevertheless, although the statements from the interviews indicated that attitude towards virtual agents might also moderate the effect of appearance variables, no such findings have been found in Study 2.

As mentioned above, differences between the mere perception, where the effects of appearance only have been tested and the effects of appearance in interaction studies were observable. It was indicated that especially for trustworthiness, likability, competence, and usage intention the interaction enhanced the evaluation of this variables. Therefore, the framework should be adapted with regard to these differences, but since the findings are rather descriptive and have not been tested statistically, no such concrete conclusions can be derived from the present dissertation. Hence, this might be a beneficial research gap for future studies.

Figure 35 presents the final framework with all connections between the pivotal variables that have been found. As age was found to be a relevant moderator of the evaluation of appearance variables, dotted lines indicate an influence of the age group.

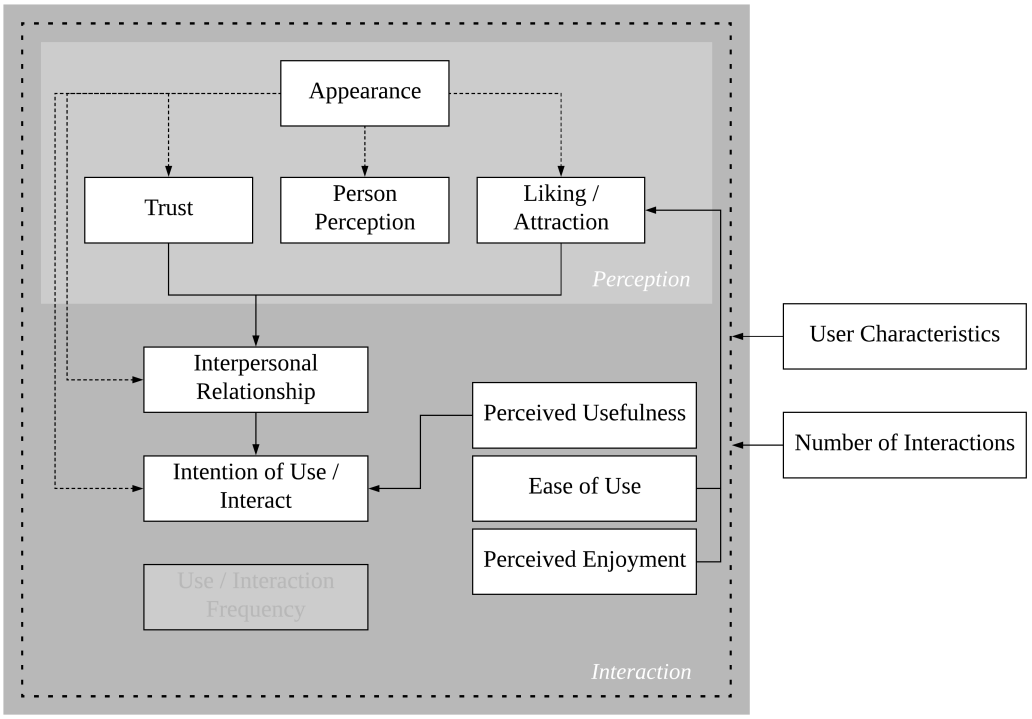


Figure 35. Adopted framework based on findings from this dissertation

11.3.2 How Do the Effects of a Virtual Agent's Appearance Develop Over Time in a Long-Term Human-Agent Interaction? (GRQ6)

With respect to long-term interactions, the number of interactions also assumed to influence the described processes. Building on the mere-exposure effect (Zajonc, 1968), it was assumed that the evaluation of the agent increases with the number of interactions. Results of Study 4 indicate an increase in liking, trustworthiness, closeness, and ease of use. In contrast, the attractiveness of the agent and the trust towards the agent decreased over time.

While first studies hint that the effect of the agent's appearance might change over time (Bergmann et al., 2012), appearance effects have been investigated over time. Unlike this prior evidence, no interaction effects of appearance variables and some interactions have been found. This is also against the assumption, that appearance is more important at the beginning of social relationships, while it becomes less pivotal over time. However, since overall Study 4 indicated only little effects of the appearance, the manipulated differences might have been too small. It has already been discussed that the species (as it has also been used in the Study from Bergmann et al., 2012) might have a more promising effect. Therefore, the long-term effect of other appearance variables should be investigated in future work.

12 IMPLICATIONS

Based on the above-discussed results promising implications can be made regarding the media equation assumption, ethical consequences, and practical recommendations.

As in the long-term study, certain variables increase over time (such as trustworthiness or closeness), it might carefully be assumed that the media equation theory (Reeves & Nass, 1996) is also valuable over time and is no phenomena only resented in the first interactions with artificial entities. Nevertheless, the time span of the long-term study was still small (due to the experimental lab setting). Therefore, these are only first hints. Nevertheless, as other long-term studies present similar results (e.g., Ring, Shi, Totzke, & Bickmore, 2015), it can be seen as a reasonable assumption.

While the social outcome that is produced by the agent's appearance and the long-term interactions can be seen as beneficial for the effectiveness of the agent, it also has negative side-effects especially with regard to ethical implications. It can be questioned, whether it is a good thing, that artificial entities can evoke feelings of bonding, trust, and closeness. This can become critical for example when a virtual assistant is embedded into users' households, and it has been removed for some reasons (e.g., financial, or technical reasons), then this bonding could cause psychological harms by the users. The bonding might also cause feelings of

commitment and dependence to the virtual agent, which can lead to negative outcomes. Users might feel that they have to follow the agent's advice more, even when they might be not positive but have negative consequences for the user and the users' environment. Nevertheless, the presented findings in this dissertation are only a first hint and it hard to make exact conclusions, whether these scenarios might occur. Therefore, field studies and long-term studies with the specific user group have to be conducted. As discussed above, it is also questionable, whether the stated bonding is the same feeling than for a human. As no comparison of humans' impacts with the effects of the agent has been made in this study, no implications can be made with regard to this. Therefore, it is still unclear if virtual entities elicit the same feelings or only the same responses (compare the debate around Kiesler and Sproull (1997), who claim that the social reactions are only demanded characteristics.

Although for some user groups, the appearance might enhance feelings of bonding. Those feelings can only hardly be prevented when a virtual agent is used. As in Study 3 also the non-embodied agent evoked higher feelings of bonding, not only appearance but also the agent itself triggers bonding. This is in line with the assumption that some social cues like speech output are enough to trigger social outcomes (Nass et al., 1993). In the conducted interaction studies of this dissertation, the agent was designed in a relational way (Bickmore, Gruber, & Picard, 2005). This has also to be mentioned critically since it might foster the social outcomes even more. Designers must decide carefully, which among of social cues they implement in the virtual agent. As presented in some circumstances those social outcomes might be positive, while they come along with ethical restrictions.

As age has been found to be a critical factor in the evaluation of appearances, the comparability of multiple studies over a long time span has to be questioned. It depends on the explanation for the presented findings. One could assume that the perception and evaluation might also be due to trends and not only caused by lower experience and literacy with the handling of technical entities of older people. Thus, it has to be examined if the findings are caused by fix preferences of specific generations or if it is demanding on the recent technical state-of-the-art. When fix generation-related preferences are found, then the results of studies where multiple years are in between are not comparable, since the (most likely) students that participated in earlier studies are nowadays nearly seniors.

As practical implications, design recommendations, especially with regard to the aimed target-group, can be made. Developers of virtual agents should specifically define and investigate their aimed target group. When seniors are the group of interest, the appearance of the agent is highly critical. Depending on the aimed effects that the agent should have, a

humanoid appearance might be promising since it was found to foster social outcomes for seniors. This might be helpful in applications, were longitudinal interactions are aimed, and therefore the usage intention can be enhanced via this effects. Additionally, in more sensitive topics and applications such as health-related contents, the agent could be designed humanoid. This extends former research, where realistic agents were found to fit better into medical applications (Ring et al., 2014). In contrast, situations where this social process should be excluded, non-humanoid appearances might be more promising. For example for uncomfortable topics, a humanoid appearance and its social outcomes might inhibit reactions of the users. While it has already shown that people disclose more sensitive information towards computers than towards humans (Weisband & Kiesler, 1996), this effect might be applicable for virtual humans with more artificial appearances.

As hints have been found that when pictures only are presented the appearance has a bigger role than in the interaction, this has practical implications for applications where only a picture is presented. It has to be tested, whether the reciprocal speech-based interaction enhanced the outcomes or if a text-based interaction with pictures has the same effects. Moreover, when virtual agents are market mostly the appearance is presented, thus in terms of economic outcomes, the appearance should be designed favorably, although the interaction might balance the effect of appearance when users interact with it.

13 LIMITATIONS

While specific methodological limitations of the single studies have already been discussed in the respective discussion Sections, there are overarching limitations of the whole approach that have to be mentioned.

The present dissertation built a categorization of appearance variables and investigated the effects of appearance variables in social interactions between human and agents. While in the qualitative approach a holistic view of all categories was gathered, the subsequent studies focused on specific (sub-) categories. Because of the experimental methods that have been used, categories had to be excluded, so that only evidence for a limited number of appearance variables exist. Mainly, species and realism have been focused, as those tuned out to be especially interesting concerning virtual assistants and the related target-group specific differences (compare results of Study 1). Consequently, other variables as feature specifications remain understudied and no implications could be derived about those variables. Nevertheless, it would be beneficial to include such variables in further studies especially with regard to possible interaction effects. Although the approach aimed to

investigate the appearance variables more systematically and there was a special focus on the controlled manipulation of the agent's appearances, other variables that have not been in focus might still interfere this investigation. For example, the used humanlike stimuli still presented other appearance variables than only species and realism, as also information about the sex, perceived age or clothing style are obtainable. To minimize those confounding effects, in the present work those variables were held constant between all manipulated appearances (e.g., same casual clothing style was used). However, this design decision, even when it was constant, might influence the perception and evaluation of the agent. Moreover, these underlying variables that were not in focus of the manipulation might interact with the pivotal variables. For instance, as a cartoon-stylization was described to be "kids stuff" (see results of Study 1), the stylization might also interact with the agent's perceived age. It could be possible that for a more childlike agent a cartoon-stylization fits more and is therefore evaluated more positive. So far, only a little evidence about the interaction of feature specifications with other variables exist, but this kind of confounding effects have to be scrutinized. Nevertheless, this limitation cannot be ruled out in experimental examinations about appearance factors, since appearance is a multifarious construct and only a limited number of appearances can be compared, and therefore specific design decisions have to be made.

Furthermore, it has to be mentioned that while the same categories have been used, different stimuli were used in the studies, which was mainly caused by technical restrictions. Since for a presentation of solely pictures more systematic manipulations are possible than for moving agents in an interaction study (e.g., pictures are easier to manipulate with the use of image editors and a higher variety of pictures than dynamic models exist). While this is especially with respect to the aforementioned confounding variables a weakness, the replication of findings based on different stimuli helps to enhance the power of the implications.

A second confounding factor occurred mostly for the following two studies, in which interactions between the agent and the user have been observed. Within those interactions, more cues and factors than just the mere appearance are obtainable by the users such as the agent's voice, its nonverbal behavior, the interaction content or response time to the user's statements. Accordingly, all these variables can affect the perception and evaluation of the agent. To control these factors, overall the same interaction scripts, nonverbal behavior, and voice have been used. Nevertheless, especially with perceived realism, the agent's behavior was found to influence the effect of appearance variables (Bailenson et al., 2005).

In terms of the used interactions and the overall topic, that has been applied (virtual assistance with health-related contents), a further limitation occurred. For all studies the same application field and even interaction contents have been chosen, to enhance the comparability of the studies and present implications for virtual assistants. Besides these positive effects, it has also to be marked critically, because the found effects are not easily transferable to other application fields since prior research highlighted the moderating effect of the application (Ring et al., 2014). On the one hand, it is important to differentiate which appearance variables are the most effective for a certain application. On the other hand, examinations of the described effects within different application fields could also highlight potential similarities, which could strengthen the robustness of the evidence regarding social processes.

While a wide range of methodological approaches was used (qualitative, quantitative, interaction study and long-term study), all have been conducted under artificial conditions (e.g., based on scenarios or lab studies). While this is a strength to exclude as many confounding variables as possible, still it has to be investigated how the presented findings are transferable into the field. Especially with regard to the long-term effects and actual usage behavior, field studies are needed.

Like most studies in the realm of the media equation theory (Reeves & Nass, 1996), the outcomes and effects of the virtual agents have not been tested in comparison to outcomes evoked by a real human. It can be assumed, that the outcomes of a real human are stronger compared to virtual agents. However, the current dissertation does not aim to exchange a human by a virtual agent but to supplement the users' life with a virtual agent. Therefore, it is not necessary that the same outcomes that a human might produce also occur for a virtual agent.

14 FUTURE WORK

As discussed with regard to the definition of virtual agents, the agency was found to affect the social effects an artificial entity evokes (see for a recent overview Fox et al., 2015). Especially regarding appearance variables, different processes might underlay the evaluation of the appearance of an agent compared to an avatar. Since avatars represent the users, factors like identification or similarity might be even more important. Moreover, impression management processes might occur. As in the present study, only the appearance of virtual agents have been investigated; it would be interesting to compare these findings with

outcomes respective avatars. Hence, future research could investigate the effects of the agency on the evaluation of species and realism.

The present work delivers a categorization of appearance variables. Although this categorization is well constructed based on prior research, it is still incomplete and to some respects inaccurate. To systemize appearance outcomes even more and give developers a stricter guideline with regard to the agent's appearance, it would be very promising to specify the categories more. Here, designers, psychologists, and computer scientists should work together to accomplish a catalog of appearance variables with strict descriptions. For instance, concerning realism, it has to be defined what the highest degree of realism is and in which steps the manipulations can be done. It needs further be defined at which point proportions have to be described as stylized or when the shade of a virtual figure is stylized. To construct such categorization, the definition of the appearance categories by the potential user should also be taken into account. Among the user's different expectations and different definitions of for instant realism or stylization are observable. The users' definitions might supplement the concepts constructed by the researchers and designers. When such a precise catalog of appearance variables would exist, research that is even more systematic could be done. So far, every author defined the forms of appearance variables (e.g., the degree of realism) on his or her own. As a consequence, aspects described as realistic in one study might already be described as stylized in another study. Building on appearances categorizations (such as the one presented in this work or future catalogs that have been refined), the comparability and generalizability of the whole research field could be enhanced. Additionally, higher comparability of the results leads to more precise design implications.

As described in the limitation section (Chapter 13), only some of the multiple appearance categories have been investigated. Thus, future research should investigate the remaining categories to receive broader insights in appearance effects. Moreover, replications are needed. Especially since the appearance of a virtual agent is always confounded by other variables that are not in focus, studies that can replicate the findings of this work would help to enhance the generalizability of the results.

Especially regarding the long-term component of the present work, field studies are needed. The artificial nature of the lab studies might infer the usage of and interaction with the agent, which leads to different social outcomes. Field studies could help to give valuable insights into the effect of appearances within a real interaction at the user's homes. When the agent is embedded in users' households, other variables might affect the preferences about the agent's appearance. As for instance, the agent might also be observable for guests or other

persons than the actual user; social desirability could affect the selection and resonance to a specific appearance. Moreover, the effects with regard to social outcomes such as trust and the interpersonal relationship have to be investigated within the users' homes and over a longer period than used in Study 4.

As the impact of user characteristics such as age or users' attitude has been demonstrated in this dissertation, it is needed to adapt the appearance to the target-group when a maximal beneficial outcome is aimed. However, when virtual assistants are embedded into the users' homes, mostly not only one specific user might interact with the agent. Especially in families, in which people of different age groups live together and might interact all with the same virtual assistant, the question has to be posed how an appearance have to be designed that fits all users. One could assume, that multiple agents are used or one agent with multiple appearances to meet all users' needs. However, this might have other undesired effects, and the evaluation of such a multi-agent system has not been investigated yet. With respect to tasked related differences, the same question occurs. Do users wish for multiple agents or is there an appearance that applies to multiple tasks and users? Future research could investigate these effects also combined with the aforementioned field studies.

As the overall results indicate that the interaction with an agent affects its perception, it is also important to investigate the effect of interaction cues on the evaluation of the agent's appearance. Prior studies indicate that the helpfulness of a virtual agent interacts with the similarity of the appearance: the effect of similarity was found to turn into the negative when the agent was unhelpful (van Vugt et al., 2010). While in the study of Vugt and co-workers (2010) facial similarity was used, it should be investigated, whether this effect also confounds a broader sense of similarity. Do people excuse interaction failures more to non-humanoid agents than to humanoid (which can be assumed to be perceived as more similar)? Additionally, the nonverbal behavior was also found to interact with appearance variables (Bailenson et al., 2005). Against this background, it is also promising to investigate further interaction cues that might interact with the agent's appearance.

The implications regarding the customization of the agent's appearance derived from Study 1, call for an experimental examination of potential age-related differences in the design process and the acceptance of such a process. It has to be investigated, whether the customization process has positive effects on the evaluation of the agent for seniors, although they stated to dislike such a procedure. In contrast, when the selection out of options is more promising, the number of options has to be investigated. In accordance with the research by Diehl and colleagues (2017), it has to be investigated which number of options has the most

positive effects for seniors and whether this number differs from other user groups. Additionally, it might be worth to investigate, if users like to customize only specific appearance variables, while others should better be pre-defined.

15 CONCLUSION

The present dissertation investigated the effect of a virtual agent's appearance on social relations between the user and the agent in a long-term interaction. Since prior appearance research was found to use mostly a rather unsystematic approach, a categorization of appearance variables has been constructed. Along these categories, the effects of a virtual agent's appearance with respect to social outcomes have been investigated by a multi-methodological approach with four empirical studies (ranging from qualitative interviews to a long-term lab study). In qualitative interviews (Study 1), the categories species and realism have been found to be of special interest. Moreover, the results indicate the importance to consider the user group since consistent differences between seniors and students have been found. Seniors were found to rely on the appearance and prefer an appearance that is familiar to them, while students want to differentiate the real world from the virtual. Study 2 supported these findings partly. While seniors mostly evaluated robot appearances less positive than human and animals, a younger sample showed higher liking and usage intention for robot appearances. Additionally, it became apparent that seniors used the agent's species for their evaluation, while students focused on realism. The findings from the first two studies regarding seniors are also transferable to a real human-agent interaction (Study 3). Humanoid appearances evoke the feeling of familiarity, and the higher social cues of a humanoid appearance foster trust and bonding towards the agent, which enhance usage intention. However, unlike the statements presented in the interview study, no realistic appearance is needed to trigger these effects. This dissertation investigated the effect of the number of interaction on appearance effects for the first time (Study 4). Unlike prior assumptions from former research (Bergmann et al., 2012), no interaction of appearance and number of interactions occurred. However, the number of interactions effected the evaluation of the agent (such as trustworthiness or closeness). Nevertheless, it has to be mentioned that in the long-term study overall little effects of appearance emerged. Therefore, it is important to investigate the effect of other appearance variables over time. One striking finding from the long-term study is that attractive agents can foster users' intention to live healthily. Therefore, an attractive appearance can also have promising effects for students, even when they are less sensitive for appearance effects.

Hence, not all users seem to have their eyes on the agent's appearance, when they evaluate the agent, but still, appearance is a crucial variable to create promising outcomes.

V REFERENCES

Ajzen, I. (1985). From Intentions to Actions: A Theory of Planned Behavior. In *Action Control* (pp. 11–39). Berlin, Heidelberg: Springer. http://doi.org/10.1007/978-3-642-69746-3_2

Appel, J., von der Pütten, A. M., Krämer, N. C., & Gratch, J. (2012). Does humanity matter? Analyzing the importance of social cues and perceived agency of a computer system for the emergence of social reactions during human-computer interaction. *Advances in Human-Computer Interaction*, *13*. http://doi.org/10.1155/2012/324694

Aron, A., Aron, E. N., & Smollan, D. (1992). Inclusion of Other in the Self Scale and the Structure of Interpersonal Closeness. *Journal of Personality and Social Psychology*, *63*(4), 596–612. http://doi.org/10.1037/0022-3514.63.4.596

Bailenson, J. N., Swinth, K., Hoyt, C., Persky, S., Dimov, A., & Blascovich, J. (2005). The Independent and Interactive Effects of Embodied-Agent Appearance and Behavior on Self-Report, Cognitive, and Behavioral Markers of Copresence in Immersive Virtual Environments. *Presence: Teleoperators and Virtual Environments*, *14*(4), 379–393. http://doi.org/10.1162/105474605774785235

Balakrishnan, K., & Honavar, V. (2001). Evolutionary and Neural Synthesis of Intelligent Agents. In M. J. Patel, V. Honavar, & K. Balakrishnan (Eds.), *Advances in the evolutionary synthesis of intelligent agents* (pp. 1–27). MIT Press.

Bartneck, C., Kulić, D., Croft, E., & Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics*, *1*(1), 71–81. http://doi.org/10.1007/s12369-008-0001-3

Baylor, A. L. (2009a). Promoting motivation with virtual agents and avatars: role of visual presence and appearance. *Philosophical Transactions of the Royal Society of London.*, *364*, 3559–3565. http://doi.org/doi:10.1098/rstb.2009.0148

Baylor, A. L. (2009b). Promoting motivation with virtual agents and avatars: role of visual presence and appearance. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *364*(1535), 3559–65. http://doi.org/10.1098/rstb.2009.0148

Baylor, A. L., & Kim, Y. (2003). Validating Pedagogical Agent Roles: Expert , Motivator , and Mentor. In *Proceedings of the World Conference on Educational Multimedia,*

Hypermedia and Telecommunications (pp. 463–466).

Baylor, A. L., & Kim, Y. (2004). Pedagogical agent design: The impact of agent realism, gender, ethnicity, and instructional role. *Intelligent Tutoring Systems*, (1997), 592–603. http://doi.org/10.1007/978-3-540-30139-4_56

Baylor, A. L., & Kim, Y. (2005). Simulating Instructional Roles through Pedagogical Agents. *International Journal of Artificial Intelligence in Education*, *15*, 95–115. http://doi.org/10.1007/BF02504991

Baylor, A. L., Shen, E., & Warren, D. (2004). Supporting learners with math anxiety: the impact of pedagogical agent emotional and motivational support. In *Proceedings of the International Conference on Intelligent User Interfaces, San Diego, CA*.

Baylor, A. L., Warren, D., Park, S., Shen, E., & Perez, R. (2005). The impact of frustration-mitigating messages delivered by an interface agent. In *Proceedings of the International Conference on Artificial Intelligence in Education* (p. 1012).

Beer, J. M., Smarr, C.-A., Fisk, A. D., & Rogers, W. A. (2015). Younger and older users′ recognition of virtual agent facial expressions. *International Journal of Human-Computer Studies*, *75*, 1–20. http://doi.org/10.1016/j.ijhcs.2014.11.005

Bergmann, K., Eyssel, F., & Kopp, S. (2012). A second chance to make a first impression? how appearance and nonverbal behavior affect perceived warmth and competence of virtual agents over time. In *Intelligent Virtual Agents* (pp. 126–138). http://doi.org/10.1007/978-3-642-33197-8-13

Berscheid, E., & Walster, E. (1974). Physical attractiveness. *Advances in Experimental Social Psychology*, *7*(1), 167–215.

Bickmore, T. W. (2005). Establishing and Maintaining Long-Term Human-Computer Relationships. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *12*(2), 293–327. http://doi.org/10.1145/1067860.1067867

Bickmore, T. W., Caruso, L., & Clough-Gorr, K. (2005). Acceptance and usability of a relational agent interface by urban older adults. *CHI '05 Extended Abstracts on Human Factors in Computing Systems - CHI '05*, 1212. http://doi.org/10.1145/1056808.1056879

Bickmore, T. W., Caruso, L., Clough-Gorr, K., & Heeren, T. (2005). "It"s just like you talk to a friend' relational agents for older adults. *Interacting with Computers*, *17*(6), 711–735. http://doi.org/10.1016/j.intcom.2005.09.002

Bickmore, T. W., Gruber, A., & Picard, R. (2005). Establishing the computer – patient working alliance in automated health behavior change interventions. *Patient Education*

and Counseling, *59*, 21–30. http://doi.org/10.1016/j.pec.2004.09.008

Bickmore, T. W., Mitchell, S. E., Jack, B. W., Paasche-Orlow, M. K., Pfeifer, L. M., & O'Donnell, J. (2010). Response to a relational agent by hospital patients with depressive symptoms. *Interacting with Computers*, *22*(4), 289–298. http://doi.org/10.1016/j.intcom.2009.12.001

Bickmore, T. W., Pfeifer, L. M., Byron, D., Forsythe, S., Henault, L. E., Jack, B. W., … Paasche-Orlow, M. K. (2010). Usability of conversational agents by patients with inadequate health literacy: evidence from two clinical trials. *Journal of Health Communication*, *15*(2), 197–210. http://doi.org/10.1080/10810730.2010.499991

Bickmore, T. W., & Schulman, D. (2009). A virtual laboratory for studying long-term relationships between humans and virtual agents. In *Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems-Volume* (pp. 297–304). Retrieved from http://dl.acm.org/citation.cfm?id=1558054

Bickmore, T. W., Schulman, D., & Yin, L. (2010). Maintaining Engagement in Long-Term Interventions With Relational Agents. *Applied Artificial Intelligence*, *24*(6), 648–666. http://doi.org/10.1080/08839514.2010.492259

Blascovich, J., Loomis, J., Beall, A. C., Swinth, K. R., Hoyt, C. L., & Bailenson, J. N. (2002). Immersive virtual environment technology as a methodological tool for social psychology. *Psychological Inquiry*, *13*(2), 103–124. http://doi.org/10.1207/S15327965PLI1302_01

Brigham, J. C. (1980). Limiting conditions of the "physical attractiveness stereotype": Attributions about divorce. *Journal of Research in Personality*, *14*(3), 365–375. http://doi.org/10.1016/0092-6566(80)90019-7

Byrne, D. (1997). An Overview (and underview) of research and theory within the attraction Paradigm. *Journal of Social and Personal Relationships*, *14*(3), 413–431.

Carbon, C. C., & Leder, H. (2005). The Repeated Evaluation Technique (RET). A method to capture dynamic effects of innovativeness and attractiveness. *Applied Cognitive Psychology*, *19*(5), 587–601. http://doi.org/10.1002/acp.1098

Cassell, J. (2000). More Than Just Another Pretty Face : Embodied Conversational Interface Agents. *Communications of the ACM*, *43*(4), 70–78. http://doi.org/10.1145/332051.332075

Chattaraman, V., Kwon, W.-S., Gilbert, J. E., & Shim, S. I. (2011). Virtual agents in e-commerce: representational characteristics for seniors. *Journal of Research in Interactive Marketing*, *5*(4), 276–297. http://doi.org/10.1108/17505931111191492

Colquitt, J. A., Scott, B. A., & LePine, J. A. (2007). Trust, Trustworthiness, and Trust Propensity: A Meta-Analytic Test of Their Unique Relationships With Risk Taking and Job Performance. *Journal of Applied Psychology*, *92*(4), 909–927. http://doi.org/10.1037/0021-9010.92.4.909

Davis, F. D. (1989). Perceived Usefulness , Perceived Ease of Use , and User Acceptance ofInformation Technology. *MIS Quarterly*, *13*(3), 319–340. http://doi.org/www.jstor.org/stable/249008

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User Acceptance of Computer Technology : A Comparison of Two Theoretical Models. *Management Science*, *35*(8), 982–1003. http://doi.org/www.jstor.org/stable/2632151

Davison, W. P. (1983). The Third-Person Effect in Communication. *Public Opinion Quarterly*, *47*(1), 1–15. http://doi.org/10.1086/268763

De Kort, Y. A. W., & Ijsselsteijn, W. A. (2008). People, places, and play: a research framework for digital game experiences in socio-spatial context. *ACM Computers in Entertainment*, *6*(2), 18. http://doi.org/10.1145/1371216.1371221

De Vault, D., Artstein, R., Benn, G., Dey, T., Fast, E., Gainer, A., … Morency, L.-P. (2014). SimSensei Kiosk: A Virtual Human Interviewer for Healthcare Decision Support. In *Proceedings of the 2014 International Conference on Autonomous Agents and Multi-Agent Systems* (pp. 1061–1068).

Deska, J. C., Lloyd, E. P., & Hugenberg, K. (2018). Facing humanness: Facial width-to-height ratio predicts ascriptions of humanity. *Journal of Personality and Social Psychology*, *114*(1), 75–94. http://doi.org/10.1037/pspi0000110

Diehl, C., Schiffhauer, B., Eyssel, F., Achenbach, J., Klett, S., Botsch, M., & Kopp, S. (2017). Get One or Create One: the Impact of Graded Involvement in a Selection Procedure for a Virtual Agent on Satisfaction and Suitability Ratings. In *Proceedings of the Intelligent Virtual Agents* (pp. 109–118). http://doi.org/10.1007/978-3-319-67401-8_13

Dion, K., Berscheid, E., & Walster, E. (1972). What is Beautiful is Good. *Journal of Personality and Social Psychology*, *24*(3), 285–290. http://doi.org/10.1177/0956797610388048

Doering, A., Veletsianos, G., & Yerasimou, T. (2008). Conversational Agents and their Longitudinal Affordance on Communication and Interaction. *Journal of Interactive Learning Research*, *19*(2), 251–270.

Eckes, T. (2002). Paternalistic and envious gender stereotypes: Testing predictions from the stereotype content model. *Sex Roles*, *47*(3/4), 99–114.

http://doi.org/10.1023/A:1021020920715

Emmerich, K., & Masuch, M. (2016). The Influence of Virtual Agents on Player Experience and Performance. In *Proceedings of CHI PLAY* (Vol. 412, pp. 10–21). http://doi.org/10.1088/1742-6596/412/1/012003

Fehr, B. (2000). The Life Cycle of Friendship. In C. Hendrick & S. S. Hendrick (Eds.), *Close Relationships: A Sourcebook* (pp. 70–82). Thousand Oaks, CA: SAGE Publications, Inc. http://doi.org/10.4135/9781452220437.n6

Feingold, A. (1992). Good-Looking People Are Not What We Think. *Psychological Bulletin*, *111*(2), 304–341. http://doi.org/10.1037/0033-2909.111.2.304

Field, A. (2018). Andy Field - Discovering Statistics Using SPSS. Sage Publications. http://doi.org/10.1111/j.1365-2648.2007.04270_1.x

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, *82*(6), 878–902. http://doi.org/10.4324/9781315187280

Fogg, B. J. (2000). Persuasive technology: using computers to change what we think and do. *Ubiquity*, *5*. http://doi.org/10.1016/S0749-3797(99)00093-8

Forlizzi, J., Zimmerman, J., Mancuso, V., & Kwak, S. (2007). How interface agents affect interaction between humans and computers. *Proceedings of the 2007 Conference on Designing Pleasurable Products and Interfaces - DPPI '07*, (August), 209. http://doi.org/10.1145/1314161.1314180

Fox, J., Ahn, S. J. G., Janssen, J. H., Yeykelis, L., Segovia, K. Y., & Bailenson, J. N. (2015). Avatars versus agents: A meta-analysis quantifying the effect of agency on social influence. *Human-Computer Interaction*, *30*(5), 401–432. http://doi.org/10.1080/07370024.2014.921494

Fuchslocher, A., Niesenhaus, J., & Krämer, N. C. (2011). Serious games for health: An empirical study of the game " Balance" for teenagers with diabetes mellitus. *Entertainment Computing*, *2*(2), 97–101. http://doi.org/10.1016/j.entcom.2010.12.001

Gerling, K., Fuchslocher, A., Schmidt, R., Krämer, N. C., & Masuch, M. (2011). Designing and evaluating casual health games for children and teenagers with cancer. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *6972 LNCS*, 198–209. http://doi.org/10.1007/978-3-642-24500-8_21

Groom, V., Nass, C., Chen, T., Nielsen, A., Scarborough, J. K., & Robles, E. (2009).

Evaluating the effects of behavioral realism in embodied agents. *International Journal of Human Computer Studies*, *67*(10), 842–849. http://doi.org/10.1016/j.ijhcs.2009.07.001

Gulz, A. (2005). Social enrichment by virtual characters – differential benefits. *Journal of Computer Assisted Learning*, *21*(6), 405–418. http://doi.org/10.1111/j.1365-2729.2005.00147.x

Gulz, A., & Haake, M. (2006a). Design of animated pedagogical agents - A look at their look. *International Journal of Human Computer Studies*, *64*(4), 322–339. http://doi.org/10.1016/j.ijhcs.2005.08.006

Gulz, A., & Haake, M. (2006b). Visual Design of Virtual Pedagogical Agents : Naturalism versus Stylization in Static Appearance. In *Proceedings of the International Design and Engagability Conference at NordiCHI* (Vol. 2006, pp. 1–9). http://doi.org/10.1007/11821830_44

Hampton, A. J., Fisher Boyd, A. N., & Sprecher, S. (2018). You're like me and I like you. *Journal of Social and Personal Relationships*, 1–24. http://doi.org/10.1177/0265407518790411

Hanus, M. D., & Fox, J. (2015). Assessing the effects of gamification in the classroom: A longitudinal study on intrinsic motivation, social comparison, satisfaction, effort, and academic performance. *Computers and Education*, *80*, 152–161. http://doi.org/10.1016/j.compedu.2014.08.019

Hayes, A. F. (2013). *Introduction to meditaion, moderation, and conditional process analysis: A regression-based approach*. New York, NY: Guilford Press. http://doi.org/10.5539/ass.v11n9p207

Heerink, M., Kröse, B., Evers, V., & Wielinga, B. (2010). Assessing acceptance of assistive social agent technology by older adults: The almere model. *International Journal of Social Robotics*, *2*(4), 361–375. http://doi.org/10.1007/s12369-010-0068-5

Hoffmann, L., Krämer, N. C., Lam-chi, A., & Kopp, S. (2009). Media Equation Revisited: Do Users Show Polite Reactions towards an Embodied Agent? In *Proceeding of the International Workshop on Intelligent Virtual Agents* (pp. 159–165). SpringerBerling, Heidelberg. Retrieved from http://link.springer.com/10.1007/978-3-642-04380-2_19

Horvath, A. O., & Greenberg, L. S. (1986). The development of the Working Alliance Inventory. In L. S. Greenberg & W. Pinsoff (Eds.), *The psychotherapeutic process: A research handbook* (pp. 529–556). New York, NY: Guildford Press.

Isbister, K., & Nass, C. (2000). Consistency of personality in interactive characters: Verbal cues, non-verbal cues, and user characteristics. *International Journal of Human*

Computer Studies, *53*(2), 251–267. http://doi.org/10.1006/ijhc.2000.0368

James, T. W., Potter, R. F., Lee, S., Kim, S., Stevenson, R. A., & Lang, A. (2015). How Realistic Should Avatars Be? *Journal of Media Psychology*, *27*(3), 109–117. http://doi.org/10.1027/1864-1105/a000156

Johnson, D. M., & Gardner, J. (2007). The Media Equation and team formation: Further evidences for experience as a moderator. *International Journal of Human-Computer Studies*, *65*(2), 111–124. http://doi.org/https://doi.org/10.1016/j.ijhcs.2006.08.007

Johnson, W. L., & Lester, J. C. (2015). Face-to-Face Interaction with Pedagogical Agents, Twenty Years Later. *International Journal of Artificial Intelligence in Education*, *25*(1), 25–36. http://doi.org/10.1007/s40593-015-0065-9

Kahn, P. H., Kanda, T., Ishiguro, H., Freier, N. G., Severson, R. L., Gill, B. T., … Shen, S. (2012). "Robovie, you'll have to go into the closet now": children's social and moral relationships with a humanoid robot. *Developmental Psychology*, *48*(2), 303–14. http://doi.org/10.1037/a0027033

Khan, R., & De Angeli, A. (2009). The attractiveness stereotype in the evaluation of embodied conversational agents. In *Lecture Notes in Computer Science: Vol. 5726 Conference on Human-Computer Interaction* (pp. 85–97). http://doi.org/10.1007/978-3-642-03655-2_10

Khan, R., & Sutcliffe, A. (2014). Attractive Agents Are More Persuasive. *International Journal of Human-Computer Interaction*, *30*(2), 1–9. http://doi.org/10.1080/10447318.2013.839904

Kiesler, S., & Sproull, L. (1997). "Social" human-computer interaction. In B. Friedman (Ed.), *Human values and the design of computer technology* (pp. 191–199). Stanford, CA: CSLI Publications. Retrieved from https://dl.acm.org/citation.cfm?id=278224

Kim, Y., & Sundar, S. S. (2012). Anthropomorphism of computers: Is it mindful or mindless? *Computers in Human Behavior*, *28*(1), 241–250. http://doi.org/10.1016/j.chb.2011.09.006

Klimmt, C., Hefner, D., Vorderer, P., Roth, C., & Blake, C. (2010). Identification with video game characters as automatic shift of self-perceptions. *Media Psychology*, *13*(4), 323–338. http://doi.org/10.1080/15213269.2010.524911

Koda, T., & Maes, P. (1996). Agents with faces: the effect of personification. *Proceeding of the 5th IEEE International Workshop on Robot and Human Communication. RO-MAN'96 TSUKUBA*, 189–194. http://doi.org/10.1109/ROMAN.1996.568812

Kopp, S., Brandt, M., Buschmeier, H., Cyra, K., Krämer, N. C., Kummert, F., … Wall, E.

(2018). Conversational Assistants for Elderly Users – The Importance of Socially Cooperative Dialogue. *AAMAS Workshop on Intelligent Conversation Agents in Home and Geriatric Care Applications*.

Krämer, N. C. (2008). *Soziale Wirkungen virtueller Helfer: Gestaltung und Evaluation von Mensch-Computer Interaktion*. Stuttgart: W. Kohlhammer Verlag.

Krämer, N. C., Rosenthal-von der Pütten, A. M., & Hoffmann, L. (2015). Social Effects of Virtual and Robot Companions. In S. S. Sundar (Ed.), *The Handbook of the Psychology of Communication Technology* (pp. 137–159). John Wiley & Sons. http://doi.org/10.1002/9781118426456.ch6

Krumhuber, E., Kappas, A., Hume, C., Hall, L., & Aylett, R. (2015). From Non-human to Human: Adult's and Children's Perceptions of Agents Varying in Humanness. In *Intelligent Virtual Agents* (pp. 471–474). http://doi.org/10.1007/978-3-319-21996-7_50

Langer, E. J. (1989). *Mindfulness*. Reading, MA, US: Addison-Wesley/Addison Wesley Longman.

Langlois, J. H., Kalakanis, L., Rubenstein, A. J., Larson, A., Hallam, M., & Smoot, M. (2000). Maxims or myths of beauty? A meta-analytic and theoretical review. *Psychological Bulletin*, *126*(3), 390–423. http://doi.org/10.1037/0033-2909.126.3.390

Leary, M. R., & Kowalski, R. M. (1990). Impression management: A literature review and two-component model. *Psychological Bulletin*, *107*(I), 34–47. http://doi.org/10.1037/0033-2909.107.1.34

Lee, E., & Nass, C. (2002). Experimental Tests of Normative Group Influence and Representation Effects in Computer-Mediated Communication When Interacting Via Computers Differs From Interacting With Computers. *Human Communication Research*, *28*(3), 349–381.

Lee, K. M., Jung, Y., Kim, J., & Kim, S. R. (2006). Are physically embodied social agents better than disembodied social agents?: The effects of physical embodiment, tactile interaction, and people's loneliness in human-robot interaction. *International Journal of Human Computer Studies*, *64*(10), 962–973. http://doi.org/10.1016/j.ijhcs.2006.05.002

Leiner, D. J. (2016). Our research's breadth lives on convenience samples A case study of the online respondent pool "SoSci Panel." *Studies in Communication | Media*, *5*(4), 367–396. http://doi.org/10.5771/2192-4007-2016-4-367

Levin, D. Z., Whitener, E. M., & Cross, R. (2006). Perceived trustworthiness of knowledge sources: The moderating impact of relationship length. *Journal of Applied Psychology*, *91*(5), 1163–1171. http://doi.org/10.1037/0021-9010.91.5.1163

Li, D. D., Liau, A. K., & Khoo, A. (2013). Player-Avatar Identification in video gaming: Concept and measurement. *Computers in Human Behavior*, *29*(1), 257–263. http://doi.org/10.1016/j.chb.2012.09.002

Li, J., Kizilcec, R., Bailenson, J. N., & Ju, W. (2015). Social robots and virtual agents as lecturers for video instruction. *Computers in Human Behavior*, *55*(B), 1222–1230. http://doi.org/10.1016/j.chb.2015.04.005

Liszio, S., Emmerich, K., & Masuch, M. (2017). The influence of social entities in virtual reality games on player experience and immersion. In *Proceedings of the International Conference on the Foundations of Digital Games - FDG '17* (pp. 1–10). http://doi.org/10.1145/3102071.3102086

Lunardo, R., Bressolles, G., & Durrieu, F. (2016). The interacting effect of virtual agents' gender and dressing style on attractiveness and subsequent consumer online behavior. *Journal of Retailing and Consumer Services*, *30*, 59–66. http://doi.org/10.1016/j.jretconser.2016.01.006

MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. *New York: Erlbaum and Taylor Francis Group*.

Marathe, S., & Sundar, S. S. (2011). What Drives Customization ? Control Or Identity ? *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 781–790. http://doi.org/10.1145/1978942.1979056

Mathes, E. W. (1975). The Effects of Physical Attractiveness and Anxiety on Heterosexual Attraction over a Series of Five Encounters The Effects of Physical Attractiveness and Anxiety on Heterosexual Attraction Over a Series of Five Encounters. *Source Journal of Marriage and Family*, *37*(4), 769–773. http://doi.org/10.2307/350828

Matsui, T., & Yamada, S. (2017). Two-Dimensional Mind Perception Model of Humanoid Virtual Agent. In *Proceedings of the 5th International Conference on Human Agent Interaction* (pp. 311–316). http://doi.org/10.1145/3125739.3125761

McAdams, D. P., & Constantian, C. A. (1983). Intimacy and affiliation motives in daily living: An experience sampling analysis. *Journal of Personality and Social Psychology*, *45*(4), 851–861. http://doi.org/10.1037/0022-3514.45.4.851

McDonnell, R., Breidt, M., & Bülthoff, H. H. (2012). Render me real? *ACM Transactions on Graphics*, *31*(4), 1–11. http://doi.org/10.1145/2185520.2185587

McGoldrick, P. J., Keeling, K. A., & Beatty, S. F. (2008). A typology of roles for avatars in online retailing. *Journal of Marketing Management*, *24*(3-4), 433–461. http://doi.org/10.1362/026725708X306176

McPherson, M., Smith-lovin, L., & Cook, J. M. (2001). Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology*, *27*, 415–444. http://doi.org/10.1146/annurev.soc.27.1.415

Meullemans, C. G. E., & Abraham, J. (2017). The Prediction of Sensitivity towards Uncanniness: An Applied Psychological Study on Novel Readers in the Light of Anthropomorphism and Attitudes towards Thrownness. *IOP Conference Series: Materials Science and Engineering*, *180*(1). http://doi.org/10.1088/1742-6596/755/1/011001

Montoya, R. M., Horton, R. S., & Kirchner, J. (2008). Is actual similarity necessary for attraction? A meta-analysis of actual and perceived similarity. *Journal of Social and Personal Relationships*, *25*(6), 889–922. http://doi.org/10.1177/0265407508096700

Moorman, C., & Matulich, E. (1993). A Model of Consumers ' Preventive Health Behaviors : The Role of Health Motivation and Health Ability, *20*(2), 208–228. http://doi.org/www.jstor.org/stable/2489270

Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley. *IEEE Robotics and Automation Magazine*, *19*(2), 98–100. http://doi.org/10.1109/MRA.2012.2192811

Nass, C., Fogg, B. J., & Moon, Y. (1996). Can computers be teammates? *International Journal of Human-Computer Studies*, *45*(6), 669–678. http://doi.org/10.1006/ijhc.1996.0073

Nass, C., & Moon, Y. (2000). Machines and Mindlessness: Social Responses to Computers. *Journal of Social Issues*, *56*(1), 81–103. http://doi.org/10.1111/0022-4537.00153

Nass, C., Moon, Y., & Carney, P. (1999). Are People Polite to Computers? Responses to Computer-Based Interviewing Systems1. *Journal of Applied Social Psychology*, *29*(5), 1093–1109. http://doi.org/10.1111/j.1559-1816.1999.tb00142.x

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. *Computer-Human Interaction (CHI) Conference: Celebrating Interdependence*, 72–78. http://doi.org/10.1145/259963.260288

Nass, C., Steuer, J., Tauber, E., & Reeder, H. (1993). Computers Agency , & Ethopoeia : as Social Actors. In *CHI '93 Conference Companion on Human Factors in Computing Systems* (pp. 111–112). http://doi.org/http://doi.acm.org/10.1145/259964.260137

Neave, N., Jackson, R., Saxton, T., & Hönekopp, J. (2015). The influence of anthropomorphic tendencies on human hoarding behaviours. *Personality and Individual Differences*, *72*, 214–219. http://doi.org/10.1016/j.paid.2014.08.041

Nguyen, H., Ishmatova, D., Tapanainen, T., Liukkonen, T. N., Katajapuu, N., Makila, T., &

Luimula, M. (2017). Impact of Serious Games on Health and Well-being of Elderly: A Systematic Review. *Proceedings of the 50th Hawaii International Conference on System Sciences*, 3695–3704. http://doi.org/10.24251/HICSS.2017.447

Nomura, T., Kanda, T., & Suzuki, T. (2006). Experimental investigation into influence of negative attitudes toward robots on human-robot interaction. *AI and Society*, *20*(2), 138–150. http://doi.org/10.1007/s00146-005-0012-7

Nomura, T., Kanda, T., Suzuki, T., & Kato, K. (2008). Prediction of human behavior in human-robot interaction using psychological scales for anxiety and negative attitudes toward tobots. *IEEE Transactions on Robotics*, *24*(2), 442–451. http://doi.org/10.1109/TRO.2007.914004

Norton, M. I., Mochon, D., & Ariely, D. (2012). The IKEA effect: When labor leads to love. *Journal of Consumer Psychology*, *22*(3), 453–460. http://doi.org/10.1016/j.jcps.2011.08.002

O'Connor, K. M., & Gladstone, E. (2018). Beauty and social capital: Being attractive shapes social networks. *Social Networks*, *52*, 42–47. http://doi.org/10.1016/j.socnet.2017.05.003

Parise, S., Kiesler, S., Sproull, L., & Waters, K. (1999). Cooperating with life- like interface agents. *Computers in Human Behavior*, *15*(2), 123–142.

Pfeifer, L. M., & Bickmore, T. W. (2011). Is the media equation a flash in the pan?: the durability and longevity of social responses to computers. *Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems*, 777–780. http://doi.org/http://doi.acm.org/10.1145/1978942.1979055

Phan Tran, M. K., Robert, P., & Bremond, F. (2016). A Virtual Agent for enhancing performance and engagement of older people with dementia in Serious Games. In *Paper presented at the Workshop Artificial Compagnon-Affect-Interaction*. http://doi.org/<http://www.enib.fr/wacai/>. <hal-01369878>

Poulsen, F. O., Holman, T. B., Busby, D. M., & Carroll, J. S. (2013). Physical attraction, attachment styles, and dating development. *Journal of Social and Personal Relationships*, *30*(3), 301–319. http://doi.org/10.1177/0265407512456673

Preacher, K. J. (2015). Advances in Mediation Analysis : A Survey and Synthesis of New Developments. http://doi.org/10.1146/annurev-psych-010814-015258

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, *40*(3), 879–891. http://doi.org/10.3758/BRM.40.3.879

Qiu, L., & Benbasat, I. (2010). A study of demographic embodiments of product

recommendation agents in electronic commerce. *International Journal of Human Computer Studies*, *68*(10), 669–688. http://doi.org/10.1016/j.ijhcs.2010.05.005

Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality*, *41*(1), 203–212. http://doi.org/10.1016/J.JRP.2006.02.001

Reeves, B., & Nass, C. (1996). *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. CSLI Publications and Cambridge University Press. http://doi.org/10.1109/MSPEC.1997.576013

Ren, J., Schulman, D., Jack, B., & Bickmore, T. W. (2014). Supporting longitudinal change in many health behaviors. *Proceedings of the Extended Abstracts of the 32nd Annual ACM Conference on Human Factors in Computing Systems - CHI EA '14*, 1657–1662. http://doi.org/10.1145/2559206.2581201

Rickenberg, R., & Reeves, B. (2000). The Effects of Animated Characters on Anxiety, Task Performance, and Evaluations of User Interfaces. In *SIGCHI conference on Human Factors in Computing Systems* (pp. 49–56). http://doi.org/10.1016/j.foodchem.2017.03.106

Riegelsberger, J., Counts, S., Farnham, S. D., & Philips, B. C. (2006). Sounds good to me: effects of photo and voice profiles on gaming partner choice. In *Proceedings of ACM CSCW'06 Conference on Computer-Supported Cooperative Work* (pp. 159–162). http://doi.org/10.1145/1180875.1180899

Ring, L., Barry, B., Totzke, K., & Bickmore, T. W. (2013). Addressing Loneliness and Isolation in Older Adults. *IEEE Humaine Association Conference on Affective Computing and Intelligent Intreraction*, 61–66. http://doi.org/10.1109/ACII.2013.17

Ring, L., Shi, L., Totzke, K., & Bickmore, T. W. (2015). Social support agents for older adults: longitudinal affective computing in the home. *Journal on Multimodal User Interfaces*, *9*(1), 79–88. http://doi.org/10.1007/s12193-014-0157-0

Ring, L., Utami, D., & Bickmore, T. W. (2014). The Right Agent for the Job ? The Effects of Agent Visual Appearance on Task Domain. In *Intelligent Virtual Agents* (pp. 374–384). http://doi.org/10.1007/978-3-319-09767-1_49

Ritchie, Jane & Lewis, J. (2003). *Qualitative Reasearch Practice A Guide for Social Science Students and Reasearchers*. (J. Ritchie, Jane & Lewis, Ed.)*Qualitative Research Practice - A Guide for Social Science Students and Researchers*. London: Sage Publications. http://doi.org/March 10, 2016

Rizzo, A. A., Scherer, S., Devault, D., Gratch, J., Artstein, R., Hartholt, A., … Morency.

(2014). Detection and computational analysis of psychological signals using a virtual human interviewing agent. In *Proceedings of the 10th Intl Conf. Disability, Virtual Reality & Associated Technologies* (pp. 2–4). http://doi.org/10.1111/j.1472-765X.2004.01606.x

Robertson, S., Solomon, R., Ridl, M., Wicklin Gillespie, T., Chociemski, T., Master, V., & Mohan, A. (2015). The Visual Design and Implementation of an Embodied Conversational Agent in a Shared Decision-Making Context (eCoach). In *Lecture Notes in Computer Science: Learning and Collaboration Technologies* (pp. 427–437). http://doi.org/10.1007/978-3-319-20609-7

Rocha, T., Bessa, M., Bastardo, R., & Magalhães, L. (2018). Image-type representation: A preliminary study on preferences of users with intellectual disabilities. *International Journal of Human Computer Studies*, *110*, 1–11. http://doi.org/10.1016/j.ijhcs.2017.09.003

Roe, B., Whattam, M., Young, H., & Dimond, M. (2001). Elders' needs and experiences of receiving formal and informal care for their activities of daily living. *Journal of Clinical Nursing*, *10*(3), 389–397. http://doi.org/10.1046/j.1365-2702.2001.00485.x

Rosenberg-Kima, R. B., Baylor, A. L., Plant, E. A., & Doerr, C. E. (2008). Interface agents as social models for female students: The effects of agent visual presence and appearance on female students' attitudes and beliefs. *Computers in Human Behavior*, *24*(6), 2741–2756. http://doi.org/10.1016/j.chb.2008.03.017

Rosenthal-Von Der Pütten, A. M. (2014). *Uncannily Human. Empirical Investigation of the Uncanny Valley Phenomenon*.

Rosenthal-von der Pütten, A. M., Bock, N., & Brockmann, K. (2017). Not Your Cup of Tea ? How Interacting With a Robot Can Increase Perceived Self-efficacy in HRI and Evaluation. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction - HRI '17* (pp. 483–492). http://doi.org/10.1145/2909824.3020251

Rosenthal-von der Pütten, A. M., Straßmann, C., & Krämer, N. C. (2018). Dominant and submissive nonverbal behavior of virtual agents and its effects on evaluation and negotiation outcome in different age groups. *Computers in Human Behavior*, (Advance online publication). http://doi.org/10.1016/J.CHB.2018.08.047

Rosenthal-von der Pütten, A. M., & Weiss, A. (2015). The uncanny valley phenomenon: Does it affect all of us? *Interaction Studies*, *16*(2), 206–214. http://doi.org/10.1075/is.16.2.07ros

Ryokai, K., Vaucelle, C., & Cassell, J. (2003). Virtual peers as partners in storytelling and

literacy learning. *Journal of Computer Assisted Learning*, *19*(2), 195–208. http://doi.org/10.1046/j.0266-4909.2003.00020.x

Saidatul Maizura, S., Farah, M. Z., Nabila, A. N. K., Noorizdayantie, S., Zuraidah, A. R., Omar, M., … Wong, S. L. (2010). The pedagogical agent in online learning : Effects of the degree of realism on achievement in terms of gender. *Contemporary Educational Technology*, *1*(2), 175–185. Retrieved from http://dergipark.gov.tr/download/article-file/252129

Sandoval, E. B., Mubin, O., & Obaid, M. (2014). Human robot interaction and fiction: A contradiction. In *International Conference on Social Robotics* (pp. 54–63). Springer, Cham. http://doi.org/10.1007/978-3-319-11973-1_6

Schwind, V., Leicht, K., Jäger, S., Wolf, K., & Henze, N. (2017). Is there an Uncanny Valley of Virtual Animals? A Quantitative and Qualitative Investigation. *International Journal of Human-Computer Studies*, *111*, 49–61. http://doi.org/10.1016/j.ijhcs.2017.11.003

Schwind, V., Wolf, K., Henze, N., & Korn, O. (2015). Determining the Characteristics of Preferred Virtual Faces Using an Avatar Generator. *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '15*, 221–230. http://doi.org/10.1145/2793107.2793116

Scopelliti, M., Giuliani, M. V., & Fornara, F. (2005). Robots in a domestic setting: A psychological approach. *Universal Access in the Information Society*, *4*(2), 146–155. http://doi.org/10.1007/s10209-005-0118-1

Shaked, N. A. (2017). Avatars and virtual agents – relationship interfaces for the elderly. *Healthcare Technology Letters*, *4*(3), 83–87. http://doi.org/10.1049/htl.2017.0009

Shiban, Y., Schelhorn, I., Jobst, V., Hörnlein, A., Puppe, F., Pauli, P., & Mühlberger, A. (2015). The appearance effect: Influences of virtual agent features on performance and motivation. *Computers in Human Behavior*, *49*, 5–11. http://doi.org/10.1016/j.chb.2015.01.077

Sobieraj, S., & Krämer, N. C. (2014). What Is Beautiful in Cyberspace? Communication with attractive avatars. In *International Conference on Social Computing and Social Media* (pp. 125–136). Springer, Cham. http://doi.org/10.1007/978-3-319-07632-4_12

Söbke, H., & Streicher, A. (2016). Serious games architectures and engines. In *Entertainment Computing and Serious Games* (pp. 148–173). Springer, Cham. http://doi.org/10.1007/978-3-319-46152-6_7

Soliman, M., & Guetl, C. (2010). Intelligent Pedagogical Agents in Immersive Virtual Learning Environments: A Review. In *Proceeding of the 33rd International Convention*

MIPRO (pp. 827–832). IEEE.

SpiegelOnline. (2018). Pflege: Krankenhäusern fehlen 80.000 Pflegekräfte. Retrieved October 27, 2018, from http://www.spiegel.de/wirtschaft/soziales/pflege-krankenhaeuser-muessten-stellen-um-22-prozent-aufstocken-a-1213558.html

Spiekman, M. E., Haazebroek, P., & Neerincx, M. A. (2011). Requirements and Platforms for Social Agents that Alarm and Support Elderly Living Alone. In *International Conference on Social Robotics* (pp. 226–235). Springer, Berlin, Heidelberg. http://doi.org/10.1007/978-3-642-25504-5_23

Stirrat, M., & Perrett, D. I. (2010). Valid facial cues to cooperation and trust: Male facial width and trustworthiness. *Psychological Science*, *21*(3), 349–354. http://doi.org/10.1177/0956797610362647

Sträfling, N., Fleischer, I., Polzer, C., Leutner, D., & Krämer, N. C. (2010). Teaching learning strategies with a pedagogical agent: The effects of a virtual tutor and its appearance on learning and motivation. *Journal of Media Psychology*, *22*(2), 73–83. http://doi.org/10.1027/1864-1105/a000010

Straßmann, C., & Krämer, N. C. (2017). A Categorization of Virtual Agent Appearances and a Qualitative Study on Age-Related User Preferences. In *Intelligent Virtual Agents* (pp. 413–422). Springer, Cham. http://doi.org/10.1007/978-3-319-67401-8_51

Straßmann, C., & Krämer, N. C. (2018). A Two-Study Approach to Explore the Effect of User Characteristics on Users' Perception and Evaluation of a Virtual Assistant's Appearance. *Multimodal Technologies and Interaction*, *2*(66). http://doi.org/10.3390/mti2040066

Straßmann, C., Rosenthal-von der Pütten, A. M., Yaghoubzadeh, R., Kaminski, R., & Krämer, N. C. (2016). The Effect of an Intelligent Virtual Agent's Nonverbal Behavior with Regard to Dominance and Cooperativity. In *International Conference on Intelligent Virtual Agents* (pp. 15–28). Springer, Cham. http://doi.org/10.1007/978-3-319-47665-0_2

Stronks, B., Nijholt, A., van der Vet, P., & Heylen, D. (2002). Designing for friendship: Becoming friends with your ECA. In *Workshop Embodied Conversational Agents - Let's specify and evaluate them!, AMAAS* (pp. 91–96). Retrieved from http://doc.utwente.nl/63371/

Sundar, S. S., & Nass, C. (2000). Source Orientation in Human-Computer Interaction. *Communication Research*, *27*(6), 683–703.

Terada, K., Jing, L., & Seiji Yamada. (2015). Effects of Agent Appearance on Customer

Buying Motivations on Online Shopping Sites. *CHI' 15 Extended Abstracts, Apr 18-23,* 929–934.

Tomasello, M. (2014). The ultra-social animal. *European Journal of Social Psychology*, *44*(3), 187–194. http://doi.org/10.1002/ejsp.2015

Trinh, H., Ring, L., & Bickmore, T. W. (2015). DynamicDuo. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*, 1739–1748. http://doi.org/10.1145/2702123.2702584

Tsiourti, C., Ben-Moussa, M., Quintas, J., Loke, B., Jochem, I., Albuquerque Lopes, J., & Konstantas, D. (2016). A Virtual Assistive Companion for Older Adults: Design Implications for a Real-World Application. *SAI Intelligent Systems Conference (IntelliSys 2016)*, 556–566.

Tsiourti, C., Joly, E., Wings, C., Ben Moussa, M., & Wac, K. (2014). Virtual Assistive Companions for Older Adults: Qualitative Field Study and Design Implications. In *Proceedings of the 8th International Conference on Pervasive Computing Technologies for Healthcare* (pp. 57–64). http://doi.org/10.4108/icst.pervasivehealth.2014.254943

Urber, K. A., Degirmencioglu, S. M., & Tolson, J. M. (1998). Adolescent Friendship selcetion and Termination: The Role Of Similarity. *Journal of Social and Personal Relationships*, *15*(5), 703–710.

van Vugt, H. C., Bailenson, J. N., Hoorn, J. F., & Konijn, E. a. (2010). Effects of facial similarity on user responses to embodied agents. *ACM Transactions on Computer-Human Interaction*, *17*(2), 1–27. http://doi.org/10.1145/1746259.1746261

van Vugt, H. C., Konijn, E. A., Hoorn, J. F., & Veldhuis, J. (2006). Why Fat Interface Characters Are Better e-Health Advisors. In *Intelligent Virtual Agents* (pp. 1–13). Springer, Cham. http://doi.org/10.1007/11821830_1

van Wissen, A., Vinkers, C., & van Halteren, A. (2016). Developing a Virtual Coach for Chronic Patients: A User Study on the Impact of Similarity, Familiarity and Realism. In *International Conference on Persuasive Technology* (pp. 263–275). Springer, Cham. http://doi.org/10.1007/978-3-319-31510-2_23

Vardoulakis, L. P., Ring, L., Barry, B., Sidner, C. L., & Bickmore, T. W. (2012). Designing Relational Agents as Long Term Social Companions for Older Adults. In *Intelligent Virtual Agents* (Vol. 7502, pp. 289–302). Springer, Cham. http://doi.org/10.1007/978-3-642-33197-8_30

Venkatesh, V., & Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal Studies. *Management Science*, *46*(2), 186–205.

http://doi.org/10.1287/mnsc.46.2.186.11926

Volante, M., Babu, S. V., Chaturvedi, H., Newsome, N., Ebrahimi, E., Roy, T., … Fasolino, T. (2016). Effects of Virtual Human Appearance Fidelity on Emotion Contagion in Affective Inter-Personal Simulations. *IEEE Transactions on Visualization and Computer Graphics*, *22*(4), 1326–1335. http://doi.org/10.1109/TVCG.2016.2518158

von der Pütten, A. M., Krämer, N. C., Gratch, J., & Kang, S. H. (2010). "It doesn't matter what you are!" Explaining social effects of agents and avatars. *Computers in Human Behavior*, *26*(6), 1641–1650. http://doi.org/10.1016/j.chb.2010.06.012

Wakefield, J. (2016). Would you want to talk to a machine? - BBC News. Retrieved October 26, 2018, from https://www.bbc.com/news/technology-36225980

Walster, E., Aronson, V., & Abrahams, D. (1966). Importance of Physical Attractiveness in Dating Behavior. *Journal of Personality and Social Psychology*, *4*(5), 508–516.

Weisband, S., & Kiesler, S. (1996). Self disclosure on computer forms. In *Proceedings of the SIGCHI conference on Human factors in computing systems common ground - CHI '96* (pp. 3–10). http://doi.org/10.1145/238386.238387

Wells, J. D., Campbell, D. E., Valacich, J. S., & Featherman, M. (2010). The Effect of Perceived Novelty on the Adoption of Information Technology Innovations: A Risk/Reward Perspective. *Decision Sciences*, *41*(4), 813–843. http://doi.org/10.1111/j.1540-5915.2010.00292.x

Wooldridge, M., & Jennings, N. R. (1995). Intelligent agents: theory and practive. *The Knowledge Engineering Review*, *10*(2), 115–152. http://doi.org/10.1017/S0269888900008122

Xu, K., & Lombard, M. (2016). Media are social actors. In *Paper presented at the International Communication Assosiaiton Conference 2016*.

Yaghoubzadeh, R., Kramer, M., Pitsch, K., & Kopp, S. (2013). Virtual Agents as Daily Assistants for Elderly or Cognitively Impaired People. In *International Conference on Intelligent Virtual Agents* (pp. 79–91). Springer. http://doi.org/10.1007/978-3-642-40415-3_7

Yang, Z., Ou, C. X., & Zhou, Z. (2017). Investigating the Impact of Recommendation Agents on E-commerce Ecosystem. In *Americas Conference on Information Systems* (pp. 1–5).

Yee, N., Bailenson, J. N., & Rickertsen, K. (2007). A meta-analysis of the impact of the inclusion and realism of human-like faces on user experiences in interfaces. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '07*, 1. http://doi.org/10.1145/1240624.1240626

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology Monograph Supplement*, *9*(2), 1–27. http://doi.org/10.1016/0022-1031(71)90078-3

Zell, E., Aliaga, C., Jarabo, A., Zibrek, K., Gutierrez, D., Mcdonnell, R., & Botsch, M. (2015). To Stylize or not to Stylize ? The Effect of Shape and Material Stylization on the Perception of Computer-Generated Faces. *CM Transactions on Graphics*, *34*(6).

Zibrek, K., & McDonnell, R. (2014). Does render style affect perception of personality in virtual humans? *Proceedings of the ACM Symposium on Applied Perception - SAP '14*, 111–115. http://doi.org/10.1145/2628257.2628270

Ziemke, T. (2001). Disentangling notions of embodiment. In *Workshop on Developmental Embodied Cognition* (pp. 83–88).

VI APPENDIX

Please contact the author if you are interested to obtain the following material:

Study 1: Interviewer guideline
Study 2: Questionnaire and complete stimulus material
Study 3 & 4: Questionnaire and interaction script

The only source of knowledge is experience.

ALBERT EINSTEIN

What we know is a drop, what we don't know is an ocean.

SIR ISSAC NEWTON