The Effect of Embodied Agents within the User Interface

A thesis submitted to The University of Manchester for the degree of
Doctor of Philosophy
In the Faculty of Humanities

2011

Rabia Fatima Khan

Manchester Business School

Table of Contents

List of Tables

List of Figures

Abstract

of thesis submitted by Rabia Fatima Khan for the degree of PhD and entitled

"The Effect of Embodied Agents within the User Interface"

The University of Manchester, January 2011

The thesis explores the trend in recent years by HCI designers to create an interface which is increasingly more anthropomorphic in nature, due to advances in computer graphics and interface technologies. The thesis has researched the effects of one such manifestation of this anthropomorphic trend on the human user, which embodies the human persona, in the form of embodied agents.

The thesis is anchored in the growing area of human-agent interaction studies; and how the agent's appearance in terms of their visual cues (i.e. gender, ethnicity, realism, and attractiveness levels), affects the human user interacting with these artificial entities. The aim of this thesis is to explore how the agents' visual appearance can elicit change in the user's perception and behaviour, in order to improve human-agent design, and the interaction experience for the user.

The thesis extends HCI studies investigating the effect of embodied agents, by highlighting the effect of the attractiveness stereotype which can elicit various impressions, stereotypes and behavioural changes within the human user. The thesis results demonstrate that attractive agents were perceived and evaluated more positively, as well being more persuasive than the unattractive agents. Hence, the agents' attractiveness was the main visual cue which played a major role in affecting the participants' opinion and behaviour towards the agents.

The thesis advances the current understanding of CASA, by providing evidence to suggest that although users may respond socially to agents; this human-agent experience is not always equal to human-human experience. The thesis concludes by stating that the CASA methodology and Media Equation require some modification and needs to be adapted when applied to human-agent interaction, and especially within the interaction-based context.

Declaration

No portion of the work referred to in the dissertation has been submitted support of an application for another degree or qualification of this or any other university or other institute of learning.

(Rabia Fatima KHAN)

Copyright Statement

I. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyrights or related rights in it (the "Copyright") and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

II. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made **only** in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

III. The ownership of certain Copyright, patents, designs, trade marks and other intellectual property (the "Intellectual Property Rights") and any reproductions of copyright works in the thesis, for example graphs and tables ("Reproductions"), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property Rights and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

IV. Further information on the conditions under which disclosures, publications and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://www.campus.manchester.ac.uk/medialibrary/policies/intellectual-property.pdf), in any relevant Thesis restriction declarations deposited in the University Library, The University Library's regulations (see http://www.manchester.ac.uk/library/aboutus/regulations) and in The University's policy on presentation of Theses.

Statement of Contribution

This disclaimer is to state that the research reported in this thesis is primarily the work of the author and was undertaken as part of his doctoral research. Referenced papers of which the student is not the sole author represent the role of the supervisors in the research, to direct the work and enhance the written style of these papers.

The work reported in Chapters 3, 4 and 5 has been published as follows. The content of these papers has been re-interpreted and rewritten in the thesis.

The census study reported in Chapter 3 was published in parts as Khan, R., & De Angeli, A. (2007). Mapping the Demographics of Virtual Humans. *Proceedings of British HCI* (pp.149 – 152). Lancaster, UK: ACM (Winner of the 'Best Student Paper' at British HCI 2007).

The attractiveness stereotype study reported in Chapter 4 and the interaction-based attractiveness stereotype study in Chapter 5 was published as Khan, R., & De Angeli, A. (2009). The attractiveness stereotype in the evaluation of embodied conversational agents. *Proceedings of INTERACT 2009 International Conference on Human-Computer Interaction* (pp. 405-417). Uppsula, Sweden: Springer.

Dedication

To my father and mother, Dr Ahmad Zaman Khan and Mrs. Shahida Zaman Khan

Acknowledgements

I would like to thank Professor Alistair Sutcliffe and Dr. Antonella De Angeli for their invaluable guidance and assistance through this journey. Furthermore, I would like to thank my family (Dr. A. Z. Khan, Mrs. S. Z. Khan, Aisha and Muhammad Khan) for being so patient with me and supporting me from the beginning to the end of my PhD. Plus a huge thanks to all my friends in Liverpool and Manchester for always encouraging and having faith in me; and especially fellow PhD colleagues who kept each other going and always there when you needed someone to talk to: Evi Mansoor, Mahmood Alvi, Joanne Hinds, Nazian Juhamri, Zamhar Ismail, Hamda Shihi. Plus, Abdallah Namoune who has always been so helpful in giving invaluable advice and guidance; and Daniel Morrell the amazing and helpful friend in AV. Thank you all for making this an unforgettable experience of a life-time.

1 Chapter 1: Introduction

This chapter intends to provide an in depth discussion of the problem space within the thesis. It introduces and outlines the research area in which the thesis work is embedded. The main research questions are defined, including a brief summary of the rest of the thesis report.

1.1 Introduction

A growing consensus within the human-computer interaction (HCI) community states that traditional WIMP (windows, icons, mouse and pointer) interfaces need to evolve and display a more adaptive, believable, flexible and human-oriented presence (Ball & Breese, 2000). In doing so, one of the main aims for HCI designers has been to develop a more natural human-computer interface (Holtgraves, Ross, Weywadt, & Han, 2007). A primary approach to address this has been a trend for interface designers to make the interface more anthropomorphic and utilise agents which behave in a manner that is more social, which would previously only have been accredited to humans.

A manifestation of this anthropomorphic trend has been for HCI designers to embody the human persona in the form of embodied agents within the computer interface. Embodied agents are described as being 'synthetic computer generated representations of humans (Nowak & Rauh, 2005), which can to some degree display life-like behaviour such as speech, emotions and gestures within the interface' (Cassell, Sullivan, Prevost, & Churchill, 2000). Whilst, anthropomorphism is defined as 'the user's attribution of human traits and characteristics to non-human objects and events' (Nass & Moon, 2000). We redefine this definition to 'anthropomorphism is the user's attribution of human traits and characteristics to embodied agents within the user interface.'

Embodied agents can be defined as: 'Interfaces based on the anthropomorphic metaphor, which look human-like and mimic a face-to-face interaction style.' Examples of various embodied agents used in HCI research are: embodied conversational agents (ECA's), relational agents (RA's), pedagogical agents (PA's), and chat-bot agents.

There are a numerous number of potential uses for embodied agents in our every day lives as well as in a wide arrange of application areas: in e-commerce, entertainment, online negotiations, as virtual tutors and trainers for children or adults, virtual customer relations managers, online sales assistants, virtual friends, online health advisors, in entertainment/games, virtual mortgage advisors for a large on-line bank, or as helpers on regular desktop applications (i.e. helping users file away their photographs).

An additional category of an embodied agent type is called a virtual avatar or avatar. The definition for a virtual avatar is identical to that of an embodied agent; the main difference being that a human user is behind and in control of the avatar. Whereas, an

embodied agent functions via a pre-programmed piece of software with fixed physical characteristics, gestures and speech; which cannot be modified in any way by the user interacting with it. Examples of avatars are in online environments such as 'second life', or in online multi-user role playing games such as 'World of Warcraft;' where the user can either create or select an avatar to represent them within the virtual world. Therefore, allowing the user to have greater control over the avatar; such as the movement, speech, appearance, clothing and physical gestures.

1.2 Research Motivation

The prime motivation of this research is four fold. The first is due to the lack of understanding and research amongst the HCI community as to how the visual appearance of the agent affects the user's perception and behaviour towards the agent. Secondly, no census has previously been undertaken to highlight the prevalent design trends in terms of the demographic (i.e. age, gender and ethnicity) and physical attributes (i.e. facial attractiveness and realism/anthropomorphism levels) pertaining to embodied agents, which would enable researchers to predict biases as well as difficulties in human-agent interaction.

Thirdly, the research is aimed at investigating the CASA model and whether it can be successfully applied to the human-agent context within this thesis, and if any adaptations would be required to this methodology. Fourthly, from a design perspective it is envisaged that the results of this research aim to develop a set of design guidelines to inform and improve the efficiency and credibility of both embodied agent design and human-agent interaction.

1.2.1 The Embodiment Revolution

The initial interest to develop a more anthropomorphic interface was generated by early research comparing the effect of a text based and talking face interface displays on the user (Walker, Sproull, & Subramani, 1994; Sproull, Subramani, Kiesler, Walker, & Waters, 1996). The findings suggested that the users found the talking face more engaging, whereby they wrote more comments and made fewer mistakes when interacting with it. Furthermore, the users responded more positively towards the talking face, which evinced a positive change in their behaviour as well as attributing personality traits towards the talking face display (Sproull et al., 1996).

These studies laid the foundations for a new wave of research, within the HCI design community, investigating how anthropomorphic attributes elicited a response from the user. As current research in graphics and software technology advanced, there was an increased use of embodied agents (Dehn & Van Mulken, 2000).

The long term goal for interface agent designers and practitioners has been to develop embodied agents that can aid users in a variety of domains including e-commerce, education, tourism, online help, virtual worlds and entertainment. Hence, their roles can vary from online advisors and tutors to virtual friends (Cassell et al., 2000). Interface designers have established various categories of embodied agents: embodied conversational agents (ECA's), relational agents (RA's), pedagogical agents (PA's), chatterbot or chat-bot agents and avatars. Conversation is increasingly becoming an integral part of agent design; as it defines humanness in terms of human face-to-face interaction (Cassell et al., 2000). Designers anticipate that once such technology becomes available, spoken communication will be the preferred means of interaction with the computer by the user (Ball & Breese, 2000).

1.2.2 The Impact of Visual Cues and Stereotypes in Social Interaction

Studies in social psychology demonstrate how people react to other individuals based on their subconscious assumptions on the target person's facial features, i.e. visual cues such as the individual's age, gender, ethnicity and attractiveness level (De Meuse, 1987; Gulz & Haake, 2006). These visual cues are exploited by humans as they are the main focal points which the observer uses to form impressions as well as attributing various personality traits to the target individual. Therefore, allowing the observer to form expectations to guide their interaction with others (Haake & Gulz, 2008).

The human face is the main vehicle used in human-human communication with others, with profound effects on the observer's impressions and behaviour towards the target person (Dipaola, 2002; Haake & Gulz, 2008). Social interaction studies have shown that age, gender, ethnicity, and physical attractiveness encode information about people (Taylor, Fiske, Etcoff, & Ruderman, 1978). The perceiver analyses visual cues of a target person to determine personal categories, a process referred to as stereotyping (Fiske & Taylor, 1991). These visual stereotypes are culturally shared socio-cognitive schema which summarise our beliefs about others, and act as a default setting within the social perception process (Haake & Gulz, 2008).

1.2.3 Attractiveness and the Attractiveness Stereotype

This thesis investigates the role of attraction, and how the embodied agent's attractiveness level elicits the attractiveness stereotype. Social psychology has explored the attractiveness stereotype in great detail; by highlighting how human observers perceive attractive individuals more positively (in terms of personality traits) than unattractive individuals (Dion, Berscheid, & Walster., 1972; Eagly, Ashmore, Makhijani, & Longo, 1991; Feingold, 1992; Fiske, 1993; Langlois, Kalakanis, Rubenstein, Larson, Hallam, & Smott, 2000).

In a seminal study, from the field of social interaction, by Dion et al. (1972); provide evidence for the presence of the attractiveness stereotype which was elicited from static images of college students. The subjects rated images of fellow college students on various personality traits including future outcomes in terms of married life and employability. They found that attractive individuals are assumed to have more socially desirable personalities, as well as happier and more successful lives than unattractive counterparts. Additionally, A number of similar studies investigating the attractiveness stereotype also suggest that the more attractive a child or adult is perceived to be then the more positive personality traits are attributed to that individual by both strangers and family/friends (Eagly, et al., 1991; Feingold, 1992; Fiske, 1993; Langlois et al., 2000).

This contrasts with agent only studies within the HCI community addressing the effect of embodied agent attractiveness; some of which have highlighted the positive attitude of the user towards attractive online agents and avatars within the interface (Holzwarth, Janiszewski, & Neumann, 2006; Nowak & Rauh, 2008; Vasalou, Joinson, Banziger, Goldie, & Pitt, 2008; Vasalou & Joinson, 2009).

1.2.4 Embodied Agent Research

Inclusion of embodied agents within the interface is increasing; however, little empirical work has investigated the effect of these representations on the user, and has yet to come up with a whole set of conclusive results. This is perhaps due to researchers using a wide variety of methods to evaluate different types of agents which carry out very different types of tasks (Dehn & Van Mulken, 2000; Catrambone, Stasko, & Xiao, 2004; Berry, Butler & de Rosis, 2005; Cowell & Stanney, 2005).

Researchers have investigated the effects of certain visual stereotypes, i.e. age, gender and ethnicity (Baylor & Kim, 2003; Baylor, Shen, & Huang, 2003; Baylor & Kim,

2004; Baylor & Plant, 2004), and physical attractiveness (Holzwarth, Janiszewski, & Neumann, 2006; Nowak & Rauh, 2008; Vasalou et al., 2008; Vasalou & Joinson, 2009) of an agent on the user. They found that in general, the participants preferred to interact with agents that were from a similar background and ethnicity to themselves. Whilst, female agents were regarded as being more agreeable and friendly than male agents. Additionally, the more anthropomorphic agents were perceived by participants as more attractive, trustworthy and credible; whereas users in second life had a general tendency to create more attractive avatars than themselves, especially in dating scenarios where their avatar represented them during interactions with other users.

Furthermore, a growing number of studies (Holtgraves, Ross, Weywadt, & Han, 2007; De Angeli & Brahnam, 2008; Veletsianos, Scharber, & Doering, 2008) have also investigated how users perceive, interact, as well as give abuse to chat-bot agents whilst conversing with them; but rarely have such studies considered how the physical attributes of an agent (i.e. the sex or attractiveness levels) could impact and affect the user's behavioural response.

The developing area 'captology' or 'persuasive computing' (Fogg, 2003; Holzwarth et al., 2006; Zanbaka, Goolkasian, & Hodges, 2006) suggests that users are persuaded by computers. Additionally, social psychology studies such as that by Chaiken (1979) suggest that attractive individuals are more persuasive in changing the behaviour of others than unattractive individuals. These discussions on embodied agents highlight the need for further empirical work to examine how such visual cues affect user perception and behaviour towards embodied agents.

A growing number of HCI interface researchers and designers (Walker et al., 1994; Sproull et al., 1996; Nass, Isbister, & Lee, 2000; Baylor & Kim, 2003; Fogg, 2003; Baylor & Kim, 2004; Zanbaka et al., 2006; Haake & Gulz, 2008) are drawing on the psychological literature in human-human interaction which elaborates on the visual cues that elicit various perceptions, stereotypes and behavioural changes between humans. The knowledge obtained from such literature is being used to help predict and explain human-agent interaction. One such example is the use of the CASA model (computers are social actors), which posits that people respond to computers much the same way as they do to other people (Nass, Moon, Fogg, Reeves, & Dryer, 1995). The CASA model is a result of empirical investigations and findings by Reeves & Nass (1996) in their seminal work named 'The Media Equation.' They present two main overriding

conclusions in this work: ‘people’s responses to media are fundamentally social and natural,’ and ‘media experiences equal human experiences.’

The basic premise of the Media Equation is that humans subconsciously interact with mediated entities such as computers, televisions and new media as real people and places. Furthermore, they point out that users treat computer personalities as psychologically real and will respond to a computer endowed with personality in the same way as they would towards a human (Figure 1).

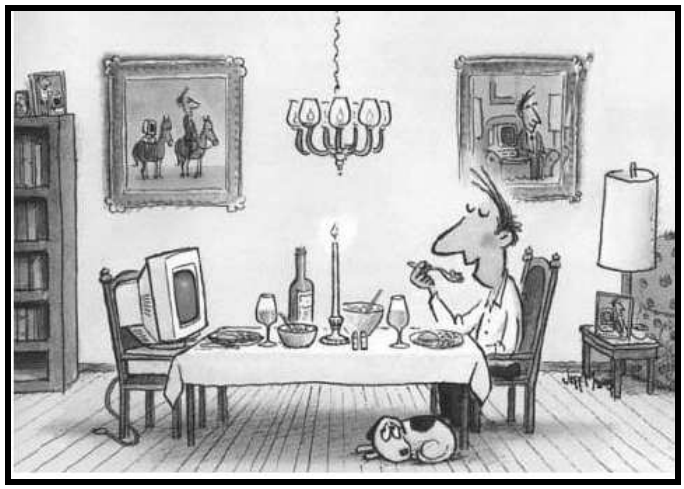


Figure 1. **The ‘Media Equation’ by Reeves & Nass (1996).**

Earlier studies by interface practitioners (Sproull et al., 1996; Nass, Moon, & Green, 1997) have pointed out that even subtle social cues, such as a taped human voice emanating from a computer, could elicit a social response from the user; as well as a change in the user’s perception and even behaviour towards the computer.

However, not all HCI practitioners are in favour of the Media Equation or CASA model, and do not agree with the belief that people will always attribute human-like characteristics to technology. For example, Morkes, Kernal & Nass (1999) conducted two experiments where one group of participants worked with another individual linked by computer mediated communication (CMC), whilst the remaining participants in the HCI condition were told they were going to interact with another computer in a different room. The CMC participants responded more sociably towards the other individual, and rated the other individual as more likeable with the tendency to cooperate with them more; whereas the HCI subjects behaved less sociably and felt less similar towards this interaction partner, as well as spending less time on the task at hand. This is further supported by Couper, Tourangeau, & Steiger (2001), who conducted a web survey investigating how anthropomorphic features would impact the

participant's response. Their results suggested that anthropomorphic characteristics such as the voice reduced or eliminated social desirability effects from the user's perspective; and further concluded that they found little support for the Media Equation within this study.

De Angeli, Johnson, & Coventry (2001) point out that HCI practitioners should not completely 'adopt' social psychology theories and directly apply them to HCI scenarios; but they should practice caution by 'adapting' these theories within a new HCI context. This argument is valid especially in light of studies demonstrating negative reactions from the user; e.g. where users have been abusive and rude whilst chatting to chat-bot agents (De Angeli, Brahnam, & Wallis, 2005; De Angeli & Brahnam, 2008; Veletsianos et al., 2008).

1.3 Research Questions

The research employs a variety of methods: online survey studies and controlled lab based experiments, with analysis of agents present in commercial applications, and research tools. The thesis addresses the following research questions:

- *What design trends are prevalent amongst the embodied agent population?*

  Previous work has not investigated the design trends in terms of the demographic and physical attributes such as gender, age, ethnicity, attractiveness and realism levels of embodied agents that have been developed for research. Only by conducting a detailed agent census can we then make observations on any significant patterns amongst the embodied agent population. This will then allow HCI practitioners to understand design trends and potential biases and difficulties which can manifest during human-agent interaction.

- *How does the attractiveness of an agent positively affect the participant's perception and behaviour towards it?*

  The thesis aims to investigate how an embodied agent's visual cues (i.e. the gender, ethnicity, anthropomorphic and attractiveness levels) could elicit any visual stereotypes (i.e. ethnic, gender or attractiveness) which will affect the user's perception and behaviour towards the agent. Hence, the main focus within the thesis is on the attractive stereotype, which will be tested within two contexts: stranger-based (where no interaction takes place between the

participant and embodied agent) and interaction-based (where the participants interact with the embodied agent; i.e. via chat sessions, or use the agent to help them solve a task, i.e. the desert survival task), between the human participant observers and embodied agent targets. The findings will add to the present discussion amongst HCI practitioners, on whether the benefits of making the user interface more anthropomorphic by using such embodied anomalies far outweigh the disadvantages, or visa versa. Additionally, the results of the empirical studies are aimed at further highlighting and extending the CASA paradigm and the Media Equation, and whether it can be directly applied to human-agent interaction; or if this needs to be modified and adapted in some way within this context.

1.4 The Report Outline

This thesis is structured into 7 Chapters (Figure 3). Chapter 2 reviews literature in both social psychology and current research developments in HCI on embodied agents. Chapters 3 to 6 report the empirical investigations conducted within the thesis. Chapter 7 concludes with a discussion of all the findings, and hence the implications for HCI research and design; including design guidelines and limitations in the thesis, as well as recommendations for further research.

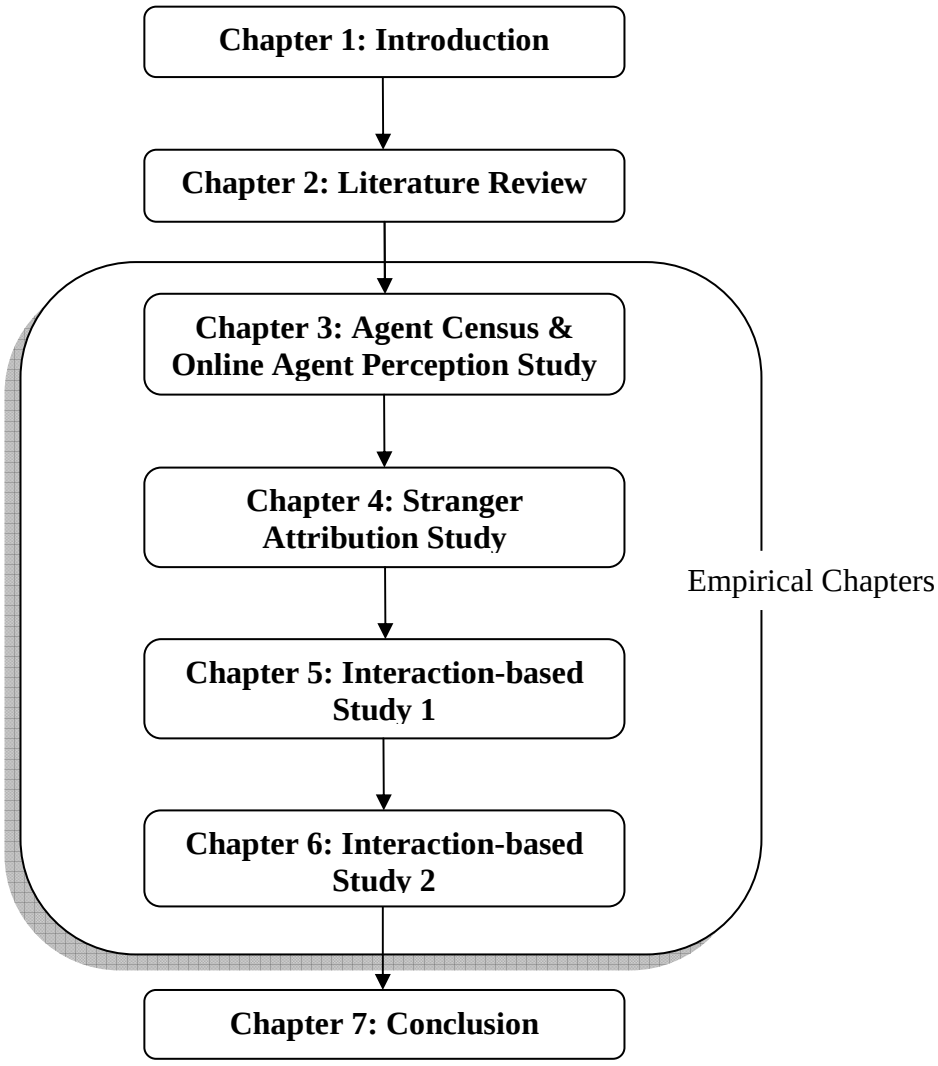


Figure 2. Structure of Thesis Report.

Chapter 2 presents the current debate on anthropomorphism and the impact of embodiment in the form of embodied agents on the user. This chapter reviews two sets of literature. The first is based on social psychology highlighting the visual cues (demographic and physical) which play a vital role in face to face communication; including the visual stereotypes. This interaction is divided into that which is either stranger-based (Dion et al., 1972; Eagly et al., 1991; Feingold, 1992) or interaction-based (Langlois et al., 2000). Secondly, the chapter reviews literature within the HCI community on the increasing use of embodied agents and how they elicit various stereotypes and behavioural changes during human-agent interaction. The chapter pays particular attention to the role of attractiveness in the social domain and its positive influence on the observer's perception and behaviour towards attractive targets.

Chapter 3 presents the first two investigations. The first is a census study on 188 embodied agents; whilst the second is an empirical online agent perception survey

examining how participants perceived embodied agent attractiveness and realism. Both of these studies illustrate the prevalent designer bias and stereotypes which favour the development of younger looking agents from an ethnically white background; as well as attractive and realistic looking female embodied agents. The online survey highlighted the power of agent embodiment, whereby their mere physical presence could influence user perception. Whilst, further revealing how young female agents were perceived by the participants as being more attractive and realistic looking than male agents. Comments made by the participants at the end of the survey were also reviewed, giving an in depth acumen into the way in which these agents were perceived by them; allowing other HCI designers to take note of their comments and suggestions to advance embodied agent design and hence human-agent interaction. Finally, a framework of analysis was devised in respect to the agent attributes, and is proposed for use by other HCI agent designers and practitioners.

Chapter 4 extends the findings from Chapter 3 by replicating Dion et al.'s (1972) 'What is beautiful is good' experiment. Hence, this study explores the effect of the attractiveness stereotype on first impressions between embodied agents and the participant, within the stranger-based context. Approximately 30 Manchester University students evaluated two sets of three static images of female embodied agents, varying in terms of their attractiveness levels (unattractive, average and attractive). The results indicated that, on first impressions, the embodied agents did elicit the attractiveness stereotype from the participant; thus highlighting the relationship between an increase in agent attractiveness and a more positive agent evaluation. The debrief interviews between the author and participants were also conducted; leading to a greater understanding of how these participants perceived the embodied agents, and ways in which to improve their design and interaction with them.

Chapter 5 presents the next empirical investigation which is composed of two main sections, looking into the effects of the attractiveness stereotype in human-embodied agent interaction within the interaction-based paradigm. Approximately 48 Manchester University students participated in this experiment. The first part of this study required participants to evaluate the static image of one of four randomly designated female embodied agents called Alex, differing in their attractiveness levels (attractive vs. unattractive) and ethnicity (black vs. white). Participants evaluated the assigned image of the embodied agent both before and after chatting to them. The results point to a strong effect of the attractiveness stereotype

both before and after this interaction between human and agent. Additionally, a larger proportion of positive personality traits were assigned to the attractive agents than towards the unattractive agents by the participants. The second section investigated the 48 agent-participant conversations, remonstrating significantly greater positive behaviour (i.e. positive inputs) by the participants towards the attractive white and black female chat-bot agents; whilst also pointing to more negative behaviour by the participants towards the unattractive chat-bot agents, which is also known as the flaming effect (Lea, O'Shea, Fung, & Spears, 1992); where users behave in an anti-social way towards computer technology. Furthermore, a coding system was also developed by the author in order to analyse these conversations.

Chapter 6 reports the final empirical investigation into the effects of attractiveness on the persuasiveness levels of the four sets of female embodied agents. Participants were divided randomly into two groups of 15; one group interacted with an attractive female embodied agent and the other with an unattractive female agent. The participants interacted with the agent by initially listening to her read out the desert survival task, before ranking ten items in order of importance. The participants were then given the opportunity to listen to the agent's advice for each of the 10 items, whereby they had the option to re-rank the items if they so wished. The results demonstrated that the attractive embodied agent was evaluated more positively in terms of her advice as well as personality traits. Subsequently, the attractive agent was more persuasive in changing the participant's decision (behaviour), as well improving the performance of the participants than those who interacted with the unattractive agent. This helped explain the reason why the attractive agent's advice was also listened to by a greater number of times than that of the unattractive agent. These findings provide additional evidence highlighting the powerful influence of agent embodiment and physical attractiveness on user perception and behaviour.

Chapter 7 summarises and discusses the findings, implications, as well as the contributions of the thesis. Furthermore, this chapter also points to limitations and proposes future work, as well as design guidelines for HCI agent designers and researchers, by finally then concluding the thesis report.

2 Chapter 2: Literature Review

The chapter reviews literature within the HCI community on the use of anthropomorphic embodied agents. This section reviews social psychology literature highlighting the visual cues (both (demographic and physical) which play a vital role in: perception and impression formation, as well as eliciting visual stereotypes during social interaction. The potential stereotypes and behavioural changes embodied characters elicit in human-agent interaction are discussed in order to obtain a deeper understanding of how users perceive and evaluate them; thus leading to an improvement of their design and communication with the user.

2.1 Introduction

The past decade has seen a considerable advancement in computer graphics and multimedia systems, resulting in computer users witnessing new paradigms in human computer interface technology. One such technology which the HCI community is increasingly focusing its attention on is that of the embodied agents, which have become more synonymous with anthropomorphic design (Berry et al., 2005). A number of HCI practitioners have taken up this quest in making the interface more anthropomorphic with the intention of allowing the interaction between the user and the interface to resemble that of human face-face communication (Cassell et al., 2000).

2.2 The Anthropomorphic Interface

The initial drive towards creating a more anthropomorphic interface was based on early studies addressing social responses to text based and talking face interfaces (Walker et al., 1994; Sproull et al., 1996). Walker et al. (1994) demonstrated that users responded more positively, spent more time, made fewer mistakes, and wrote more comments to a talking face than to a plain text display on the screen. Furthermore, the talking face was perceived as being more engaging than the text based interface. Sproull et al. (1996) pointed out that users enjoyed the experience when they interacted with an expressive face, and responded in a more positive manner towards the face display than towards the text based interface.

These studies illustrated how the embodiment of the human face elicited change in the user's behaviour. Additionally, enjoyment was greater when the users interacted with a face display rather than the plain text interface display. Sproull et al. (1996) concluded that users found the anthropomorphic interface emotionally more satisfying due to the interaction being regarded as easier, comfortable and more natural to use. Hence, such interfaces have the ability to make a computer more entertaining, engaging, approachable and understandable to the user, and make the user feel more relaxed with the computer (Catrambone et al., 2002).

However, Shneiderman (Shneiderman & Maes, 1997), argued that adding anthropomorphic qualities to an interface not only undermines the responsibility of the user, but raises their expectations which can lead to disappointment if the system fails to live up to them. Furthermore, opponents of the anthropomorphic interface regard them

as annoying distractions which are misleading for both the designer and user; increase user anxiety, undermine user responsibility, reduce user control, as well as destroying a user's sense of accomplishment (Catrambone et al., 2002; Shneiderman & Maes, 1997).

Subsequently, designers have developed such anthropomorphic interfaces which embody the human persona within the interface. This embodiment may vary in levels of realism, but appear human-like in the form of either a full body representation or a facial display. The term 'embodied agent' is used to refer to such anthropomorphic human-like instances within this thesis.

The main advocates of embodied agents within the HCI community state that by adding agents to the interface not only makes the computer system more human-like, but also allows for the interaction between the user and interface to run more smoothly (Berry et al., 2005). Furthermore, Catrambone et al. (2002) report that embodied agents are perceived as being more entertaining, engaging, understandable, and approachable by the user. Koda & Maes (1996) point out that a user will pay more attention to an embodied agent face within the interface and rate it more positively; as these agents allow the user to become more engaged and tentative towards it.

Embodied agents have also been reported to reduce user frustration; where female agents are more effective than male agents (Hone, 2006). Moreno & Flowerday (2006) conclude that students preferred to have a social visual representation in a computer-based learning scenario, in the form of an embodied agent, rather than a text based one. Agents have great potential in many domains, and their deployment could benefit users of all age groups; such as in health and behaviour change domains (i.e. helping people with their dieting or to stop smoking), counselling, coaching, therapy, in education, online shopping or as kiosk assistants and virtual receptionists (Bickmore & Picard, 2005).

2.2.1 The Impact of using Realistic Looking Agents

The use of more realistic embodied agents is justified by studies conducted by Koda and Maes (1996), and Baylor & Kim (2004). Koda and Maes (1996) conducted an experiment to see the affects of applying a face and facial expressions onto an interface indicating that users perceived a realistic looking face as being more likable, engaging and intelligent. Baylor & Kim (2004) pointed out that more realistic images of pedagogical agents had a greater impact on the transfer of learning on students.

Similar results are suggested by Luo, McGoldrick, Beatty, & Keeling (2006), where users perceived human-like characters to be more trustworthy, appropriate, and likeable than cartoon-like characters. Furthermore, male human-like characters in general made the web site (bank card details) they appeared on seem more trustworthy and pleasant to the user.

2.2.2 The Impact of Using Conversation

Cassell (2000) stated 'it is conversation which defines humanness and human interaction.' This has lead HCI designers to develop a number of conversational interface technologies; examples of which are embodied conversational agents (ECA's) and chat-bot agents.

HCI practitioners are incorporating chat-bots within their investigations, to demonstrate how this interface technology affects user perception and behaviour towards them. In a series of experiments by Holtgraves, Ross, Weywadt, & Han (2007); participants conversed with an embodied female bot agent. The studies revealed that after conversing with this chat-bot, the users perceived her to have human-like qualities and personality akin to that of a human.

2.2.3 The Negative Impact of Using Chat-Bot Agent Technology

The downside to human chat-bot interaction is that the chat-bot may be prone to abuse by users, which was not anticipated by the media equation (Veletsianos, Scharber, & Doering, 2008). De Angeli & Brahnam (2008) point out that it is rare for studies using the CASA paradigm to investigate these negative user responses and behaviour. Furthermore, their study (De Angeli & Brahnam, 2008) illustrated how users verbally abused the Jabberwacky chat-bot by using expletives, as well as being sexually explicit towards it. Similar results were presented by De Angeli, Graham, Johnson, & Coventry (2001) where users made unpleasant remarks towards the ALICE chat-bot which were not only rude, but also inferred some sort of stupidity on behalf of the chat-bot, as well as treating the chat-bot as though it was subservient to humans. The abuse of chat-bots is an area HCI researchers need to investigate further.

Veletsianos et al. (2008) conducted an experiment whereby a group of school children between 14-15 years of age interacted with an AI (artificial intelligence) chat-bot called Joan by chatting to her. Joan was embodied as an attractive blonde female character

who was also an expert in geography. Her role was to present and answer student's questions on an online assignment. Analysis of these conversations between Joan and the students revealed that the chat-bot suffered high levels of abuse from the students. The abuse consisted of vulgar language, sexually explicit comments, flirtation, as well references to drugs and violence. Additionally, they state that users may tend to give more abuse when online; as the internet is generally considered as a safe environment which reduces their human inhibitions.

De Angeli & Brahnam (2006) discuss how the physical characteristics of embodied agents can elicit stereotypes based on the embodiment's gender, age and ethnicity. The focus is usually on the sex stereotype and its link with aggression; whereby the main observation was that female embodiment tends to trigger a greater sex stereotype than male embodiment, which leads to verbal aggression that is sexual in origin.

2.2.4 The Negative Impact of Using Embodied Agents

Some researchers consider agents as being rather inappropriate and impractical (Catrambone et al., 2002), whereby once the agent's novelty has worn off then these animated interfaces reduce their appeal in comparison to other user interfaces which don't use embodied agents (Don, Brennan, Laurel, & Shneiderman, 1992; Shneiderman, 1995).

Such critics also regard agent interfaces as being misleading to both the designer and user, by tricking the user into believing the agent is intelligent; therefore removing the feeling of control from the user (Shneiderman & Maes, 1997). Whilst, Luo et al., (2006) report that if a user detects any limitations within the embodied agent; this can soon contribute to users feeling frustrated and irritated with the agent. Furthermore, Shneiderman (Shneiderman & Maes, 1997) claims that users want to feel as if they are in control of the interface; and by allowing an agent to complete certain tasks will somehow instil a feeling of loss of control and inability by the user to do a good job.

The subsequent section discusses the various categories of embodied agents that have been developed by HCI researchers.

2.2.5 Embodied Conversational Agents

In recent years there has been a growing interest in ECA's. Due to advances in human computer interface technology, there has been a rapid expansion in the use of these ever

increasingly looking human-like embodied agents. One of the main pioneers in ECA research is Justine Cassell who defines an ECA as ‘an animated anthropomorphic computer character which is able to engage a user in real-time by emulating face-to-face conversation through the use of a facial display, head motion, gaze behaviour, body posture, hand gesture, speech intonation, as well as speech content’ (Cassell et al., 2000). Hence, these agents can take on a number of different roles, for example, as an assistant, tutor, information provider, or as customer service agents.

Figure 3. REA, the virtual Real Estate Agent (Cassell & Miller, in press).

A growing number of animated agents that converse with human users have been developed by researchers, such as REA the real estate agent and Steve (Soar Training Expert for Virtual Environments). REA, shown in Figure 3, is an ECA who has the ability to interview potential home buyers and show them around virtual houses for sale. She was developed by the MIT Media Lab with the ability to conduct natural face-to-face conversations with users; as well as utilising her human-like body to perform hand gestures and various body postures and facial displays such as eye gaze (Bickmore & Cassell, 2001).

Whilst Steve (shown in Figure 4) is used to teach students how to operate and maintain gas turbine engines aboard naval ships. He can teach in both individual tasks and team task scenarios and has the ability to demonstrate physical tasks, such as the operation and repair of equipment onboard (Rickel & Johnson, 2000).

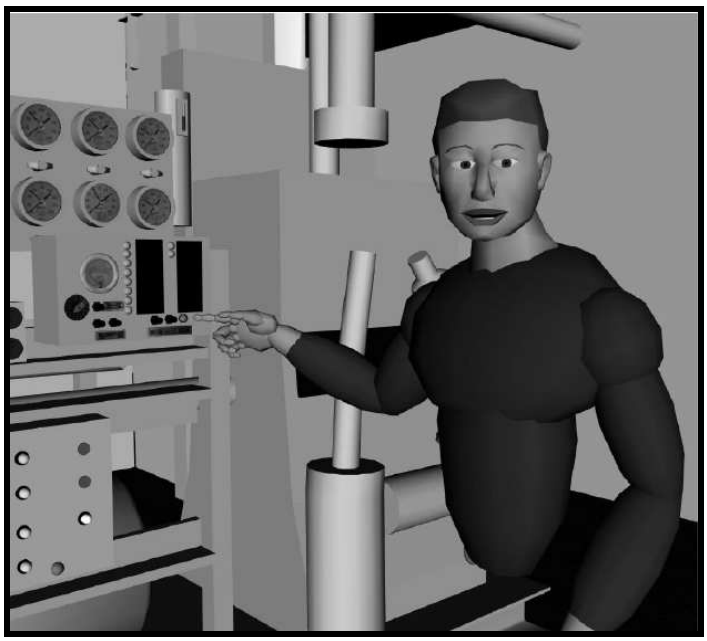


Figure 4. Steve describing a power light (Rickel & Johnson, 2000).

Both Rea and Steve provide at least 3 functions: acknowledgment of the user's presence, feedback, turn-taking, as well as some small talk with the user. ECA's were initially developed for research purposes, and now an increasing number are being used for commercial use. Examples of commercial ECA's are Ana Nova the online newsreader, and Anna the friendly online IKEA sales assistant. Cassell et al., (2000) conclude that 'the primary goal of ECA research is to produce an intelligent agent which is in some sense complete, i.e. capable of certain social behaviours enabled by an input-recognition system, as well as possessing some model of personality and emotion.'

2.2.6 Relational Agents

Timothy Bickmore is one of the leading researchers on RA's and defines them as 'computational artefacts designed to build and maintain long-term social-emotional relationships with users' (Bickmore, Caruso, Clough-Gorr & Heeren, 2005). Relational agents differ from other embodied agents in that the user interacts with the agent over a certain period of time, which in due course develops the user's trust and likeability towards them. This long term relationship between the RA and user can also become a more potent way of persuading the user to alter their behaviour or even change misinformed beliefs; i.e. agents giving informative advice in terms of drinking habits, or how to eat more healthily (Bickmore & Picard, 2005).

Relational agents are primarily designed to remember past events or history, as well as manage any future expectations in their interactions with users (Bickmore et al., 2005). Their uses are currently being researched in the subsequent areas of counselling, coaching, psychotherapy and healthcare.

An example of a RA is Laura (Figure 5) the virtual exercise advisor, which has been developed in order to increase physical activity in overweight adults, due to the fact that over two-thirds of Americans are currently overweight or obese (Bickmore et al. 2005). The software named the FitTrack system was part of a longitudinal study entailing Laura to interact with participants on a daily basis, by chatting to them about their fitness program and physical activities; utilising both hand and facial gestures during the interaction. At the end of the month most participants who used the FitTrack system expressed their desire to continue using it; as they had built a friendly relationship with Laura, highlighting her successful role as an advisor. Whilst, the control group in this study seemed to get bored earlier on in their fitness program, and exercised less regularly than participants who interacted with Laura.

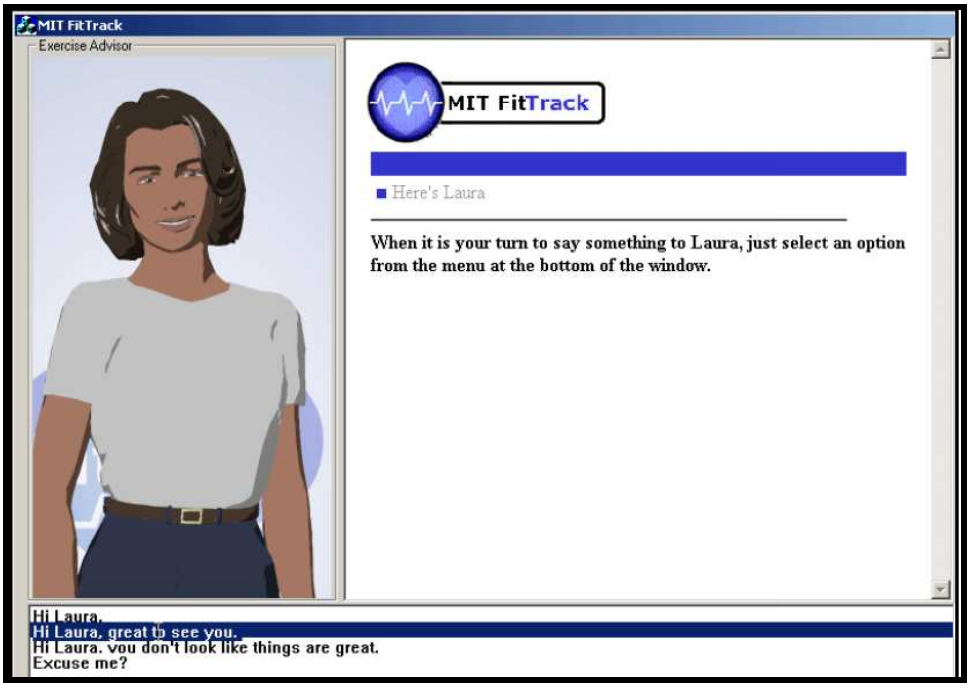


Figure 5. Laura the exercise advisor (Bickmore & Picard, 2005).

As the use of embodied agents is becoming more ubiquitous, the presence of RA's on mobile devices would be a very powerful combination for building relationships and eliciting behavioural change in the user. Hence, RA's such as Laura the exercise advisor could be deployed onto a PDA system or even a mobile phone, which means the user is

not restricted to using her or other RA's exclusively just at home (Bickmore & Picard, 2005).

2.2.7 Pedagogical Agents

Baylor & Kim (2003) define PA's as 'animated life-like characters designed to facilitate learning in computer-mediated learning environments.' They are an extension of an intelligent tutoring system providing students with a tutor through the use of artificial intelligence. An example of one such PA is 'Baldi,' a classroom language training tutor for children with hearing loss (Cassell et al., 2000).

Baylor and colleagues investigated the effect of the PA's visual appearance on user perception and behaviour; and have highlighted the benefits of using PA's by stating how helpful, credible, and entertaining students have found them. Visual cues such as the PA's age, ethnicity, realism and gender have also been shown to influence the user's approach to a certain subject matter; by influencing the student's transfer of learning (Baylor & Kim, 2003; Baylor et al., 2003; Baylor & Kim, 2004). Further work on pedagogical agents is discussed later in section 2.6.

2.2.8 Chat-Bot Agents

A chat-bot is a chat robot, or an AI (artificial intelligence) agent which has the ability to use conversation as a means of communicating with the user in real time (Holtgraves et al., 2007). In general, bots are designed to elicit various anthropomorphic attributions by the user; which include personality and affect, as well as emotional and social intelligence (De Angeli, Graham, Johnson & Coventry, 2001). The advantage of such conversational interfaces is that it appears to be a natural dialog style where the user does not need to learn complex command structures and functionality (Catrambone et al., 2002).

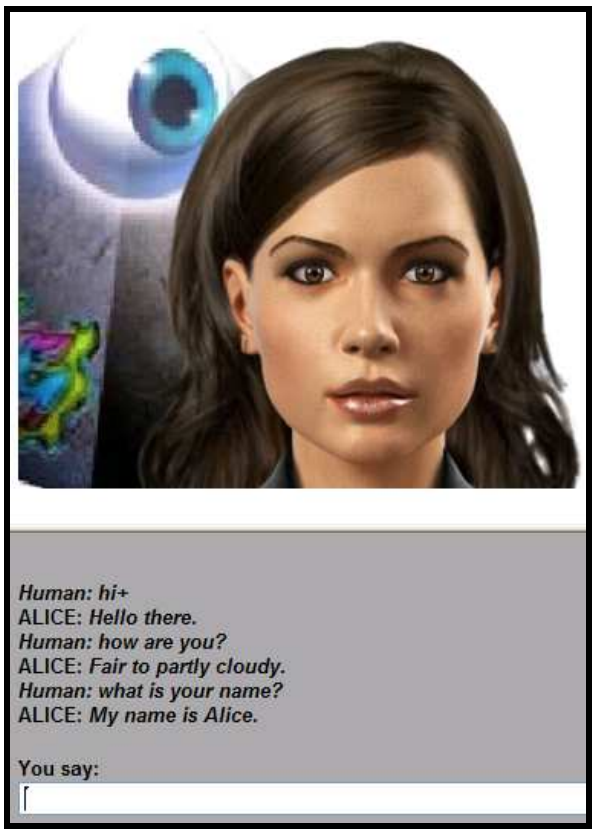


Figure 6. Embodied version of the ALICE chat-bot.

The user communicates by typing their input into a dialog box, whereby the chat-bot will reply either verbally or via readable text on screen. An example of an embodied chat-bot is ALICE (Artificial Linguistic Internet Computer Entity), represented in Figure 6, who was developed by the ALICE AI Foundation and hosted by Pandorabots (2008). She also utilises speech to communicate with the user. Figure 7 illustrates the original disembodied version of ALICE (2008); whereby ALICE uses a text based communication system which is displayed on the screen.

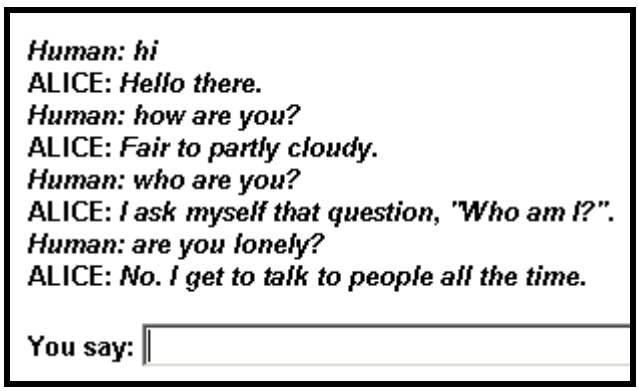


Figure 7. Disembodied version of the ALICE chat-bot.

ALICE was originally created by Richard Wallace in 1995 for the purpose of entertaining individuals. In 2000 and 2004, ALICE won the Loebner prize (an

instantiation of the Turing Test), which is a prize for AI technologies such as chat-bots with the most human-like personality. The main reason behind Alice's personality and behaviour is due to AIML (Artificial Intelligence Markup Language) coding. AIML is an XML-compliant language which allows the developer to modify the chat-bot's personality and personal details; i.e. name, age, gender, likes and dislikes.

Due to the development of e-commerce there has been a renewed interest in the development of chat-bots and an improvement of their technology (Holtgraves et al., 2007); especially in terms of a chat-bot having some sort of embodiment. This has lead the ALICE AI Foundation to collaborate with an online company called Sitepal (www.sitepal.com). Sitepal specialises in creating embodied agents and AI chat-bots called virtual characters that can be personally tailored and designed by the client. The client can select the gender, age range, ethnicity, clothing, hair and makeup style and clothing for their virtual character. These characters can be placed onto a website and welcome users as well as inform visitors about the site and answer any questions they may have regarding the company site (Figure 8).

Sitepal allows the client to create these AI enabled chat-bots using AIMC (Artificial Intelligence Management Center) technology. AIMC is a powerful tool which allows the client to teach the new chat-bot how to answer various questions which could be asked of the chat-bot by potential users, as well as editing personal details by using AIML coding.

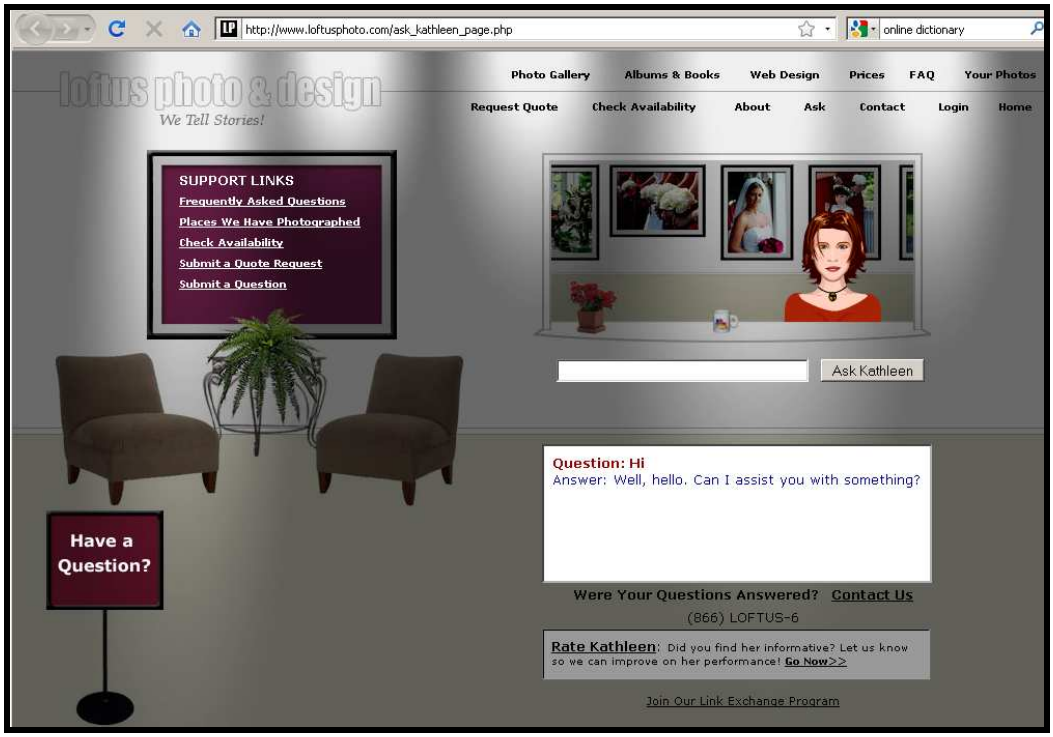


Figure 8. The Loftus home page.

Figure 8 represents an AI chat-bot called Kathleen which is used to assist potential customers, for a photography company called Loftus (2008), with product related questions on the website. Whilst, Figure 9 is a screen shot of a website called Active History (2008) which offers award-winning methods for teaching history in the classroom. Students can click on one of the speaking characters (chat-bots) such as William the Conqueror, and then ask questions regarding his life; whilst listening to the character give its answers.

Figure 9. The Active History home page.

The following sections (2.3 - 2.6) incorporate and discuss two main areas of research: The first draws on social psychology theories on how people perceive and behave towards other individuals based on their visual appearance in social interactions; and the second investigates literature within the HCI spectrum, on the impact of embodied agents' visual appearance on the user.

2.3 The Impact of Embodiment

Although embodied agents are populating the digital society in increasing numbers, the HCI literature based on how this embodiment impacts user perception and behaviour is sparse (Nowak & Rauh, 2005). Furthermore, the results of such studies have not provided consistent and conclusive results (Cassell, 2000; Dehn & Van Mulken, 2000; Catrambone et al., 2002).

In human-human social interaction the visual appearance of an individual's face is known to play a central role, with profound effects on our impressions as well as our behaviour towards them (Gulz & Haake, 2006). Similarly, the embodiment of the interface which incorporates the human-like face is a strong social cue which transmits

its own unique message (i.e. age, gender, ethnicity, attractiveness) that can be interpreted by the observer in order to form impressions as well as elicit behavioural change within the user (Sproull et al., 1996); thus leading the user to perceive the interface as being more social (Nass, Steuer & Tauber, 1994).

It is more than likely that an embodied agent's face will play a central role in interaction with the user (Gulz & Haake, 2006); in the same way the human face is important for personal communication (Dipaola, 2002). Humans use of two main methods of communication: verbal and non-verbal cues. In order to categorise these cues, De Meuse (1987) designed a conceptual framework which represented how a perceiver would receive and interpret non-verbal cues during face-to-face interaction. Subsequently, Cowell and Stanney (2005) adapted and improved the original taxonomy by making it more descriptive, whereby they detailed all the various types of cues involved in this interaction, as shown in Figure 10.

De Meuse's Taxonomy (Figure 10) illustrates how non-verbal cues are divided into cues which are either behavioural, or cues that are non-behavioural in origin. Non-verbal cues can be defined as "all communication except that which is coded in words" (De Meuse, 1987). Behavioural cues can be classified into five variables: facial expression, eye contact, posture, gesture and paralanguage. Some of these behaviours are more controllable by an individual, such as nodding of the head or smiling.

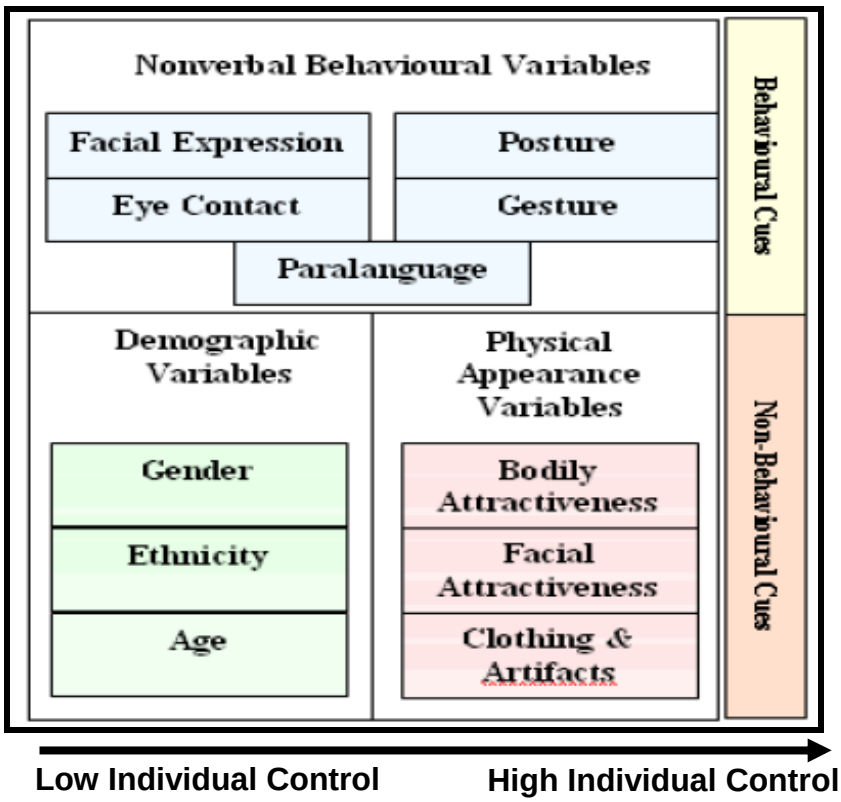


Figure 10. De Meuse's Taxonomy (Adapted and improved by Cowell & Stanney, 2005).

Non-behavioural cues are grouped into: demographic variables and physical appearance variables. Demographic cues consist of gender, ethnicity and age which are not under an individual's control. Physical appearance cues are classified into three groups: bodily attractiveness, facial attractiveness, clothing and artifacts; which individuals have greater control over. For example, the type of clothing, hair style and make-up can all modify ones physical appearance and therefore effect the perceiver's perception and attitude (De Meuse, 1987).

De Meuse's framework evolved from a review of 46 empirical studies looking at the effects of three groups of non-verbal cues (i.e. non-verbal behaviour, demographic and physical appearance) on performance appraisal. De Meuse's (1987) review suggested that the effects of non-verbal cues on personal perception of individuals were actually quite significant, highlighting the fact that both demographic and physical appearance indicators play a large role in affecting interpersonal perceptions and behaviour.

2.4 Forming Impressions

First impressions influence social interaction, as people form impressions from visually prominent facial features (Fiske & Taylor, 1991). The perceiver will assign meaning to these cues, as well as forming impressions based on a combination of these available indicators (Taylor et al., 1978; Sproull et al., 1996). These cues play an instrumental role in social cognition, by transmitting information and communicating systematic images to the perceiver (Fiske, 1993).

2.4.1 Stereotypes within the Social Domain

When individuals form unconscious impressions about groups of people, this is known as stereotyping (Taylor et al., 1978; Fiske & Taylor, 1991). The process of categorising people is usually based on widely shared and simplified generalisations about individuals as members of a certain social group. Individuals can distinguish a target person's race, gender and age within milliseconds of the first meeting, and categorise an individual as being either an in-group or an out-group member (Fiske, 2000). An in-group member may belong to the same ethnic group, gender or age group as other in-group members, whilst out-group members would belong to another ethnic group, gender, and age group (Fiske & Taylor, 1991).

Individuals identified as part of the in-group are perceived as having the same values as other members, as well as being perceived as more socially attractive, trustworthy, competent, and favoured more than out-group individuals (Nass et al., 2000). However, not all stereotypes are based on a negative bias about a certain group of people. For example, the French are thought to have the best wine and fashion sense, whilst African-Americans are perceived to be better at basketball and more musically talented than fellow White Americans (Aronson, Wilson, & Akert, 2005).

In summary, the close connection between everyday human-human interaction and the anthropomorphised human-agent interaction brings with it a whole baggage of positive or negative stereotypes, which help frame our expectations and are used to build common references within conversations about other individuals (Haake & Gulz, 2008).

2.4.2 Effect of Gender Stereotypes

Williams & Bennet (1975) define gender stereotypes as 'the constellation of psychological traits generally attributed to men and women respectively.' Gender

stereotypes can present themselves in various domains, such as in occupational roles and the work place, in advertising, parenting roles, and so on.

Traditionally, society has certain gender-role expectations and perceives men and women to possess particular attributes which fit their role. Hence, females are expected to be more nurturing, expressive, sensitive and friendlier than males. This can then lead to both men and women developing a different set of attitudes and skills due to their experiences within their gender. Females generally learn to become more accommodating and polite in societies where they are less powerful and the chances for them to occupy high status roles are unlikely. A common view of a female is that they are regarded as being more talkative and emotional, whilst males are usually thought of as being rather aggressive or rational. (Eagly & Mladinic, 1989; Aronson et al., 2005).

Gender stereotypes are also quite pervasive in the advertising industry. Studies on television commercials throughout the world have shown that females are usually portrayed as having dependent roles (i.e. they have no position of power and depend on others), in comparison to males who are usually portrayed as having a higher status and being highly independent (Aronson et al., 2005). Interestingly, equal ratings, in terms of job evaluation, have also been assigned to both males and females when the job role being evaluated was less gender specific, for example, an interviewer or an academic (De Meuse, 1987).

However, in recent times it is becoming more evident that the traditional gender stereotypes for males and females is changing. These days, as the number of females in the work force has increased, so have attitudes towards the 'traditional' role of the sexes. To back this claim, Eagly & Mladinic (1989) compared two gender stereotype studies in 1957 and then in 1975; revealing that females were evaluated rather more favourably in the second study in 1975.

2.4.3 Effect of Ethnic Stereotypes

In terms of ethnicity, there is evidence to suggest that individuals assess another's ethnicity in order to determine whether they are from the same in-group (Nass et al., 2000). The media plays a significant role in influencing the viewer's perception of various ethnic groups by. This is supported by Ford's (1997) claim that young children are more likely to believe the portrayal of African Americans as being 'true to life' when highly exposed to African Americans on television. Furthermore, Ford (1997)

investigated the number of black appearances as well as cross-racial interactions present in on screen prime-time programmes. The results indicated an increase from 8.3% in 1978 to 18% in 1989.

Studies have also highlighted the attitudes by minority groups in the U.S.A., such as the Latino and African American students; who have a sense of being portrayed as being less intelligent and less diligent than fellow European American students in their class (Aronson et al., 2005). Whilst, Negative ethnic stereotypes were demonstrated by a study on nuclear energy. Here, two confederates, one African American and one White American, debate about the pros and cons of nuclear energy. During the debate a racist and derogatory comment was made about the African American debater by a member of the audience, which then activated and raised other negative stereotypes in the remaining audience members who heard this comment. This resulted in the observers to lower their ratings for the African American debater, whilst at the same time being rated as just as skilful as the white debater in the control group where no racist comment was made (Hogg & Vaughan, 2005).

2.4.4 Effect of Age Stereotypes

Ageism is defined as ‘a process of systematic stereotyping of and discrimination against people because they are old, just as racism and sexism accomplish this for skin colour and gender’ (Nelson, 2004). Overall, ageism occurs towards both males and females alike. However, there is a difference in the attitudes towards the elderly in both eastern and western cultures. In cultures where the extended family thrives and is the norm, the older members of the community are considered to be knowledgeable and wise teachers or leaders. In societies and cultures where the nuclear family has replaced this extended family (i.e. UK, USA and Canada), the elderly are regarded as powerless and worthless members of society whose special needs tend to go unattended; whilst the quality of youth is highly regarded and the young are thought of as an asset to society (Nelson, 2004; Hogg & Vaughan, 2005).

2.5 Attractiveness in the Social Domain

Attractiveness can be defined as a ‘quality which arouses the interest and admiration of onlookers’ (Prestia, Silverston, Wood & Zigarmi, 2002). There is a general consensus that the face holds the key to whether an individual is classified as attractive or

unattractive; as this is the most obvious and attainable characteristic which is effortlessly accessible to individuals (Cross & Cross, 1976).

Furthermore, De Meuse's Taxonomy (1987) highlights facial attractiveness as one of the physical appearance cues an individual has the ability to control and modify to some extent (i.e. using make-up). Not only do the facial features reveal evidence regarding the age, sex, ethnicity, physical condition, and current emotional state of an individual; but individuals will also react to others based upon their facial features, by making numerous subconscious assumptions regarding their character and personality based on their looks (Gulz & Haake, 2006).

2.5.1 The Attractiveness Stereotype

Karen Dion is one of the early pioneers in research on the physical attractiveness stereotype. In their seminal work entitled 'What is beautiful is good,' (Dion et al., 1972), the subjects rated photographs of fellow students based on various personality traits as well as questions on expected life outcomes such as career success and marital happiness. One of the main conclusions in this study is the statement, 'What is beautiful is good,' thus linking beauty with goodness. Moreover, this work inferred that attractive people possessed more socially desirable personalities than unattractive individuals. The subjects also expected attractive individuals to lead better lives in terms of occupational success, as well as being more competent husbands/wives than their unattractive counterparts. Dion's work supports the existence of the attractiveness stereotype, whereby the target individual's physical attractiveness is used as a cue in making extensive references about the personality of these individuals in question.

Research investigating the attractiveness stereotype falls into two distinct sections. The first focused on adopting the stranger-attribution paradigm; whereby observers evaluated static images of other target individuals (Eagly et al., 1991; Feingold, 1992). The second branch is an extension to the stranger-attribution literature, named the interaction-based paradigm, where the observer evaluated the target individual in a more ecological and realistic set up, such as in a classroom, hospital or night club (Langlois, Kalakanis, Rubenstein, Larson, Hallam & Smoot, 2000).

2.5.2 Attractiveness within the Stranger-Attribution Paradigm

Eagly et al. (1991) combined a total of 76 existing studies in order to examine the generality as well as the strength of the attractiveness stereotype. The aim was to determine whether observers (subjects) will infer more favourable personality traits towards an attractive rather than an unattractive target (usually depicted in a photograph), without any form of interaction taking place between them.

The personality traits identified by Eagly et al. (1991), were divided into six categories in order to measure the strength of the attractiveness stereotype: social competence (i.e. sociable, popular, and likeable), adjustment (i.e. well-adjusted, mature, and happy), concern for others (i.e. sensitive, empathetic, generous, and modest), integrity (i.e. trustworthy, honest, and faithful to spouse), intellectual competence (i.e. intelligent, rational, and ambitious), and potency (i.e. strong, self-assertive, dominant).

Eagly et al. (1991) hypothesized that the attractiveness stereotype should have a large impact on inferences relating to social competence, a moderate impact on those relating to adjustment, intellectual competence and potency; and an even weaker impact on concern for others and integrity.

Table 1. Effect sizes by Eagly et al. (1991) presented in Wheeler & Kim (1997).

Category	Mean Effect Size
Social competence	.68
Adjustment	.52
Potency	.49
Intellectual competence	.46
Modesty (from concern for others)	-.67
Integrity	.13
Concern for others	.01

As predicted, social competence showed the greatest effect size which supports Eagly et al.'s (1991) argument, that the main core of the attractiveness stereotype are traits such as popularity and sociability (Table 1). Whereas, moderate effect sizes were measured for intellectual competence, adjustment and potency. Modesty implied a strong negative mean effect size, whilst, attractiveness had an almost insignificant effect on concern for others, and very little effect on integrity.

2.5.3 Attractiveness within the Stranger-Attribution Paradigm

The second stranger-attribution based meta-analysis was conducted by Feingold (1992) using a different set of studies to Eagly et al. (1991). Three different sets of literature were explored by Feingold: first, experimental literature looking at the attractiveness stereotype; second, correlation studies which examine characteristics which are associated with physical attractiveness; and the third, studies examining how individuals rate themselves (self-rated attractiveness).

The effect sizes for Feingold's (1992) meta-analysis is highlighted in Table 2, which is very similar to the meta-analysis by Eagly et al. (1991). Social skills and social competence, scored highly in both meta-analysis; implying that attractive people are perceived as being more popular, confident and likeable. Sexual warmth scored the second highest effect size (0.78). Both Eagly et al. (1991) and Feingold (1992) found a weak relationship between integrity and attractiveness, as well as a high negative effect size for modesty.

Table 2. Effect sizes by Feingold (1992) presented in Wheeler & Kim (1997).

Category	Mean Effect Size
Social skills	.88
Sociability	.46
Mental Health	.50
Dominance	.54
Intelligence	.31
Modesty	-.34
Sexual warmth	.78
Character	-.04

The correlation studies by Feingold (1992) indicated that attractive individuals were not troubled by issues regarding the opposite sex; and were least affected by social anxiety and loneliness. Physical attractiveness was also found to be positively correlated with sexual experience, social skills, and popularity amongst the opposite sex. For females, physical attractiveness was strongly correlated with self-esteem, opposite sex popularity, and sexual permissiveness, general mental health, social comfort (except for freedom from self-consciousness), social behaviour measures (except number of same-

sex friends) than correlations related to men. The only negative correlations for females were between physical attractiveness and freedom from public self-consciousness. Subsequently, physical attractiveness was unrelated to personality disposition, i.e. dominance, for both males and females (Feingold, 1992).

Feingold's (1992) meta-analysis of self-rated attractiveness studies revealed a number of interesting findings. Correlations between self-rated attractiveness and self-esteem, intelligence, and current sexual experience were larger for females than those for males. However, in terms of number of sex partners, opposite sex popularity, and global sexual experience the correlations were greater for males than females. Females on the other hand did not show a significant correlation between sexual permissiveness and self-rated attractiveness, whereas positive correlations were obtained with most of the personality measures used except that for freedom from public self-consciousness.

Personality dimensions such as self esteem and dominance were unrelated to physical attractiveness, but were positively correlated to self-rated judgements of physical attractiveness. Additionally, social skills and freedom from public self-consciousness were correlated with physical attractiveness but not with the self-ratings of physical attractiveness. Finally, the greatest socially related measures, i.e. freedom from loneliness and social anxiety, sexual experience, and opposite sex popularity; were strongly correlated with both physical attractiveness and self-rated physical attractiveness (Feingold, 1992).

2.5.4 Attractiveness within the Interaction-Based Paradigm

A recent meta-analysis by Langlois et al. (2000) focused on more ecologically relevant judgements of individuals known to the observer, or strangers one encounters in a shopping store, bank, hospital, or at work. Langlois et al. (2000) investigated three common maxims: 'beauty is in the eye of the beholder,' 'never judge a book by its cover,' and 'beauty is only skin deep.' These maxims seem to be at odds with the stranger-attribution literature (Eagly et al., 1991; Feingold, 1992).

Looking at the first maxim: findings from the attractiveness stereotype literature demonstrate that beauty is not solely in the eye of the beholder. Both within and across ethnicity and cultures, there is general agreement amongst raters as to who is considered to be attractive and who unattractive (Eagly et al., 1991; Feingold, 1992; Langlois et al., 2000; Rhodes, 2006).

The second maxim: never judge a book by its cover. Langlois et al. (2000) supports the notion that people interact differently with others based on the target individual's attractiveness alone. Thus, revealing that the mere thought of finding an individual attractive also permeates in their actions towards the target individual, without the observer being aware of their change in behaviour (Langlois et al., 2000).

Finally, the third maxim: beauty is only skin deep. Here, literature on the attractiveness stereotype attest to the fact that this is not so; as attractive individuals have been observed to exhibit more positive traits than unattractive individuals. In terms of self-perception; attractive adults tended to exhibit greater favourable self-perceptions, such as being more mentally healthy and more competent than unattractive individuals. This demonstrates that perhaps due to the positive way these attractive individuals are treated by people around them, it somehow impacts their mental well being in giving them more confidence and self-belief in themselves; into thinking they really do possess positive personality traits, which is why it is then reflected in their rather positive behaviour. One can conclude that for both adults and children attractiveness is strongly related to popularity and to success (for adults) (Langlois et al., 2000).

Langlois et al.'s (2000) findings also suggest that there is no gender difference, in ecologically valid situations, when it comes to how important attractiveness is perceived by both males and females in most domains. These findings are also replicated in stranger-attribution literature (Eagly et al., 1991; Feingold, 1992). Furthermore, when observing age difference; again children and adults equally value attractiveness as an important fundamental feature.

The main conclusions from this meta-analysis are that attractiveness is actually a significant advantage to children and adults in almost every domain of treatment, judgment and behaviour examined by Langlois et al. (2000). In relation to the maxims that have been discussed, Langlois et al. regard them as merely myths with no strong evidential link to real life, as they state: 'beauty is more than just in the eye of the beholder; people do judge and treat others with whom they interact based on their attractiveness; and surprisingly, beauty is more than just skin deep.'

2.5.5 The Negative Side of the Attractiveness Stereotype

There is a dark side to the 'what is beautiful is good' stereotype which has been labelled as the 'what is beautiful is self-centred' stereotype (Cash & Janda, 1984; Eagly et al.,

1991). This suggests that in one respect attractive individuals are seen as being vain and conceited, selfish, self centred, with a greater tendency to engage in adultery, as well as lacking in modesty than their less attractive counterparts (Dermer & Thiel, 1975; Eagly et al., 1991). Likewise, attractive women have also been viewed as being rather egoistic as well as materialistic (Cash & Janda, 1984; Eagly et al., 1991). Negative associations in terms of concern for others have also been reported as an undesirable personality trait in attractive individuals (Dermer & Thiel, 1975).

Dermer & Thiel (1975) used female subjects to rate female stimuli (photographs), showing that unattractive female subjects actually rated the very attractive female stimuli as having reliably less desirable personality traits than the unattractive targets. This implies that the negative attractiveness stereotype appears to be attenuated for unattractive female participants. Additionally, a study by Spencer & Taylor (1988) reveals how attractiveness, in certain situations, can actually be a hindrance for the individual in the workplace. Hence, attractive males that were given poor performance ratings by their managers, were thought to receive this due to their lack of ability. Whilst those attractive males who were given a good performance rating, were thought to achieve this due to very little effort of their own. When attractive females received poor performance ratings, this was thought to be due to a lack of effort on the females behalf. It seems that individuals who expect attractive individuals to be rather successful may do so by having expectations based on false assumptions.

2.5.6 The Importance of Attractiveness in the Social Domain

Cross & Cross (1971) suggest that females (on average) are seen as being more attractive than males. Furthermore, studies also indicate that both males and females value attractiveness, but males tend to value attractiveness more than females actually do in certain situations. The physical attractiveness level becomes more potent for men when searching for a potential partner (Feingold, 1992). This is further highlighted in a study by Regan & Berscheid (1997), which indicated male student's strong preference for physical attractiveness in a marriage partner; whilst both females and males equally preferred physical attractiveness in a sexual partner.

Bar-Tal & Saxe (1976) reported that attractiveness affects the judgement of both males and females, but the strength of the physical attractiveness stereotype is more potent when applied to women than to men. Subjects were shown pictures of couples,

husbands and wives, which were then evaluated to assess impressions of personality traits, socioeconomic backgrounds, and general personal characteristics. Unattractive males married to an attractive female received quite a favourable and positive evaluation, but an unattractive wife married to an attractive male received quite a negative evaluation.

Berscheid & Walster (1974) conducted a more interactive study based on an ecological set up in a bar. They asked individuals at the bar to give their impressions of a man who walked into the bar with either an attractive or unattractive female with him. The results indicate that when the man entered the bar with an attractive female, he received the most favourable overall impression from the observers. On the contrary, when the man walked in with an unattractive female, he was then viewed quite negatively.

Eagly et al.'s (1991) suggestion that the attractive stereotype is always stronger for females than for males is not fully supported as it seems to depend on the context. This is acknowledged by Walster et al.'s (1966) investigation which highlighted the influential power of physical attractiveness in a computer degree study. The participating students were under the impression that a computer would match them up with a partner according to the answers of their questionnaires they answered prior to the dance; when in fact each student was rated by four judges on their attractiveness levels. The pairing was carried out randomly, and each partner was asked during the date whether they wished to see the other partner again. The overriding factor involved in the decision to want to date their partner again was the date's physical attractiveness.

2.5.7 What Makes a Face Attractive?

A recent meta-analysis highlighted three main factors which make a face attractive: symmetry, averageness and sexual dimorphism (Rhodes, 2006). These three factors have been shown to be reliable across cultures, and for both male and female faces.

A study conducted by Fink, Neave, Manning, & Grammer (2006) concluded that faces which are highly symmetrical received higher ratings from the observer in terms of attractiveness, status of health, and some personality attributes: lively, intelligent, self-confident, balanced, and sociable. Furthermore, less symmetrical faces were rated as being more anxious. Hence, facial symmetry is considered as a cue to an individual's quality in relation to certain personality traits.

Rhodes (2006) conducted experiments on how the user rated composite faces in comparison to individual faces. The findings are unanimous in that the average facial composites and configurations are usually rated as being more attractive than the individual faces making up these composites.

The final factor, sexual dimorphism, tends to signal the reproductive potential and sexual maturity of an individual. Findings on sexual dimorphism suggest that in general feminine traits: large eyes, high eyebrows, full lips, small nose, small chin, prominent cheekbones and narrow cheeks are rated by both males and females as being beautiful. Female beauty tends to be associated with more feminine or childlike qualities. While, masculine traits: square chins, thin lips, small eyes, and thick brows tend to signal dominance and status which enhances their mating value. During the fertile phase of the menstrual cycle, female preference tends to move towards more masculine faces at this time (Rhodes, 2006).

Additionally, Cunningham (1986) conducted an experiment where college students rated photographs of beauty contest finalists and ordinary-looking college women, and then analysed the differences in facial features between the two groups. One way in which they differed was that the beauty contestants tended to have widely spaced eyes, small noses, small chins, wide pupils, high eyebrows, and a big smile. These features were associated with positive personality ratings, such as being intelligent, sociable, and assertive.

Whilst, in a study by Cross & Cross (1971), subjects were asked to rate 72 photographs of individuals of similar ages, both sexes and various races, in terms of which facial features they found most important and attracted them the greatest. The results demonstrated the following proportion of subjects who preferred certain facial features: first the eyes, chosen by 34%; the mouth/smile by 31%; hair, by 10%; skin colour, by 5%; shape of the nose, by 5%; finally, the facial proportions of the face as a whole were chosen by 15% of subjects.

2.6 Stereotypes within HCI

Zanbaka et al. (2006) investigated the persuasiveness of speakers and reported that both male and female participants were persuaded more by an embodied agent of the opposite sex. Baylor & Kim (2003) report that learners perceived female agents as being more agreeable and extrovert in comparison to male agents. Whilst, male agents

had the potential to improve learning related results, as they seemed to impact self-regulation and learner satisfaction more than female agents Furthermore, Baylor et al. (2003) show that in comparison to male learners, female learners were more likely to choose to work with a specific agent basing their decision on previous experiences with human teachers.

Baylor & Plant's (2004) study showed that female learners (pre-service teachers) who worked with a non-stereotypical female engineer agent, called Nina, were more willing to believe that they themselves had the intelligence to become a successful engineer; and clearly wanted to learn more about this field after being positively influenced by Nina. Nina was presented as an attractive looking engineer agent, who was more influential as an agent model for engineering in comparison to a second less attractive agent.

Baylor et al. (2003) investigated the role of ethnicity (African American vs. Caucasian), gender (male vs. female), realism (realistic vs. cartoon like), and the learner's choice of agent. Approximately 183 undergraduate students were asked to choose from one of eight PA's to learn from; in regards to the topic 'Coping with College life.' The results highlight how the African American learners not only preferred to interact with an agent of a similar ethnicity to themselves, but also had a positive attitude towards these agents.

The findings of Baylor et al. (2003) were consistent with another investigation by Baylor & Kim (2003). Here, the African-American learners also reported that they preferred a PA which was similar to them in terms of gender and ethnicity. Learners working with agents of the same ethnicity viewed these agents to be warmer, friendlier and more engaging. However, Moreno & Flowerday (2006) show how learners who chose to work with PA's of the same ethnicity, seemed to be more focused on how the PA represented the actual learner rather than concentrating on the task at hand (due to being distracted by the ethnicity of the PA).

Only two studies seem to have evaluated the effects of the age of the embodied agent on the user. The first being a study by Baylor, Rosenberg-Kima, & Plant (2006) which indicated that young looking agents were classed as cool looking by the learner when the agent was of a similar age group to these participants. However, the older looking agents representing stereotypical engineers were more effective in influencing participants to pursue engineering related careers, as they were perceived as being more

experienced and wiser than the young looking agents. In the second study, Cowell & Stanney (2005) demonstrated that users preferred to interact with embodied agents of a similar age. This reflects studies in social psychology which show how humans prefer to be influenced more by their own in-group members; which in this instance would be of a similar age group (Nelson, 2004).

2.6.1 Effect of Embodied Agent & Avatar Attractiveness Stereotypes

Early observations by Sproull et al. (1996) looking at agent attractiveness indicated that attractive and pleasant looking agents were deemed by users to have more of a personality than unattractive agents. Nowak & Rauh (2005) investigated how participants viewed and rated eight static images of avatars which varied from human characters to animals, including various objects such as a bottle or hammer; and reported a preference for human avatars over the non human avatars presented to them. The second study (Nowak & Rauh, 2008), again required the participants to rate static images of 30 avatars before selecting one avatar to represent them in a chat session with another participant. Both studies suggested that participants perceived the more anthropomorphic avatars to be considerably more attractive and credible, and preferred to be represented by these avatars online; whilst feminine avatars were perceived as being more attractive than masculine avatars.

Vasalou & Joinson (2009) further investigated the perception of avatars created by participants for three scenarios: blogging, dating and gaming. Participants created an attractive avatar for the dating scenario, whilst creating a more intellectual looking avatar for the gaming scenario. Additionally, a study on Second Life (Messinger, Ge, Stroulia, Lyons, Smirnov, & Bone, 2008), found that users make their avatars not only similar, but also somewhat more attractive than themselves. These findings are further supported by Yee & Bailenson (2007); who found that participants represented by attractive avatars in online virtual environments were more intimate with and more willing to approach members of the opposite gender, as compared to participants represented by less attractive avatars.

These studies underline how the visual appearance of an avatar (i.e. gender or levels of realism/anthropomorphism and attractiveness) can effect the perception and behaviour of the user.

2.6.2 The Attractive Stereotype and Persuasion

Persuasion can be defined as 'an occurrence when an individual endeavours to induce some sort of change in the attitudes, beliefs, or behaviour of another individual or group of people' (Zanbaka et al., 2006). Chaiken (1979) reports that attractive communicators were significantly more persuasive than unattractive communicators. Fogg (2003) explains that people respond socially to computer products, and these computer products can behave as persuasive social actors which have the ability to elicit social responses and behavioural change in humans.

Zanbaka et al. (2006) investigated the persuasiveness of virtual speakers. The results suggested that virtual speakers are just as effective in changing an individual's attitude as a real person; whilst male participants were persuaded more when the speakers were actually female than male. The virtual speakers were also rated more positively in terms of the way they were perceived by the participant than the human speakers. According to Pratt, Hauser, Ugray, & Patterson (2007), users were more wiling to be persuaded and change their actions when they received advice from an embodied agent of a similar ethnic background to their own.

A significant gap in HCI research is that few studies have investigated the effects of persuasion due to the attractiveness levels of an embodied agent. However, a study by Holzwarth et al. (2006) allowed users to interact with virtual agents acting as online sales assistants. The results indicated that attractive agents were perceived by users as being more persuasive and effective sales agents, across all levels of user involvement, than the less attractive agents when purchasing goods.

2.7 Conclusion

In recent years the quest to make the interaction between the user and interface more life like and believable has resulted in the increased use of embodied agents. Researchers have found many uses for these agents within various social domains: in entertainment, e-commerce, education, as online help assistants, marketing or sales assistants, health or mortgage advisors, as well as virtual friends (Bickmore & Picard, 2005).

The stated advantages of using embodied agents has also been criticised by a group of HCI researchers who regard the claims as being misleading (Shneiderman & Maes,

1997). However, many HCI designers and practitioners (Walker et al., 1994; Sproull et al., 1996; Nass et al., 2000; Baylor & Kim, 2003; Fogg, 2003; Baylor & Kim, 2004; Zanbaka et al., 2006; Haake & Gulz, 2008) are exploring how users and agents interact, by applying the concepts of the Media Equation (Reeves & Nass, 1996) and CASA paradigm model to agent design. The basic premise of the Media Equation and CASA model is that humans respond in a similar way to computers as they do towards people. Therefore, in order to understand this human-agent relationship, one can apply theories and findings from social psychology regarding human-human interaction and assign these to human-agent interaction.

Whilst researchers have shown that the basic premise of the Media Equation and CASA model can be successfully applied to human-agent scenarios; some studies have shown that users will abuse and behave in a negative manner towards chat-bot agents (De Angeli et al., 2001; De Angeli & Brahnam, 2008; Veletsianos et al., 2008), and that people will not always attribute human-like characteristics to technology (Morkes, Kernal, & Nass, 1999).

Numerous social psychology studies have pointed to the impact visual cues have on perception, impression and stereotype formation in human-human interaction (Taylor et al., 1978; Fiske & Taylor, 1991; Fiske, 2000; Nass et al., 2000). These social interaction studies have demonstrated how an individual's gender (Williams & Bennet, 1975; Eagly & Mladinic, 1989), age (Nelson, 2004), ethnicity (Ford, 1997; Nass et al., 2000), and levels of attractiveness (Dion et al., 1972; Berscheid & Walster, 1974; Dermer & Thiel, 1975; Cross & Cross, 1976; De Meuse, 1987; Eagly, 1991; Feingold, 1992; Wheeler & Kim, 1997; Langlois et al., 2000; Prestia et al., 2002) can elicit various perceptions, stereotypes and behavioural changes within the observer. However, embodied agent researchers have only in recent years begun to explore the effect of the agent's age, gender, ethnicity and realism (Nass et al., 2000; Baylor & Kim, 2003; Baylor et al., 2003; Baylor & Kim, 2004; Baylor & Plant, 2004; Moreno & Flowerday, 2006; Zanbaka et al., 2006) on the human user. Moreover, only a handful of these researchers have examined the effect of agent attractiveness (Nowak & Rauh, 2005; Messinger et al., 2008; Nowak & Rauh, 2008; Vasalou et al., 2008; Vasalou & Joinson, 2009) on user perception, stereotypes and behaviour towards the agent.

Furthermore, there has been no analysis of design trends for agents in terms of their visual cues: such as the agent's age, gender, ethnicity, role, realism/anthropomorphism

level, or even levels of attractiveness. Such an investigation into the agent's demographic and physical attributes give a better insight into the biases and stereotypes held by these agent designers and practitioners. Therefore, more research is requested to advance the current theoretical and empirical understanding of embodied agents, in order to improve human-agent design.

3 Chapter 3: Agent Census and Online Agent Perception Study

This chapter reports the first two sets of investigations of the thesis. The first being a census of 188 embodied agents, highlighting the prevailing design trends in terms of the agent's demographic and physical attributes. These details were recorded and presented within an agent database tool (ADT). The second study was an empirical online survey study which investigated how participants perceived agent attractiveness and realism. Both investigations point to widespread designer bias which favours the development of younger looking agents from an ethnically white background; as well as realistic and attractive looking female agents. Accordingly, the online survey study demonstrated the power of agent embodiment in terms of the influence of their physical appearance on user perception; whereby young female agents were perceived by participants as being more attractive and realistic than male agents. Furthermore, a strong positive correlation was also elicited between agent attractiveness and agent realism. Participant comments from the online survey are also reviewed. A classification system in relation to the agent attributes was devised and proposed for use by other HCI agent researchers and practitioners.

3.1 Introduction

The last decade has witnessed a shift towards a more anthropomorphic interface, resulting in the increased use of embodied agents in various domains (education, e-commerce, health advisors, entertainment, etc.) Despite this growing trend, a valuable sphere of research regarding embodied agents has largely been ignored by HCI agent designers and researchers. This is the absence of an in depth census on embodied agents which have been developed by HCI practitioners since their appearance in the late nineties. In order to address this gap, the aim of the initial investigation was to report a census on 188 agents, highlighting any obvious designer bias or common design trends amongst this embodied agent corpus, which could also point to any type of bias or stereotypes designers may possess.

3.2 Study 1: Embodied Agent Census Study

The expectations for this initial study was to provide HCI designers and researchers of embodied agents an insight into the various design trends which were commonly found amongst the embodied agents. The findings of the first census (Khan & De Angeli, 2007) investigating 147 agents was updated in the second census within this thesis, whereby additional agents were added to the data and reanalysed for a total of 188 agents.

3.2.1 Method

A total of 188 embodied agent faces were collected by a process of internet searches in online journals and conference proceedings (ACM library and Science Direct), conference sites (IVA: Intelligent Virtual Agents conference from 2003), and search engines (Google Scholar) using the following keywords: embodied conversational agents (ECA's), conversational agents, pedagogical agents, social virtual humans/agents, virtual humans/agents, virtual characters, conversational virtual humans/agents, synthetic virtual humans/agents, synthetic humans/agents, synthetic characters, and avatars. A number of E-mails were also sent to various mailing lists (British HCI, CHI Announcements, and CHI Students) to invite researchers to share their pictures of embodied agents they had either developed or utilised within their research.

The images were initially selected based on the following criteria:

- Human-like (No animal characters)
- Frontal view only, and
- Good quality images which were not too small (at least 7 x 5 cm).

3.2.2 Materials

An agent database tool (ADT) was developed, using Microsoft Access 2003, in order to store and analyse the agents' data. Each agent was assigned a unique ID number, including specific attributes pertaining to each agent within the database tool; such as demographic cues: age group, ethnicity, gender; and physical appearance cues: clothing style, facial attractiveness, and anthropomorphism levels. Additional information for each agent included their pictures, author details including the paper/journal name or other sources; as well as the results from participant evaluations of the agent's perceived attractiveness and realism levels from the online survey study which was conducted after the census study.

Furthermore, in order to maintain consistency, all agent pictures were edited to show just the face and top portion of the shoulders, as well as modifying each picture to have a plain white background; reflected in the example in Figure 11. All 188 pictures were edited using Adobe Photoshop 7.0.

Figure 11. An example of one of the 188 edited embodied agents.

3.2.3 Framework of Analysis

Prior to adding the agents to the ADT, a framework of analysis was devised to construct the characteristics and attributes for each agent. Inter-rater reliability tests/observations were carried out amongst multiple observers (fellow colleagues) within the department. The initial reliability test (see Appendix A) required the observer to note the category they felt was most appropriate in describing each of the 30 agents (randomly selected by

the author); in terms of the agent's age group, ethnicity and anthropomorphism level. The categories that were given to the observers were initially devised by the author to test for three main attributes:

Age:

- *Child* – An individual between birth and puberty;
- *Young Adult* – An individual between puberty/teens and the age of 30;
- *Adult* – An individual Between ages of 30 and 50; and
- *Older Adult* – An individual over 50.

Ethnicity:

- *White* – Faces originating from Caucasian/European background;
- *Black* – Faces originating from African background;
- *Asian* – Faces originating from South Asian background; and
- *Oriental* – Faces originating from the Far East.

Anthropomorphism:

- *Cartoon* – faces which do not represent real people. They can be sketches, or humorous images often displaying some exaggeration of facial characteristics (caricatures);
- *Drawing* – 2 dimensional representational images featuring human-like faces;
- *Mannequin* –3 dimensional representational images of human-like faces; and
- *Photo realistic* - Pictures of real human beings or artificial faces which are extremely human-like, so that they could be erroneously attributed to a real person.

The inter-rater reliability was vital to maintain a high level of general agreement and consistency amongst the observers and the author. Initially, a total of six observers were used to test the reliability of the three attributes: as three of these observers gave their opinion for one set of agents (shown in Group 1 in Table 3), the remaining three observers gave their opinions of a different set of agents (Group 2 within Table 3) by filling in the observer scoring sheet shown in Appendix A.

Table 3. The Calculated Agreement Index for Group 1 and 2 ([a] Anthropomorphism Level).

% Agreement – Agent Group 1			
Observer	**Age**	**Ethnicity**	**Anth [a]**
1	90	93	60
2	83	86	66
3	93	80	50
% Agreement – Agent Group 2			
4	80	83	53
5	86	96	56
6	93	90	46
% Total	87	88	55
Total Overall Agreement		**77%**	

This was carried out whilst viewing 30 agents on a monitor screen in front of them. Once this was completed, an agreement index was calculated for each of the 3 categories for both Agent groups; and then the total overall agreement index for the combined categories as presented in Table 3.

The percentage agreement was worked out as follows:

- Percent Agreement = No. of agreements / No. of observations X 100.

Both sets of observers for Group 1 and 2 gave a high inter-rater agreement of 80% + for the age and ethnicity attributes. However, anthropomorphism scored 55%, which brought the total overall agreement down to 77%. Discussions took place with the observers regarding the anthropomorphism categories:

- *Cartoon* – faces which do not represent real people. They can be sketches, or humorous images often displaying some exaggeration of facial characteristics (caricatures);
- *Drawing* – 2 dimensional representational images featuring human-like faces;
- *Mannequin* –3 dimensional representational images of human-like faces; and
- *Photo Realistic* - Pictures of real human beings or artificial faces which are extremely human-like, so that they could be erroneously attributed to a real person.

It became apparent that three definitions required further clarification (*cartoon*, *drawing* and *mannequin*) as they seemed to be misleading and be interpreted in various ways. The three categories in question were modified and rewritten to make them sound more

comprehensible to all observers. This resulted in the removal of the drawing category which was integrated within the *mannequin* definition, resulting in three categories defining anthropomorphism:

- *Cartoon* – Faces which are based on humans which can be drawing or sketches often displaying some exaggeration of facial characteristics (caricatures).
- *Mannequin* – Representational images of human-like faces which are not *cartoon*-like, and at the same time cannot be mistaken for a real person.
- *Photo Realistic* - Pictures of either artificial faces or real human beings which are extremely human-like, so that they could be erroneously attributed to a real person.

In order to test the inter-rater reliability of these new categories, the same two groups of agents were tested again with six other observers using the observer scoring sheet in Appendix B. The results, presented in Table 4, show the calculated percentage agreement values for all three categories in Group 1 and 2.

Table 4. Calculated Agreement Index for anthropomorphism categories in Group 1 & 2.

% Agreement – Agent Group 1			
Observer	**Cartoon**	**Mannequin**	**Photo Real**
7	25	27	100
8	28	25	100
9	24	23	96
% Agreement – Agent Group 2			
10	24	28	100
11	27	24	100
12	23	28	100
% Total	**84**	**86**	**99**
Total Overall Agreement		**90%**	

The original categories gave a total percentage agreement of 55% and after modification of these categories, and a second round of inter-rater observations yielded a 35% increase to an impressive 90%. The final framework of analysis consisting of seven attributes relating to each agent is presented in Table 5.

Table 5. Finalised Framework of Analysis for all agents within the ADT.

Attribute	Category
Gender	• *Male* • *Female*
Age	• *Child* – An individual between birth and puberty • *Young Adult* – An individual between puberty/teens and the age of 30 • *Adult* – An individual between ages of 30 and 50 • *Older Adult* – An individual over 50
Ethnicity	• *White* – Faces originating from Cauc*asian*/European background • *Black* – Faces originating from African background • *Asian* – Faces originating from South Asian background • *Oriental* – Faces originating from the Far East
Dress Style	• *Casual* – Informal clothing and not dressy • *Formal* – Designed for wear or use in certain occasion/event or role • *Uniform* – A job specific outfit (i.e. Doctor or Police) • *Missing* – No outfit is visible, only face and neck displayed
Anthropomorphism Level	• *Cartoon* – Faces which are based on humans which can be drawing or sketches often displaying some exaggeration of facial characteristics (caricatures) • *Mannequin* – Representational images of human-like faces which are not cartoon-like, and at the same time cannot be mistaken for a real person • *Photo realistic* - Pictures of either artificial faces or real human beings which are extremely human-like, so that they could be erroneously attributed to a real person
Role	• *Pedagogical Agent*– Agent that facilitate the learning process; • *Actor* – Performs the role of a character within a scenario • *Storyteller* – A narrator of anecdotes, incidents, or fictitious tales • *Assistant* – An agent who assists, supports, guides and helps the user • *Presenter* – An agent which presents/reads out the daily news and weather forecast
Name	• *Name* – An agent with a personal human-like name (such as Peter, and Lucia) • *No Name* – An agent with no human like name

3.2.4 Results

The data was analysed by running queries on a total of 188 embodied agents (92 female, and 96 male) within the ADT. Gender comparisons were conducted on all seven previously mentioned agent attributes (Table 5), which were also used to produce timelines showing usage trends from 1997 to 2006.

Figure 12 represents a trend analysis regarding the development of male and female agents from 1997 to the end of 2006. Male agents were researched initially in 1997 before the interest in female agents began in 1999. There was an increase in the deployment of both male and female agents from 1999 to 2002, and then a sharp rise for both genders between 2002 and 2003, from which point their development is still expanding.

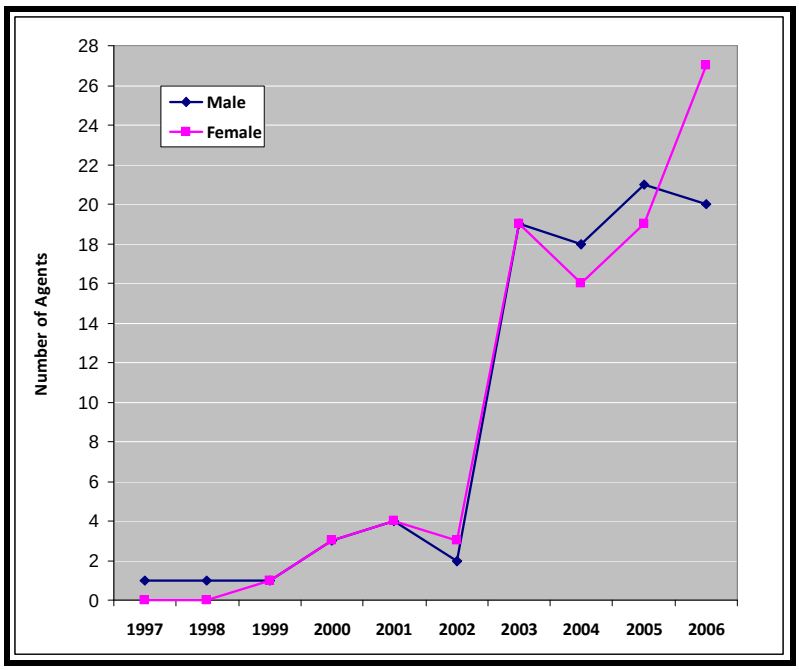


Figure 12. Frequency of agents based on gender according to Timeline.

3.2.4.1 Ethnicity

Table 6 reports the frequency values for male and female agents as a function of their ethnicity. A large bulk of embodied agents (84%) are from a *white* ethnic background, and the remaining 16% represent various other ethnic groups: *black*, *asian* and *oriental*.

Table 6. Gender by ethnicity frequency distribution.

	Male	***Female***	***Total***
White	78	80	158
Black	10	11	21
Asian	4	1	5
Oriental	4	0	4
Total	**96**	**92**	**188**

In terms of the age distribution; approximately 66% of *white*, 90% of *black*, and 100% of *asian* and *oriental* agents are classed as *young adult*s. The timeline graph in Figure

13 highlights the earliest development of *white* agents from 1997 to 2002. It is only from 2003, that the first *black* agents were created; and later still, in 2005 and 2006, the first use of *asian* and *oriental* agents were recorded.

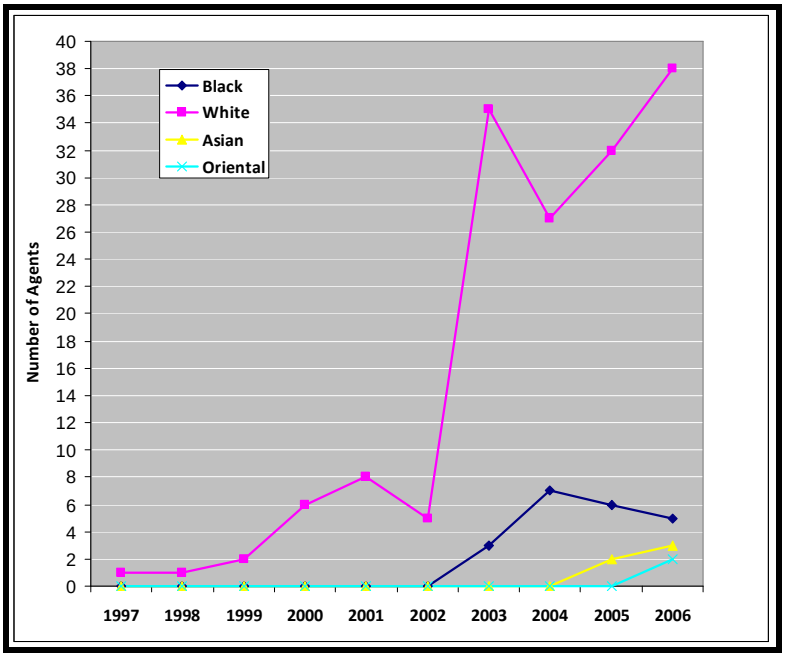


Figure 13. Frequency of agents based on ethnicity according to Timeline.

3.2.4.2 Age

Table 7 illustrates estimates of male and female agents' age. Approximately 69% of agents are *young adults*, whilst 31% of agents are of a combination of other age groups: *child*, *adult* and *older adult*; the smallest of which is the *older adult* group consisting of just 12 agents. *Young adult*s are predominantly female whilst the other older groups such as *adult* and *older adult* are dominated by male agents.

Table 7. Gender by age frequency distribution.

	Male	***Female***	***Total***
Child	10	5	15
Young Adult	54	76	130
Adult	23	8	31
Older Adult	9	3	12
Total	**96**	**92**	**188**

The predominant anthropomorphic level for *child* agents is that of *cartoon* style (88%), followed by *mannequin* style agents (12%). *Young adult*s are substantially *mannequin*-like (51%), with a modest number of *cartoon* (26%) and *photo realistic* (23%) looking

agents. The number of *photo realistic* agents tends to decrease as the age of the agent increases from *adult* (19%) to *older adult* (16%). This pattern is also the case for *cartoon*-like agents whereby the number of *cartoon*-like agents decrease as the age of the agents increase from the *adult* age group (9 *cartoon*-like agents) to the *older adult* age group (1 *cartoon*-like agent).

The timeline based on agent age group in Figure 14 illustrates how the initial use of *young adult* agents commenced in 1998 which has steadily increased to 2002, then steeply rising in their usage from 2003 onwards. The next prevalent age group, *adults*, first appeared in 1997, with only a handful of these agents being developed up to 2002, showing only a slight rise in their numbers (by 7) to 2003. *Child* agents did not appear until 2001, whilst *older adults* were not investigated or developed until 2003; both groups had lower frequencies than the rest.

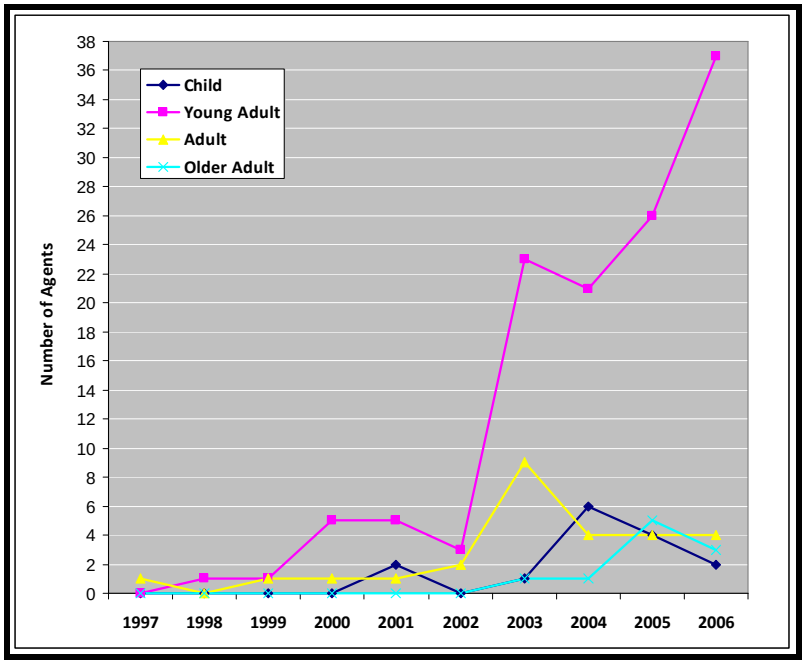


Figure 14 Frequency of agents based on age according to Timeline.

3.2.4.3 Anthropomorphism Levels

The data on agent anthropomorphism levels is illustrated in Figure 15, whereby *cartoon*-like agents are predominantly made up of male agents. As the realism levels increase, the number of female to male agents within these groups (*mannequin* and *photo real*) also increases, showing a possible gender effect. The most prevalent anthropomorphic category amongst embodied agents is that of *mannequin*s. The major ethnic group represented by embodied agents is *white* for all three anthropomorphic categories: *cartoon*- like (81% are *white*), *mannequin* (87% are *white*), and *photo real*

(82% are *white*). The second most common ethnic group is *black*: *cartoon*-like (10%), *mannequin* (11%), and *photo real* (13%). The two remaining ethnic groups: *asian* and *oriental*, both of which make up no more than 5% of the embodied agent populations for the three mentioned anthropomorphic levels.

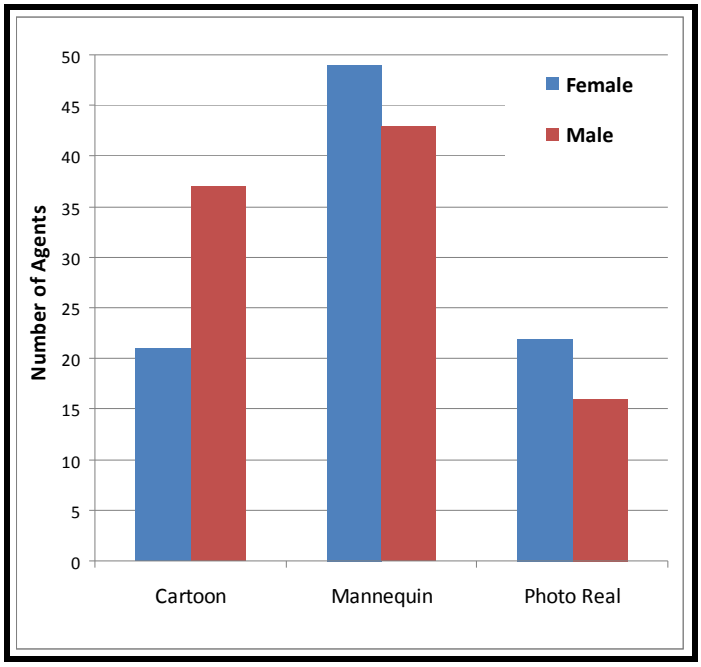


Figure 15. Frequency of agents reported by anthropomorphism based on gender.

The trend analysis in Figure 16 representing the anthropomorphism timeline shows a gradual increase in the development of *cartoon* (from 1997) and *mannequin* agents (from 1998). *Photo realistic* agents were not actually deployed until 2001. Additionally, all three groups displayed a marked increase in number from 2002 to 2003, with a steady number of agents being developed from this point.

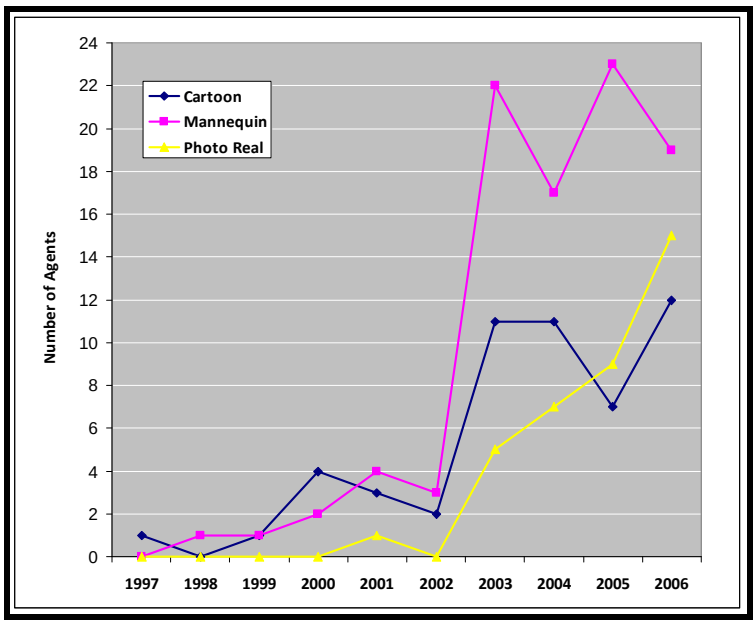


Figure 16. Frequency of agents based on anthropomorphism levels according to Timeline.

3.2.4.4 Role

Figure 17 represents the distribution of the five types of roles assigned to EA's. The most common is the *pedagogical* role, with an equal number of males (N = 30) and females (N = 30). Approximately 75% of agents were assigned a role by their developers whilst the remaining 25% had none. A gender effect may be present within the *presenter* and *storyteller* role, as these seem to be predominantly assigned to female agents.

Pedagogical agents are largely made up of *white* agents (72%) and *young adult*s (72%). These *young adult*s being mainly composed of females (N = 24) than males (N = 19). Embodied agents that play an acting role also tend to be composed of mainly *white* (93%) agents and *young adult*s (62%). This pattern is similar for embodied agents assigned with an *assistant* and *presenter* role. Additionally, *storyteller* agents are 100% *white*, 75% of which are classed as *child*ren, whilst 75% of *storyteller* agents are also considered *cartoon*-like in appearance.

In terms of the anthropomorphic representation; over half of the *pedagogical* agents (56%) are *mannequin*-like, and the remaining agents are equally divided between *cartoon* (22%) and *photo realistic* looking (22%) agents. agent *actor*s are represented by a large number of *cartoon*-like (41%) characters, followed by *mannequin* (33%) and then *photo realistic* (26%) looking agents. Agents that play the role of an *assistant* tend to be *cartoon*-like (52%), with *mannequin*-like agents making up 31%, and *photo realistic* agents around 17%. *Presenter* agents do not consist of any *cartoon* looking

characters, and are composed of mainly *mannequin* and *photo realistic* agents at a ratio of 2:1 respectively. The *storyteller* agents are mainly made up of *cartoon*-like (N = 3) and a single *photo realistic* (N = 1) looking agent.

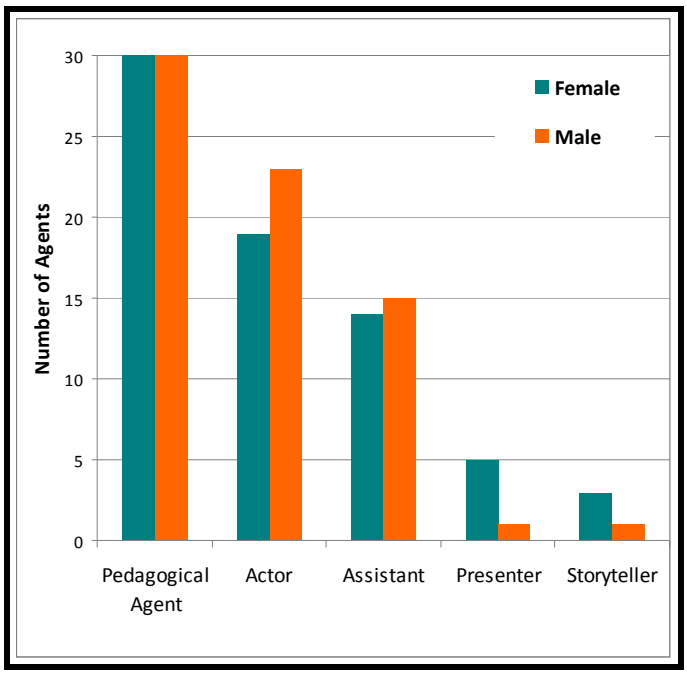


Figure 17. Frequency of agents reported by role based on gender.

Figure 18 highlights the increasing tendency by researchers to apply three main roles (*pedagogical*, *actor* and *assistant*) to agents from 2003 onwards. The *storyteller* and *presenter* roles were initially being assigned to agents in 2001 and 2002 respectively, and remain the least common roles to be assigned to embodied agents throughout the timeline journey.

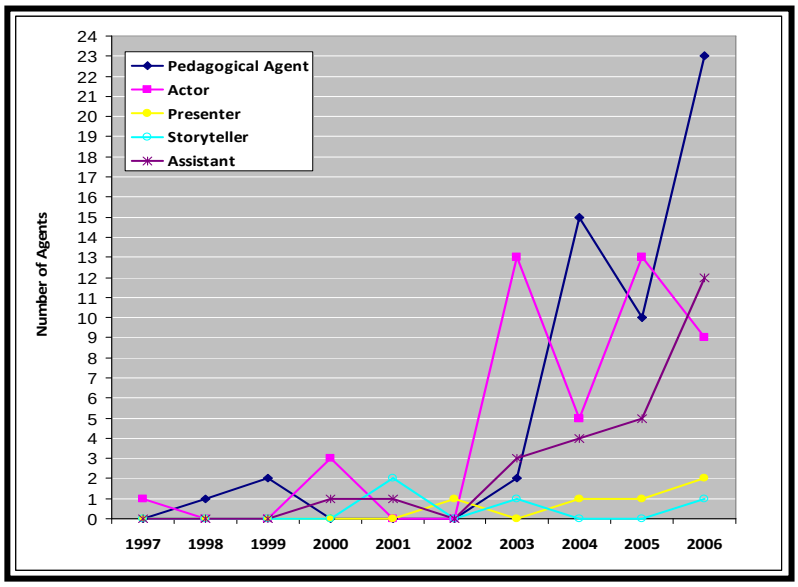


Figure 18. Frequency of agents based on role according to Timeline.

3.2.4.5 Dress Style

Figure 19 represents the three agent dress styles: *casual*, *formal* and *uniform*. The most prevalent dress style assigned to agents is *casual* (62%), followed by a *formal* (14%), and then *uniform* style (2%); leaving 22% of agents with no clothing style (called *missing*) as they are presented as just a face and neck display. Approximately 72% of the agents that were dressed *casual*ly are classed as *young adult*s, whilst 78% of *casual*ly dressed agents are from a *white* ethnic background. A similar pattern is also found with agents dressed *formal*ly, whereby 59% of these agents are *young adult*s, and 88% of them being ethnically *white*. A handful of agents which dressed in a *uniform* style (N = 3) were solely *white*, with only a single agent representing each of the remaining age groups: *child* (N = 1), *adult* (N = 1), and *older adult* (N = 1).

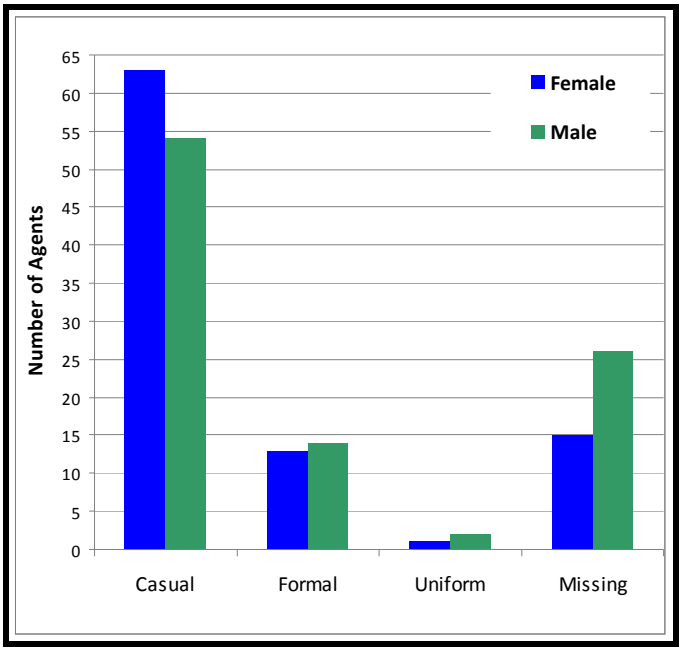


Figure 19. Frequency of agents reported by dress style based on gender.

The timeline graph in Figure 20 represents the increasing trend in using *casual* dress for agents from 1998, and a further spurt in growth from 2003. The use of *formal* dress style was minimal up to 2005, from where a slight increase has been observed in more agents wearing *formal* dress.

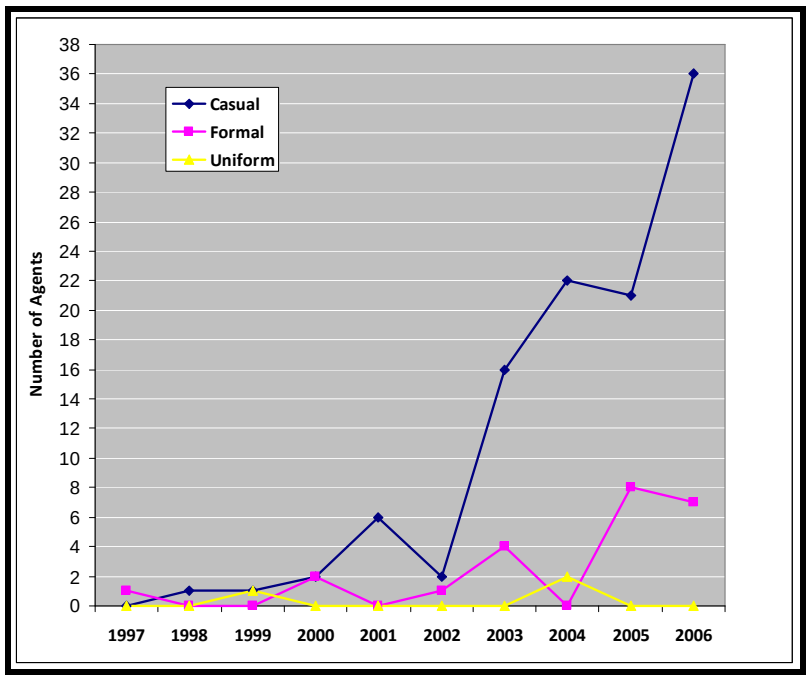


Figure 20. Frequency of agents based on dress style according to Timeline.

3.2.4.6 Name

The number of agents which have been assigned *no name* (N = 146) is greater than those which have been given *names* (N = 42), see Figure 21. Out of those that have *name*s, 93% are *white* and the remaining three agents in this group are *black* (N = 2) and *asian* (N = 1). Approximately 90% of the agents with *name*s have a role assigned to them. Out of these, the most prevalent role amongst *name*d agents is the *actor* role (37%) which is totally 100% *white*; closely followed by the *assistant* agents (34%) out of which 85% are *white*. This leads on to the more modest agent group size such as the *pedagogical* agents (13%), *presenter* agents (11%) and finally the *storyteller* agents (5%); the latter two groups being 100% *white*.

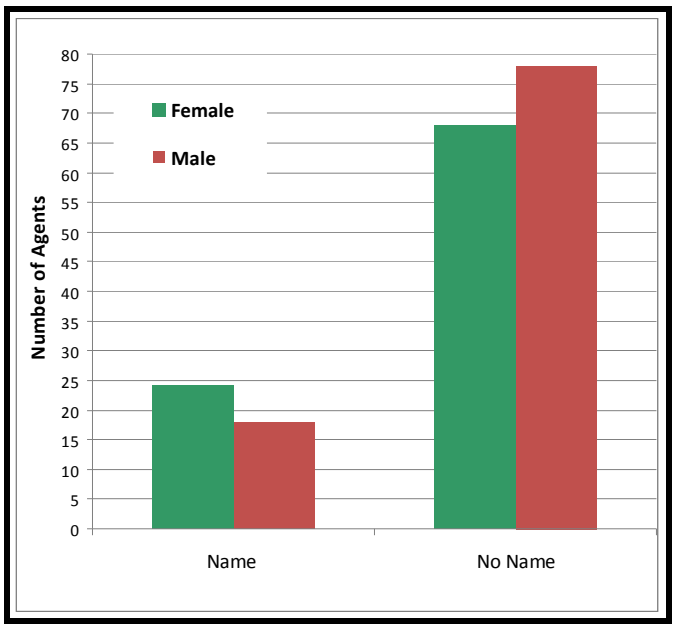


Figure 21. Frequency of agents reported by name based on gender.

Figure 22 illustrates the general tendency for designers to *name* agents, which has gradually began to change in the late 1990's. The earliest recording of agents with *name*s was in 1998, which has been steadily increasing to 2005.

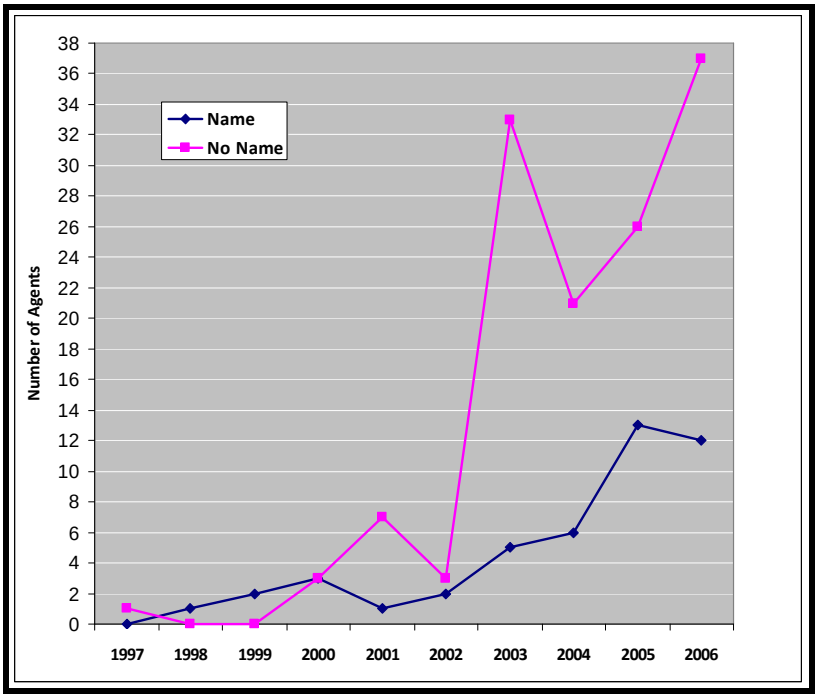


Figure 22. Frequency of agents based on name according to Timeline.

3.2.5 Conclusion

The Census clearly illustrates how the development of both male and female agents by HCI designers and practitioners has sharply increased from 2002 to 2003; with a steady increase since that time. This is probably related to the advances in graphic design technology and improved multimedia and communication software. For example, from the earlier introduction of Microsoft Agents as part of Microsoft Windows 2000, to the latest Sitepal agent development tool, which have allowed designers to create embodied agents. Since these advances in technology have assisted designers in creating more realistic or anthropomorphic agents, which is reflected by the sharp increase in the number of photo realistic looking agents since 2003. It must be noted that these agent designers have a tendency to make more photo realistic and anthropomorphic looking female agents than male agents, who form a large proportion of the cartoon-like agents group.

It may be assumed that the designers of embodied agents are predominantly male; who value female attractiveness to a larger extent than females value male attractiveness (Feingold, 1992; Regan & Berscheid, 1997). This is also a reflection of how society in general advocates and favours female beauty over male beauty; as the evaluation of a female depends so much on her physical attractiveness, whereas males tend to be evaluated on their possessions (Eagly et al., 1991; Feingold, 1992).

The results also demonstrated that the most common role that is usually given to an agent is that of a pedagogical role. However, agents have a lot of potential in playing other roles, rather than just that of a pedagogical one. Furthermore, HCI agent designers seem to be following an ethnocentric approach when it comes to developing agents of varying ethnicities. This may be due to the fact that these designers are predominantly from a white ethnic background themselves, and unassumingly create white agents representing their own ethnic makeup. The significantly greater number of white agents in comparison to agents of other ethnic backgrounds leads to a mismatch with the potential users. Designers should be aware that potential users are global and vary considerably in terms of their ethnicity.

The focus on developing young adult agents, which has been dramatically increasing since 2003, may be due to the perception by designers that their users are from a similar age group; who may also prefer to interact with agents of a similar age group. Individuals that are identified to be from the same in-group (i.e. same age group) are

perceived as having the same values as them, including attributes such as being more trustworthy, competent, and socially attractive. This is further supported by HCI studies (Cowell & Stanney, 2005; Baylor et al., 2006) which have highlighted how users prefer to interact with agents from a similar age group as themselves. Additionally, a negative bias has been reported towards older individuals (i.e. ageism); whereby in western culture they are perceived as being useless members of society (Nelson, 2004). Therefore, the increased use of agents from this older adult age group within more positive roles could possibly reduce the ageism stereotype held amongst users.

Subsequently, the recent trend for designers to give names to agents may be a way of adding personality to the increasingly realistic agents. Therefore, a named agent increases the realism level of that agent by personalising it in this manner.

3.3 Study 2: Online Agent Perception Study

The agent census study revealed a common trend for designers to place greater emphasis on developing a predominantly white and younger adult agent population. This drew attention to the bias and stereotypes held by these agent designers. The following online agent perception study examined how users evaluated and perceived these embodied agents in terms of their attractiveness and realism levels; and if there were any visible patterns or effects due to the agent's demographic attributes (i.e. gender, age, and ethnicity). A total of 145 agents (76 male, and 69 female) were rated in an online survey by 545 participants, on their levels of perceived attractiveness and realism levels.

The online agent perception study aims were:

- *To help designers to understand how users perceive agent attractiveness and realism levels.*
- *To demonstrate whether there is a relationship between agent attractiveness and realism levels.*
- *Add to the current knowledge on embodied agent design, thus improving human-agent interaction.*
- *Produce evaluated images of agents which can be utilised for future experiments: i.e. replicating the 'what is beautiful is good' study by Dion et al. (1972).*

3.3.1 Materials

The online survey study was developed using Survey Gizmo, which supports creation of online surveys, polls and quizzes.

145 agent pictures were selected from the agent database tool (ADT), and uploaded to Google Pages from where a hyperlink was provided for each agent (Figure 23), to the online survey. A total of five surveys were developed using Survey Gizmo, each containing pictures of 29 embodied agents.

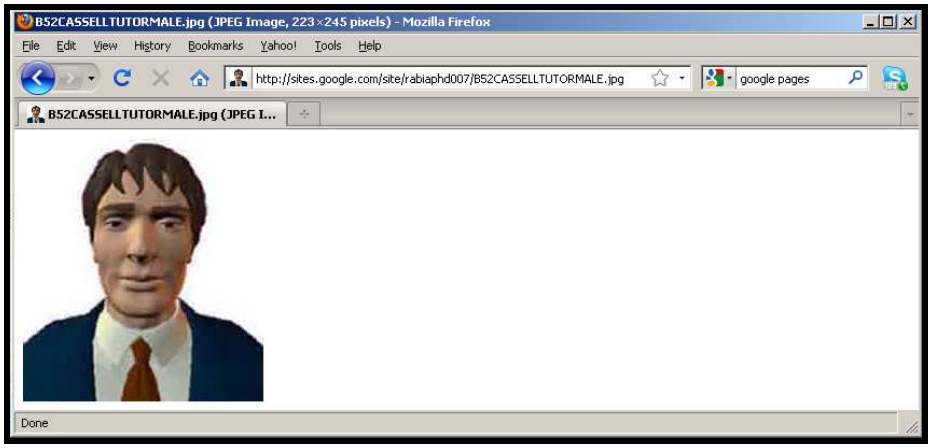


Figure 23. Example of an agent uploaded to Google Pages with provided hyperlink.

3.3.2 Method

A total agent population of 145 agents (female = 69, and male = 76) were selected, based on the following criteria:

- Only coloured images were used.
- Very good quality images (approximately 7cm x 5cm).
- All images must have a standard white background.
- The image must focus on either: the face, the face and neck, or up to the shoulders. No images of a full body.

3.3.2.1 Participants

Participants (N = 545) were mainly students and staff within Manchester University; as well as family, work colleagues, friends and their acquaintances. Out of 545 participants

who took part in the survey, 62% were female (N = 338) and 38% (N = 207) were male. Participant demographics are expanded in detail in Table 8.

Table 8. Details of participant demographics.

Participant	Category	Total (N = 545)	%
Gender	*Male*	207	38
	Female	338	62
Age group	*Below 18*	8	1.5
	18-25	333	61.1
	26-35	150	27.5
	36-45	40	7.3
	46-55	10	1.8
	56-65	4	0.7
Ethnicity	*British White*	256	47
	British non White	30	5.5
	European White	77	14
	Asian Other	36	6.6
	South Asian	30	5.5
	White Other	13	2.4
	Other non White	103	19
Nationality	*UK*	333	61
	Non UK	212	39
Profession	*Academic/Student*	355	65
	Other	190	35

3.3.2.2 Procedure

The online surveys were advertised through the Manchester University mailing system. Each participant selected one out of the five unsupervised online surveys in order to evaluate the attractiveness and realism levels of 29 embodied agents. Each survey took the participant approximately 30 minutes to complete and submit.

Welcome to this survey which is part of my PhD research on Embodied Agents: which are visual characters used in computer applications.
Embodied agent communicate with the user using natural language and can act as virtual tutors, online advisors, sales people, virtual friends, as well as providing help with online transactions.
In this survey we ask you to evaluate the attractiveness and realism of 29 pictures of existing embodied agents.
The survey will take approximately 10-15 minutes to complete.
Please do no use the browser's 'back' button, but only the buttons situated at the bottom of each of the question pages.
Failure to do so may result in your responses being lost.
You can also use the 'full screen' option of your browser for better results.
This study has been approved by the Ethnical Committee of Manchester Business School: Approval Number – I07061303.
If you wish for more information please contact me at:
Rabia.Khan@postgrad.manchester.ac.uk

Figure 24. Introduction page given at the start of each survey.

Participants were given full instructions when opening the survey, as well as an introductory statement and summary of the background and purpose of the study (Figure 24).

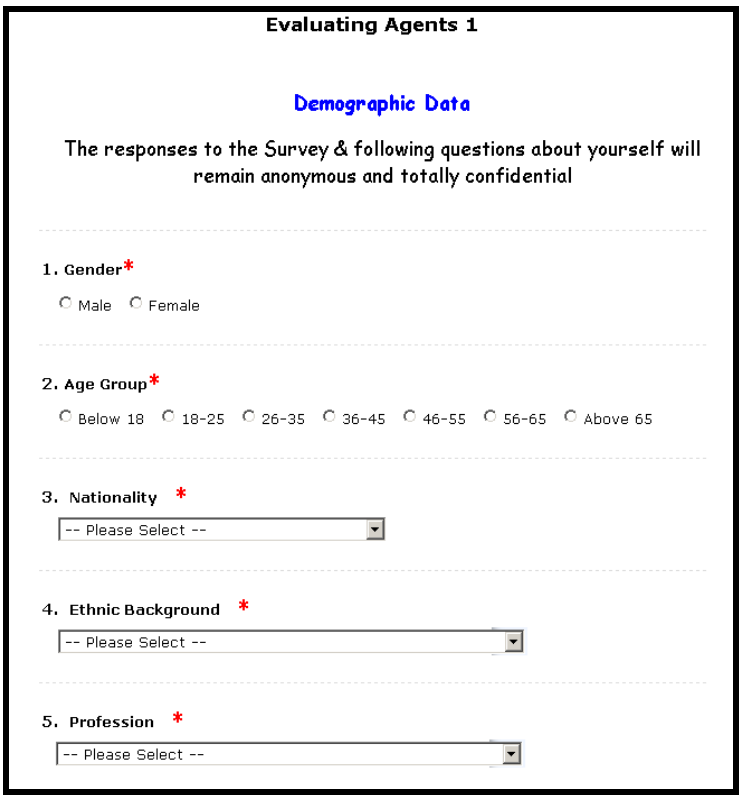


Figure 25. The demographic details page for participants.

The author's contact details were also provided for respondents to contact the author with any questions or queries. The second page in the survey consisted of closed questions inquiring about the demographic details of the participant, such as gender, age, and ethnicity (Figure 25). Participants were also reminded not to click on the back button within the browser as the results of their evaluation could be lost in this manner. Next, the participants were then presented with a set 29 images of embodied agents which appeared in random order.

Participants rated each image using a 7 point semantic-differential scale on their perception of the two traits: attractiveness and realism (Figure 26). In terms of attractiveness a rating of 1 meant that the embodied agent was perceived by the participant as unattractive, and a rating of 7 implied that this embodied agent was classed as attractive. For realism (i.e. anthropomorphism level of agent), again a rating of 1 implied that the embodied agent was perceived as unrealistic, and a rating of 7 meant that the agent in question was perceived as realistic. Once each agent was

evaluated, participants were thanked and invited to type any comments or suggestions about the online survey they had just completed.

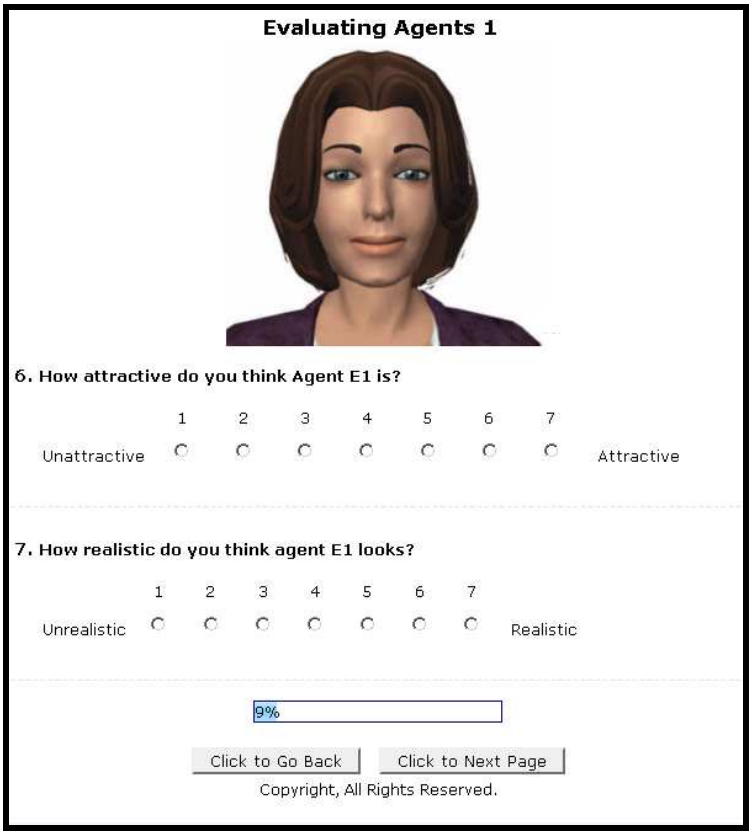


Figure 26. Rating the perceived attractiveness and realism levels of an embodied agent.

3.3.3 Results

The perceived attractiveness and realism mean ratings by 545 participants for each of the 145 agents (male = 76, and female = 69) were analysed using SPSS. Results are summarised into five sections: agent group ratings of perceived attractiveness and realism, group correlations for perceived attractiveness and realism, individual agent correlations for perceived attractiveness and realism, agents perceived as the most attractive, and participant's comments and suggestions.

3.3.3.1 Agent Group ratings of Perceived Attractiveness and Realism

Table 9 compares the perceived attractiveness ratings; whilst Table 10 compares the perceived realism levels for the various agent groups.

Table 9. Attractiveness values for agent group t-test comparisons (*M* = mean value; *SD* = standard deviation; *N* = number of agents; *NS* = not significant; *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$).

	Attractiveness			
	M	***SD***	***p***	***N***
Male Mannequin White All Ages	2.44	0.63		31
Female Mannequin White All Ages	3.16	0.88	***	31
Male Mannequin White Young Adult	2.48	0.65		19
Female Mannequin White Young Adult	3.11	0.93	*	25
Male Mannequin	2.51	0.64		37
Female Mannequin	3.2	0.83	***	37
Male Mannequin All Ethnic Young Adults	2.53	0.63		24
Female Mannequin All Ethnic Young Adults	3.16	0.86	**	31
Young Adults	2.96	.91	**	99
Older Age Groups	2.33	.76		26
Young Adults	2.96	.91	**	99
Other Age Groups	2.40	.74		46

The general trend was for female groups to be perceived by participants as being more attractive than male agents; and for young adult groups to be evaluated as more attractive than the other agent age groups.

Table 10. Realism values for agent group t-test comparisons (*M* = mean value; *SD* = standard deviation; *N* = number of agents; *NS* = not significant; *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$).

	Realism			
	M	***SD***	***p***	***N***
Male Mannequin White All Ages	2.88	.67	*NS*	31
Female Mannequin White All Ages	3.15	.57		31
Male Mannequin White Young Adult	2.74	.74	*NS*	19
Female Mannequin White Young Adult	3.05	.52		25
Male Mannequin	3.02	0.77	*NS*	37
Female Mannequin	3.2	0.59		37
Male Mannequin All Ethnic Young Adults	2.87	0.72	*NS*	24
Female Mannequin All Ethnic Young Adults	3.12	0.56		31
Young Adults	3.02	.98	*NS*	99
Older Age Groups	3.08	.99		26
Young Adults	3.02	.98	*NS*	99
Other Age Groups	2.69	.97		46
Male Agents	2.72	.99	*	76
Female Agents	3.12	.97		69
Male Young Adults	2.78	.93	*	42
Female Young Adults	3.19	.99		57
White Males	2.59	.73	**	63
White Females	3.13	.78		57
White Male Young Adults	2.59	.95	*	32
White Female Young Adults	3.2	.97		45

Female (M = 3.19, SD = .95) agents were perceived as being significantly more attractive than *male* (M = 2.46, SD = .74) agents, $t_{(143)} = 5.317$, $p < 0.001$. In terms of realism, *female* (M = 3.12, SD = .97) agents were also rated as being more realistic looking than their *male* (M = 2.72, SD = .99) agent counterparts, $t_{(143)} = 2.24$, $p < 0.05$ (see Figure 27).

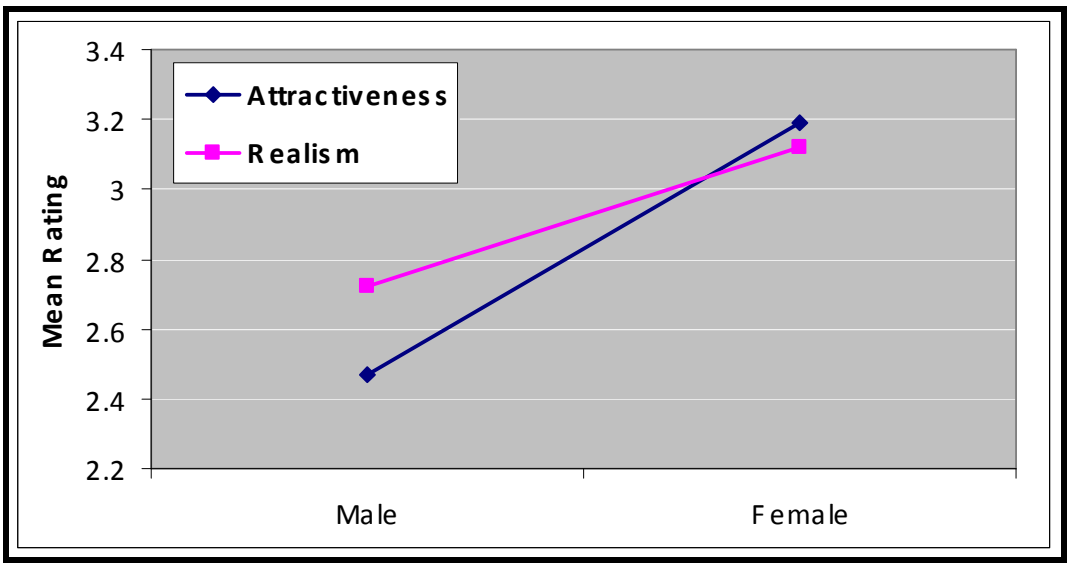


Figure 27. Attractiveness and realism ratings according to gender.

The perceived attractiveness for *female young adult*s (M = 3.23, SD = .95) was greater than that of *male young adult*s (M = 2.59, SD = .75), $t_{(97)} = 3.61$, $p < 0.001$. Furthermore, these *female young adult*s (M = 3.19, SD = .99) were also perceived as being more realistic than *male young adult* agents (M = 2.78, SD = .93), $t_{(97)} = 2.04$, $p < 0.05$.

Additionally, participants regarded *white female* (M = 3.22, SD = .94) agents as being more attractive than *white male* (M = 2.35, SD = .72) agents ($t_{(118)} = 5.58$, $p < 0.001$). Furthermore, *white male* agents (M = 2.59, SD = .73) were rated as less realistic than *white female* (M = 3.13, SD = .78) agents, $t_{(118)} = 2.67$, $p < 0.01$. The results indicated a significant difference in attractiveness between *white male young adult* (M = 2.49, SD = .77) agents and *white female young adult* (M = 3.28, SD = .99) agents, $t_{(75)} = 3.69$, $p < 0.001$. The *white female young adult* (M = 3.2, SD = .97) agents were also shown to be perceived as being significantly more realistic than the *white male young adult* (M = 2.59, SD = .95) agents, $t_{(75)} = 2.60$, $p < 0.05$.

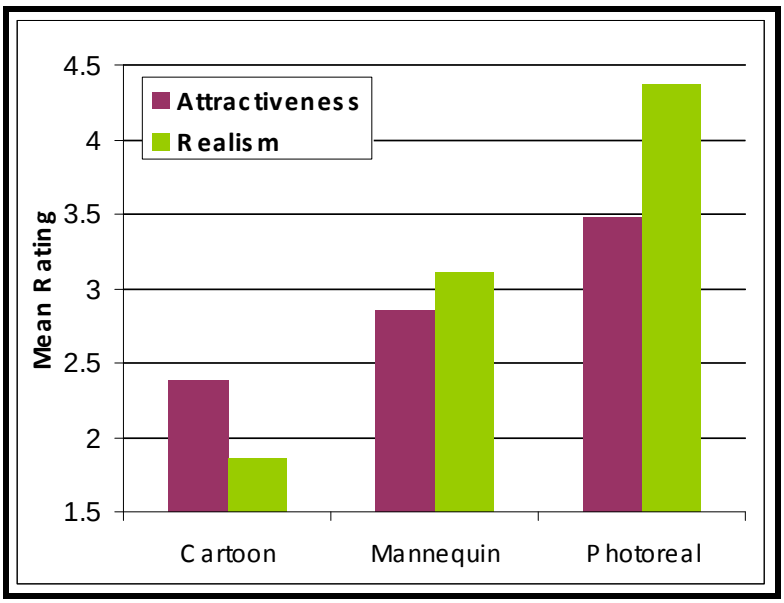


Figure 28. Mean ratings for all agents as a function of anthropomorphism.

ANOVAs revealed (Figure 28) a significant overall difference between the three anthropomorphic categories (*photo real*, *mannequin*, and *cartoon*-like) in terms of their attractiveness levels, $F_{(2,142)} = 14.11$, $p < 0.001$, partial $\eta^2 = .16$, and their realism levels, $F_{(2,142)} = 124.1$, $p < 0.001$, partial $\eta^2 = .65$. Participants perceived the *photo realistic* agents (M = 3.48, SD = .98) as the most attractive, followed by the *mannequin* looking agents (M = 2.86, SD = .81), and then *cartoon*-like agents (M = 2.39, SD = .72). The evaluation of how realistic the agents were perceived to be, also indicated the same order for the anthropomorphic categories; with the *photo real* (M = 4.37, SD = .81) agents being regarded as the most realistic, followed by the *mannequin* (M = 3.11, SD = .68) like agents, and then *cartoon*-like (M = 1.86, SD = .49) agents. In addition, post-hoc tests (Scheffe, LSD and Bonferroni) indicated a statistically significant difference between all three anthropomorphic groups in terms of their attractiveness and realism levels ($p < 0.05$ for all comparisons).

Subsequently, ANOVAs (Figure 29) yielded a significant overall difference in terms of perceived attractiveness between all three female anthropomorphic groups (*photo real*, *mannequin*, and *cartoon*-like), $F_{(2,66)} = 9.53$, $p < 0.001$, partial $\eta^2 = .42$; as well as a significant difference in relation to the realism levels of these three groups, $F_{(2,66)} = 70.61$, $p < 0.001$, partial $\eta^2 = .69$. *Photo realistic female* agents (M = 4.2, SD = .95) were rated as the most attractive anthropomorphic group, followed by *female mannequin* agents (M = 3.2, SD = .82), and *female cartoon*-like agents (M = 2.6, SD = .59). The ratings for realism revealed a similar pattern, whereby the *female cartoon*-like

agents (M = 1.86, SD = .45) were rated as the least realistic looking agents, followed by *female mannequin* agents (M = 3.2, SD = .59), and the *photo real* agents (M = 4.4, SD = .72) as the most realistic looking agents.

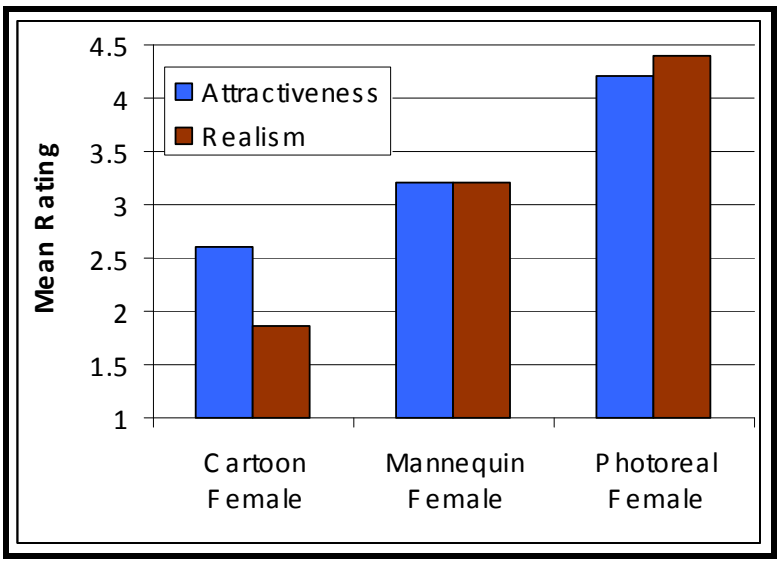


Figure 29. Mean ratings for female agents as a function of anthropomorphism.

Follow-up post-hoc tests (Scheffe, LSD and Bonferroni) were again performed on all three anthropomorphic categories of female agents, to reveal a significant difference *(p < 0.05 for all comparisons)* between these categories in terms of their perceived attractiveness and realism levels by the participants. However, ANOVAs on all three male anthropomorphic groups showed no significant difference, in terms of perceived attractiveness and realism levels, between them ($p > 0.05$).

Furthermore, a two-way ANOVA with anthropomorphism (3: cartoon, mannequin, and photo real) and gender (2: male and female) as between-subjects factors revealed a significant effect of embodiment ($F_{(2,139)} = 11.08$, $p < 0.001$, partial $\eta^2 = .14$) and gender ($F_{(1,139)} = 20.61$, $p < 0.001$, partial $\eta^2 = .13$) on the perceived attractiveness levels of the agents. Whilst, the same two way ANOVA on perceived realism levels, again showed a strong effect of embodiment, $F_{(2,139)} = 114.71$, $p < 0.001$, partial $\eta^2 = .68$, but no effect of gender, $p = 0.73$ (means and standard deviations for the agent groups in question are shown in Table 11).

Table 11. Attractiveness and realism mean values for the 2 way ANOVA (anthropomorphism x gender)
(*M* = mean value; *SD* = standard deviation; *N* = number of agents).

	Attractiveness		*Realism*		
	M	***SD***	***M***	***SD***	***N***
Male cartoon	2.24	.76	1.85	.51	30
Female cartoon	2.60	.59	1.86	.45	17
Male mannequin	2.51	.64	3.02	.77	37
Female mannequin	3.20	.82	3.21	.58	37
Male photo real	2.87	.85	4.34	.72	12
Female photo real	3.91	.96	4.40	.95	15

Consequently, an additional two-way ANOVA was not conducted, with gender (2: male and female) and age group (4: child, young adult, adult, and older adult) as between subject factors due to the low distribution of agents within the adult and older adult age groups (N < 6); as the major bulk of agents belonged to the young adult age group. Furthermore, a three-way ANOVA was also not carried out for the same reason, whereby gender (2: male and female), anthropomorphism (3: cartoon, mannequin, and photo real), and age group (4: child, young adult, adult, and older adult) were manipulated as between subject factor.

3.3.3.2 Correlations for Perceived Attractiveness and Realism

The relationship between the two variables: perceived attractiveness ratings and perceived realism ratings, were investigated using Pearson's product-moment correlation; whereby agents was the main unit of analysis. There was a strong positive correlation between all the agent's perceived attractiveness and realism levels, $r_{(145)} = .657$, $p < 0.001$, two tailed. Furthermore, female agent attractiveness and realism ratings were highly correlated, $r_{(69)} = .733$, $p < 0.001$. Additionally, a positive relationship, nearly as strong as that of female agents, was apparent between the attractive and realism levels of male agents $r_{(76)} = .571$, $p < 0.001$. Appendix C presents a number of tables highlighting the positive correlations for all agents within the five surveys, as well as between the various female and male sub-groups.

3.3.3.3 Individual Agent Correlations for Perceived Attractiveness and Realism

Individual agent correlations revealed significant ($p < 0.01$) and positive relationships between perceived agent attractiveness and realism levels (see Appendix C). A total of 36 male and female agents displayed very strong correlation values between perceived attractiveness and realism as a function of their ethnicity, age and anthropomorphism

level. Approximately 64% of these agents were females (N = 23), whilst only a modest 36% of male agents (N = 13) were observed with such strong correlations.

For both male and female agents, the number of highly correlated *white* agents outnumbered the number of highly correlated *black* agents. Whilst the vast majority of the top 36 highly correlated agents were from a *white* ethnic background (*white* female = 87%, and *white* male = 85%). Additionally, In terms of age, *young adults* were the dominating age group for both males (N = 11) and females (N = 21). Finally, *mannequin*-like agents were the most prevalent type of anthropomorphic category amongst these high correlating male (69%) and female (61%) agents.

3.3.3.4 Agents Perceived as the Most Attractive

The agents which were perceived by participants as the most attractive have been presented in Table 12. The top section of the table lists the top five rated females out of a total of 69 female agents, starting with the most attractive within the top line. All the top five attractive females were of the same age group (*young adults*), with the same ethnic background (*white*), as well as being classed as *photo realistic*. The more attractive a female agent was rated then the more realistic it was also perceived to be.

Table 12. The top five most attractive rated female and male agents.

Female					
Mean Att.	**Mean Real**	**Age**	**Ethnicity**	**Anthrop.**	**Role**
5.38	5.41	Young *Adult*	White	Photo Real	Actor
5.37	5.32	Young *Adult*	White	Photo Real	Storyteller
5.35	5.25	Young *Adult*	White	Photo Real	Ped. Agent
5.24	5.17	Young *Adult*	White	Photo Real	Presenter
5.21	5.14	Young *Adult*	White	Photo Real	Presenter
Male					
4.25	5.21	Young *Adult*	White	Photo Real	Actor
3.86	3.67	Young *Adult*	White	Mannequin	Ped. Agent
3.84	5.57	*Adult*	Oriental	Photo Real	Actor
3.84	3.06	Young *Adult*	Asian	Cartoon	Assistant
3.62	2.94	Young *Adult*	White	Mannequin	Ped. Agent

Regarding male agents, approximately four out of five were from the *young adult* age group, whilst the top two attractive male agents were from a *white* ethnic background, followed by an *asian*, o*riental* and then *white* looking agent. The top five male agents were not all *photo realistic*, as the second, fourth and fifth most attractive agents were classified as *mannequin*, *cartoon*-like, and *mannequin* respectively. The most attractive

male and female agents were: *young adults, white, photo real,* and were assigned the role of an actor by their designers.

3.3.3.5 Participant's Comments and Suggestions

Once the participants completed the evaluation of the agents within the online survey, they were then invited to write their own suggestions and opinions about the agents they had observed. Out of the 545 participants who evaluated the five agent surveys, almost a quarter of them (24%) wrote comments of their own. The following section discusses some of the participant's comments which have been divided into seven main topic areas:

Firstly, a large number of participants (N = 28) discussed which facial features made an agent appear attractive or unattractive. Approximately 53% of these participants commented on facial features that made the agent appear attractive; out of which 60% (N = 9) discussed the appearance of the eyes, and 40% (N = 6) commented on the importance of positive facial expressions such as a smile. For example:

- *I noticed the eyes first, and that made them appear attractive for me!* (eyes).
- *Big beautiful eyes look great on these agents* (eyes).
- *A nice smile helps them look attractive* (positive facial expression)
- *I think whether I found characters attractive had something to do with whether they were smiling or not* (positive facial expression).

The remaining 47% commented on facial features which made the agent appear unattractive; whereby 54% (N = 7) mentioned their discomfort in observing bald agents (mainly bald females), whilst 46% (N = 6) pointed to their dislike of negative facial expressions (i.e. unhappy looking face):

- *Your agents with hair look better than the bald ones. I found bald agents unattractive* (bald).
- *Never liked bald females-just wrong!* (bald).
- *The bald female looks weird, so rated her low on attractiveness* (bald).
- *I wasn't keen on the unhappy looking agents, and so rated them as unattractive* (negative facial expression).
- *The miserable looking agents weren't so nice to look at for me* (negative facial expression).

Secondly, 26 participants commented on which physical characteristics affected the agents' level of realism. Approximately 38% of these participants found that the eyes played an important role, another 27% mentioned the appearance of the skin (i.e. blemishes, freckles and skin tone), whilst 23% discussed the agent's hair, and the final 12% pointed to facial symmetry. For example:

- *On the most realistic agents, the most realistic thing about them are their eyes* (eyes).
- *The first think I looked at was the agents' eyes, and if they looked realistic then I rated that agent as more realistic* (eyes).
- *A face with blemishes, differing skin tones, and/or freckles gives a realistic image of a human face* (skin).
- *Agents with various skin shades, with or without spots and blemishes as in humans, I felt made them realistic* (skin).
- *If the face is too symmetrical, then naturally, that looks unrealistic* (facial symmetry).
- *As human faces aren't symmetrical, so if agent faces were not very symmetric, then it helped towards rating them as being more realistic* (facial symmetry).

Thirdly, a large number of participants (N = 22) commented on how they preferred attractive looking female agents, or their desire to view more nice looking female agents. Around 86% of these participants were male, and the remaining 24% female. Examples of some of their comments are as follows:

- *Not sure about male agents, but I love cute female agents* (male participant).
- *I would like to see more sexy executive women* (male participant).
- *Where are the babes? ;-) You don't expect that a real man is attracted by men faces do you? Real men want Catherine Jenkins! Yes!* (male participant).
- *I like females and found the females on here nice looking, did u make them?* (female participant).
- *I prefer female agents to males, and I came across some really good looking ones* (female participant).

Fourthly, 21 participants (males = 9, females = 12) also suggested other aspects that could be investigated by the experimenter:

- *Perhaps give the user chances to add comments to each individual picture as to why they like/dislike it. Or more in depth questions?* (male participant).
- *Why not ask us which agent we prefer or don't and why?* (male participant).
- *The more real looking people seem better for education* (female participant).
- *The cartoons could be ok in some situations, like for children/kids* (female participant).

Fifthly, a number (N = 11) of participants (female = 63%, male = 37%) pointed to how fearful they felt when they observed some of the agents. This is illustrated by some of the following comments:

- *Some of those would make me run for cover* (female participant).
- *The distorted faces scared me a little!* (female participant).
- *Some of them were petrifying!* (male participant).
- *Some of the more realistic agents were just plain scary* (male participant).

Additionally, 11 participants (male = 5, female = 6) highlighted their concerns due to the lack of agents that either represented them or other races and religious backgrounds:

- *I would like to suggest agents wearing head scarf, it will be great!!* (female participant).
- *None of the agents could be associated with my cultural background* (female participant).
- *How come most agents look European?* (male participant).
- *It would be nice to see agent of different races, most I see here are white* (male participant).

Finally, a total of 5 participants (male = 80%, female = 20%) made comments regarding their preference of attractiveness over realism. For example:

- *I'd choose attractiveness over looking realistic if I were to choose an agent* (male participant).
- *I think attractive agents are way better to communicate with than realistic ones* (male participant).
- *There is no point having realistic agents if they are ugly looking. They have to look good, otherwise forget it* (male participant).
- *I liked these pretty agents, and perhaps I would rather speak to them than to one that is very human like* (male participant).

- *My opinion is that having an attractive and less realistic agent online is better than having a realistic and unattractive agent* (female participant).

3.4 Conclusion

Findings from the census pointed to a trend by HCI designers to create more white young adult mannequins like agents; a high proportion of which were young looking female agents. This lead to the development of the online survey study examining how users perceived male and female agents, in terms of their attractiveness and realism levels.

Hence, the results from the survey study evinced that on the whole female embodied agents were perceived by the users to be significantly more attractive and realistic looking than male agents on the whole. Furthermore, t-tests showed that white females as well as younger adults were also rated as being more attractive than the other age groups. These findings confirm observations from the census study showing the tendency for designers to develop female agents that appeared to be generally more attractive and realistic looking than male agents.

The ANOVA's demonstrated significant differences between the three levels of agent anthropomorphism (i.e. cartoon, mannequin and photo real), showing that the more attractive an agent was rated then the more realistic it was also perceived to be. However, the difference in realism levels between the mannequin (M = 3.11, SD = .68) and photo real (M = 4.37, SD = .81) agents was not as large as was expected. It was initially assumed by the author that the photo realistic agents would be rated much higher; and a reason into why this was not so may be due to the information given within the first set if instructions (Figure 24) to the participants. These instructions described embodied agents as being 'visual characters used in computer applications.' This implied that these agents were artificial; so with this first impression in mind, the participants would have viewed the agents as being unreal, even when confronted with images of agents which could easily have been mistaken as actual humans.

Similarly, ANOVA's on female embodiment revealed that there was an effect of anthropomorphism; whereby the photo realistic females were perceived by participants as more attractive and realistic than the female mannequin and female cartoon like groups. Whilst, a further two-way ANOVA pointed to an effect of anthropomorphism and gender; showing that on the whole photorealistic females and males were rated as

being more attractive and realistic than the mannequin and cartoon like groups. Additionally, the gender effect shows that female agents are perceived as being more attractive and realistic than the male agents; which confirms the initial assumption in the census study where females were observed to be generally more attractive and realistic looking than male agents; as the emphasis by designers has so far been to create more attractive and realistic looking female agents.

Additionally individual agent correlations showed positive correlation values for female and male agents. Subsequently, it is no surprise that the five highest rated female agents in terms of their perceived attractiveness were all photo realistic, ethnically white from the young adult age group. HCI agent designers should commit their efforts into developing agents which do not constantly appease male desires; but focus on not just creating more attractive male agents but also more aesthetically appealing agents representing the non white and older age groups; which have unfortunately either been neglected or portrayed as unappealing for the user.

Approximately 28 participants commented on the large eyes and mouth which positively influenced their opinion of the agent's attractiveness. The agent's eyes were one of the main factors which contributed towards the participant's opinion of how attractive (53% of participants) and realistic (38% of participants) an agent was perceived to be. An agent which smiled was also a reason why 40% of participants found an agent to be attractive. Whilst 27% of participants discussed how the appearance of the agent's skin made it appear more realistic: such as blemishes, uneven skin tone; another 23 % found that the agent's hair played a role in making the agent look realistic, and a smaller group of participants (12%) mentioned the facial symmetry of an agent. Other participants (54%) pointed to their objection of interacting with bald female agents, which may reflect the fact that in daily interactions it is rare to come across bald or balding females; and usually it may be due to radiation or chemotherapy treatment for cancer.

A number of participants (N = 11) raised the issue of the lack of other ethnic and religious groups being presented by the embodied agents; as most were observed as being ethnically white from a prevalent European cultural background. Additionally, the census study and this online perception study have both highlighted the ethnocentric approach by designers. Consequently, HCI practitioners and designers need to be aware of the need for a more diverse range of agents representing numerous ethnicities,

cultures and religions; which will surely be a better and realistic portrayal of the world population, as well as aiding in reducing racist attitudes towards embodied agents. This is more poignant within multi cultural societies, such as in the UK.

The comments made by certain male participants (N = 22) complimenting the attractive female agents, as well as urging for more cute or sexy females, concurs with social psychology studies of male attitudes towards female attractiveness (Cross & Cross, 1971; Bar-Tal & Saxe, 1976; Feingold, 1992; Regan & Berscheid, 1997). These findings have shown that the physical attractiveness stereotype is more potent when applied to females than to males.

Some participants commented (N = 11) on feelings of fright when they evaluated some of these embodied agents. There are two plausible reasons for this fear: firstly, there actually were some rather bizarre looking agents within this study which could have elicited this reaction; and secondly studies (Mori, 1970) in the field of robotics have proposed a theory named the 'uncanny valley,' whereby the human observer feels a sense of revulsion towards the robot when it appears to be too human-like in appearance and behaviour. Regarding this study, the fear participants felt when observing agents which appeared too realistic, may have been due to the effect of the uncanny valley. HCI agent designers must keep this in mind when seeking to develop increasingly human-like agents.

To conclude, the comments and suggestions made by the participants clearly indicate the importance of the agent's visual cues (i.e. age, gender, ethnicity, attractiveness and realism levels), and the role these play in forming the participant's impressions of the agents in question.

4 Chapter 4: The Stranger-Based Attractiveness Stereotype Study

This chapter reports an experimental investigation into the role of the attractiveness stereotype using unfamiliar embodied agents within the stranger-base context. The study replicated Dion et al.'s (1972) experiment 'What is beautiful is good' in a controlled experimental setting. University students evaluated two sets of three static images of female embodied agents, varying in terms of their attractiveness levels (unattractive, average and attractive). The results demonstrated that, on first impressions, the embodied agents did elicit the attractiveness stereotype from the participant; thus highlighting the relationship between increased agent attractiveness with more positive agent evaluations. Participants' debrief interviews are also discussed giving an insight into how users perceive embodied agents.

4.1 Introduction

Results from the online perception study in Chapter 3 indicated that participants perceived white young adult females as the most attractive and realistic looking group of embodied agents. The present study extended these findings, by investigating the reliability of the attractiveness stereotype in the evaluation of embodied agents within the context of the stranger-attribution paradigm.

Social psychology studies have long established that physical attractiveness can elicit positive attributions as well as positive behaviour towards individuals when no interaction is taking place; i.e. within the stranger-attribution paradigm (Dion et al., 1972; Eagly et al., 1991; Feingold, 1992). Hence, the study reported in this chapter replicated, as closely as possible, a study on the attractiveness stereotype (Dion et al., 1972). The stimuli utilised by Dion and colleagues were images of real human faces, whilst the images employed in this attractiveness stereotype study were of embodied agent faces, selected from the evaluated agent images in Chapter 3. The selection criteria (described in section 4.2) utilised for the agent images in this experiment, resulted in a smaller range of male images to select from. Therefore, only female faces were tested in this study as the variance in attractiveness and realism levels of male faces was limited.

The investigation required participants to record their impressions of each agent face along 7 dimensions, which was conducted based on a 3 x 2 experimental design; with attractiveness being manipulated within-subjects, and agent-set as the control set-up. Two sets of debriefing interviews were also conducted with the aim of gaining an insight into how participants viewed agents

In order to determine the presence of the attractiveness stereotype, it was hypothesised that:

- H1: *Attractiveness will affect the initial impressions of agents. Hence, the more attractive the agent, the more positive the participant preference and evaluation of the agent.*

4.2 Stimuli

This study employed the use of two agent sets: agent set 1 (Table 13) and agent set 2 (Table 14); which were the control set-ups for the experiment. Both agent sets consisted of three female embodied agents, each of which represented one of the three levels of attractiveness: unattractive, average and attractive. These selected agents were previously rated for perceived attractiveness and realism levels by 545 independent evaluators in the online agent perception study (Chapter 3).

Table 13. Agent Set 1.

Unattractive	Average	Attractive
		
Agent 1 ***Attractiveness:*** *1.68* ***Realism:*** *1.98*	**Agent 2** ***Attractiveness:*** *3.09* ***Realism:*** *3.71*	**Agent 3** ***Attractiveness:*** *4.63* ***Realism:*** *4.39*

Table 14. Agent Set 2.

Unattractive	Average	Attractive
		
Agent A ***Attractiveness:*** *2.02* ***Realism:*** *2.39*	**Agent B** ***Attractiveness:*** *3.23* ***Realism:*** *3.04*	**Agent C** ***Attractiveness:*** *4.15* ***Realism:*** *3.37*

The following criteria were applied to narrow down the number of agents for selection in this attractiveness stereotype study:

- Faces from a white ethnic background.
- Faces from the young adult category.
- Faces from the mannequin style category.
- Faces representing the very attractive and the very unattractive target were not at the extreme end of the attractiveness distribution.

- Faces with neutral expressions.
- Neck and shoulders displayed within the image.

The rationale for selecting faces from the white, young adult and mannequin style categories was firstly that, these were perceived as being the most attractive set of agents, as well as having a more normal distribution of perceived attractiveness and realism values; in addition to being the most abundant group of embodied agents (Chapter 3). A large number of male agents were removed due to these selection criteria, as many male images (from the white, young adult and mannequin style category) only consisted of a face; with no neck and shoulders which was a requirement for this experiment. This lead to a skewed distribution of male agents which matched the selection criteria, in terms of their perceived attractiveness and realism levels, due to the small range of male agents to pick from.

Hence, female agents were then selected, as there was a greater number of female agents in comparison to male agents which passed the selection criteria, showing normal distribution; as well as being generally rated as more attractive than male agents (Chapter 3). The final selection of female agents was then divided into two agent sets: agent set 1 (Table 13) and agent set 2 (Table 14), which were the control set-ups; as no effect of agent set was observed. Here, a total of 15 randomly selected participants rated agent set 1, and the remaining 15 participants rated agent set 2. The following explains the reasoning behind the agent selection process and the procedure that was adhered to for this investigation.

In their seminal study by Dion et al. (1972), 50 yearbook pictures of Minnesota undergraduates were collected to be evaluated by fellow students on their physical attractiveness levels. A total of 12 images were selected for their experiment from the original 50, half female and half male students. These images were divided into four sets; two of the sets composed of female images and the remaining two sets of male students. Thus, each set consisted of three images; out of which one was rated as attractive, one average looking, and the third image rated as unattractive by the undergraduate students. For the main experiment, each participant rated and evaluated just 1 set of agents, whereby 30 participants rated the female images and the remaining 30 rated the male images. Additionally, the images Dion et al. (1972) opted for within the main experiment were not at the extreme ends of attractiveness or unattractiveness.

4.3 The Dependent Variables

The first dimension, *physical attractiveness*, was taken from the Interpersonal Interaction measures scale developed by McCroskey and McCain (1972), which have been widely utilised by researchers in communication and psychology (Wheeler & Kim, 1997). This scale consisted of a six 5-point Likert items:

- She is pretty
- She is somewhat ugly
- She is very sexy looking
- I find her attractive physically
- I don't like the way she looks
- She is not very good looking.

Perception of the embodied agents was measured on a 7-point semantic differential scale taken from Wheeler & Kim (1997):

- *Intellectual competence*: unintelligent – intelligent, emotional – rational, unambitious – ambitious
- *Social competence*: unsociable – sociable, unfriendly – friendly, introvert – extrovert
- *Social adjustment*: unstable – stable, immature – mature, poorly adjusted – well adjusted
- *Integrity*: dishonest – honest, untrustworthy – trustworthy, insincere – sincere
- *Potency*: weak – strong, unassertive – assertive, submissive – dominant.

A*nthropomorphism* was measured on a two-item, 5-point Likert scale adapted from Baylor and Ryu (2005), who initially used this measure to evaluate participant opinion of pedagogical agents:

- The agent is human-like
- The agent is machine-like.

4.4 Method

The study was based on a two-way 3 x 2 mixed ANOVA design, whereby attractiveness (3: unattractive, average, and attractive) was manipulated as the within-subjects factor whilst agent-set was the control set-up (2: agent set 1, and agent set 2).

4.4.1 Participants

The participants were mainly recruited via e-mail invitation. A total of 30 students at the University of Manchester (15 Male, and 15 Female) took part in the experiment, see Table 15.

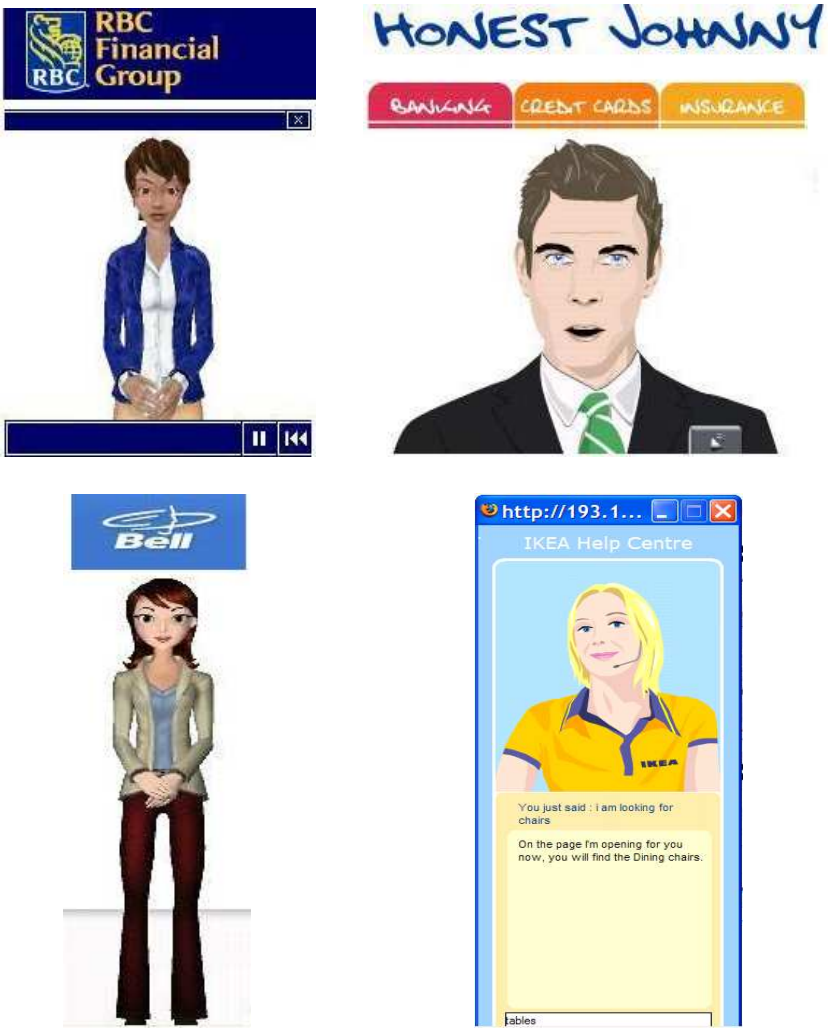
Table 15. Details of participant demographics.

Participant	Category	%
Gender	*Male*	50
	Female	50
Age group	*18-25*	36
	26-35	64
Ethnicity	*British White*	23
	Middle Eastern	20
	East Asian	10
	South East Asian	10
	Hispanic	10
	Other	27
Nationality	*UK*	27
	Non UK	73
Student Status	*Postgraduate*	87
	Post Doc	13

4.4.2 Procedure

The experiment was introduced to participants as a study looking into the user's opinion of embodied agents. The framing hypothesis was not discussed with the participants. These subjects were randomly divided into two groups of 15, whereby one group evaluated agent set 1 and the other group evaluated agent set 2. Before the experiment commenced, all participants were shown a 2 minute video to give them a better awareness and insight into what exactly embodied agents were.

(a) May (b) Honest Johnny



(c) Emily (d) Anna

Figure 30. The online embodied agent examples.

.

The video showed examples of four online agents (Figure 30) which were available on the internet:

a) May - the online bank advisor by the RBC Financial Group,

b) Honest Johnny – the online comparisons advisor,

c) Emily – an online assistant for Bell Canada, and

d) Anna – IKEA's online help assistant.

At the end of the video, a short interview took place, inviting the participant to provide comments and opinions of the agents they had just observed, and which agent they preferred. The entire debrief interview was audio recorded via Camtasia Studio.

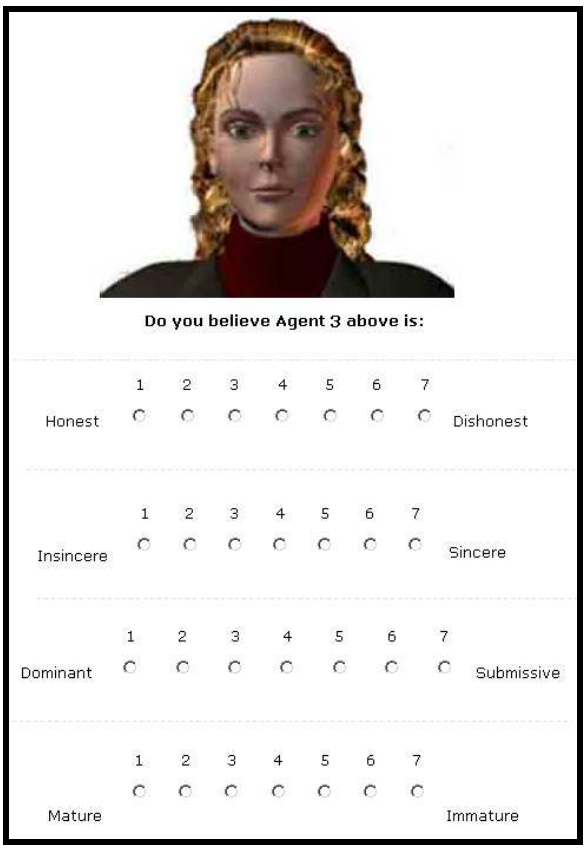


Figure 31. Online agent evaluation page.

Each participant was then shown one of the three pictures from the assigned agent set on a computer screen; and then invited to evaluate each one by filling an on-line questionnaire (Figure 31). The presentation order was randomised for each of the three pictures, which remained on screen whilst the participants evaluated each image. On completion, participants were presented with all three images of the agents they had evaluated and asked further questions in a second debrief interview, about their perceptions and opinions about the three agents in question, and agents in general.

4.5 Results

All seven scales revealed high reliability (Cronbach's Alpha > 0.80) for each of the three attractiveness levels: unattractive, average, and attractive. Seven indexes were then computed, providing mean scores on individual items for each level of attractiveness. These means were analysed as dependent variables in a 3 x 2 mixed-design ANOVAs. Hence, a general linear model with repeated measures was used. The partial eta-squared (η^2) was computed as an estimate of effect size: partial η^2 = .01 indicated small effects, partial η^2 = .06 medium effects, and partial η^2 = .14 large effects (Pallant, 2007).

4.5.1 Manipulation Check

The ANOVA on *physical attractiveness* indicated a very strong effect on perceived attractiveness ($F_{(2,56)}$ = 135.88, p < 0.001, partial η^2 = .83) of the agents; further showing a significant interaction between attractiveness and agent-set ($F_{(2,56)}$ = 12.29, p < 0.001, partial η^2 = .31). The interaction was a result of the unequal distribution of attractiveness levels between the two agent-sets (Figure 32).

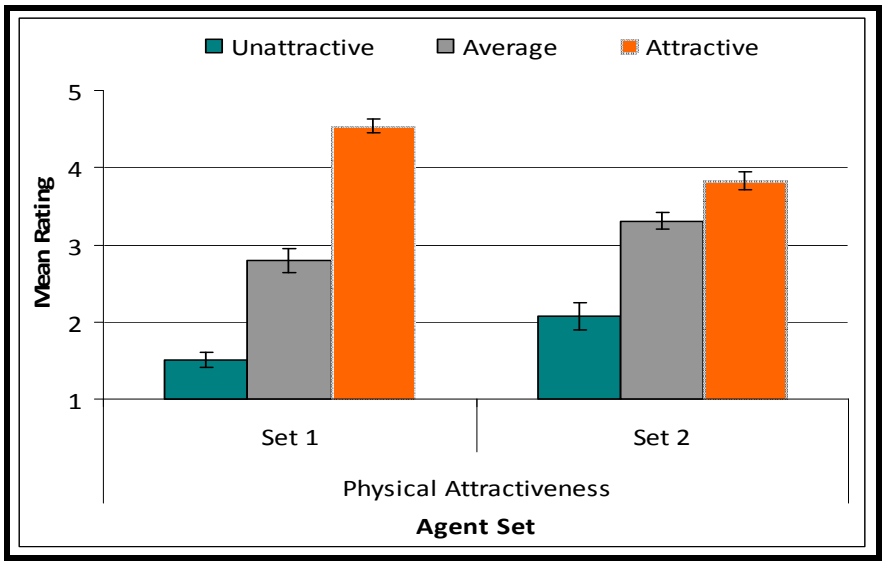


Figure 32. Physical attractiveness mean rating scores as a function of experimental conditions.

4.5.2 Test of Hypotheses

Similar patterns were confirmed for the evaluation of *intellectual competence* and *social adjustment* (Figure 33) by participants. ANOVAs for both of these dimensions revealed a main effect of attractiveness (in the order: $F_{(2,56)}$ = 51.99, p < 0.001, partial η^2 = .65;

$F_{(2,56)}$ = 45.52, p < 0.001, partial η^2 = .62); as well as a significant 2-way interaction between attractiveness and agent-set (in the order: $F_{(2,56)}$ = 7.22, p < 0.01, partial η^2 = .21; $F_{(2,56)}$ = 9.19, p < 0.001, partial η^2 = .25).

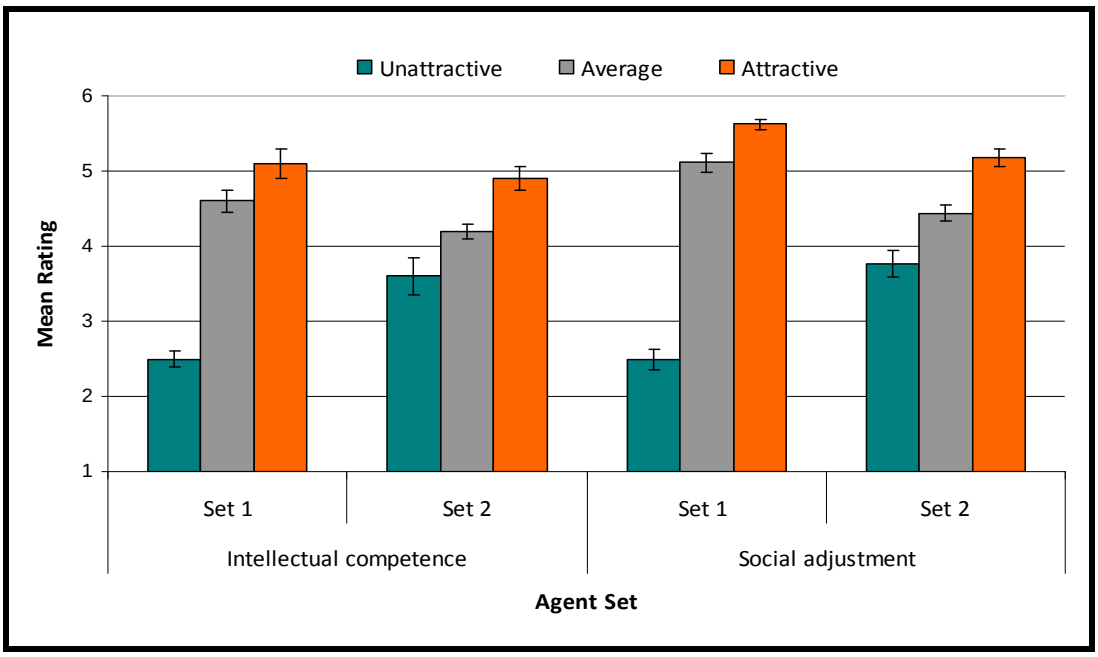


Figure 33. Intellectual competence and social adjustment mean rating scores as a function of experimental conditions.

The more attractive agents were systematically evaluated better within the *integrity* dimension, $F_{(2,56)}$ = 54.5, p < 0.001, partial η^2 = .66 (Figure 34). The difference between the two agent sets was reflected by the significant interaction between attractiveness and agent set, $F_{(2,56)}$ = 3.3, p < 0.05, partial η^2 = .12.

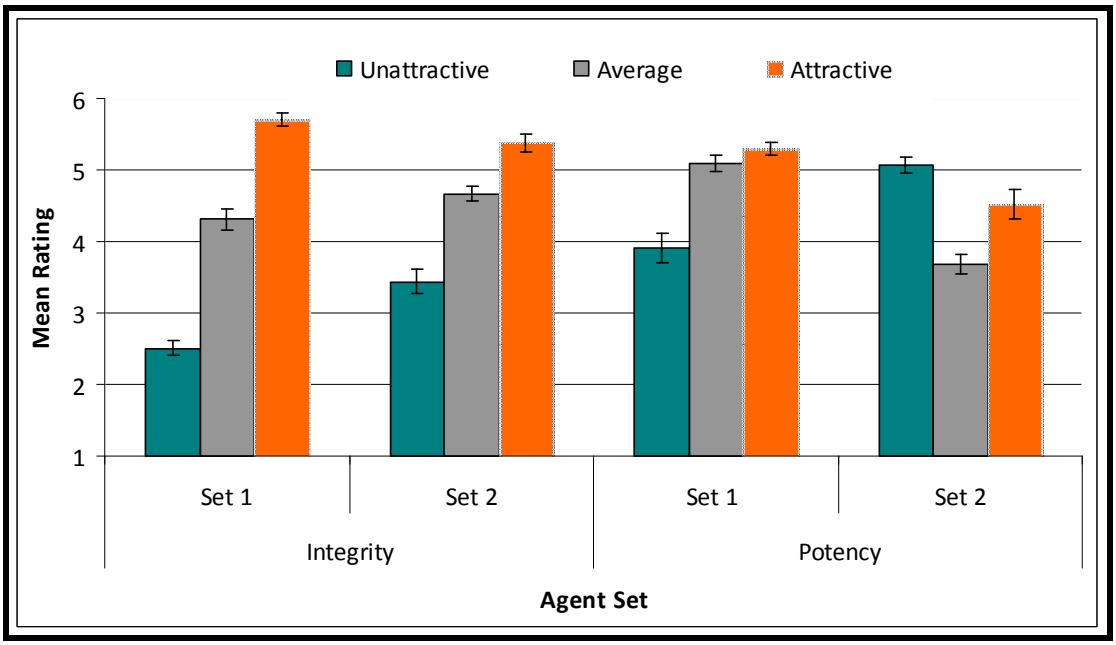


Figure 34. Integrity and potency mean rating scores as a function of experimental conditions.

Perception of *potency* (Figure 34) reflected a different trend, due to the considerable interaction effect between attractiveness and agent-set ($F_{(2,56)} = 7.52$, $p < 0.05$, partial $\eta^2 = .21$); whereby no effect of attractiveness was observed. Whilst, *social competence* (Figure 35) was strongly influenced by attractiveness, $F_{(2,56)} = 97.68$, $p < 0.001$, partial $\eta^2 = .77$; but showed no interaction effect.

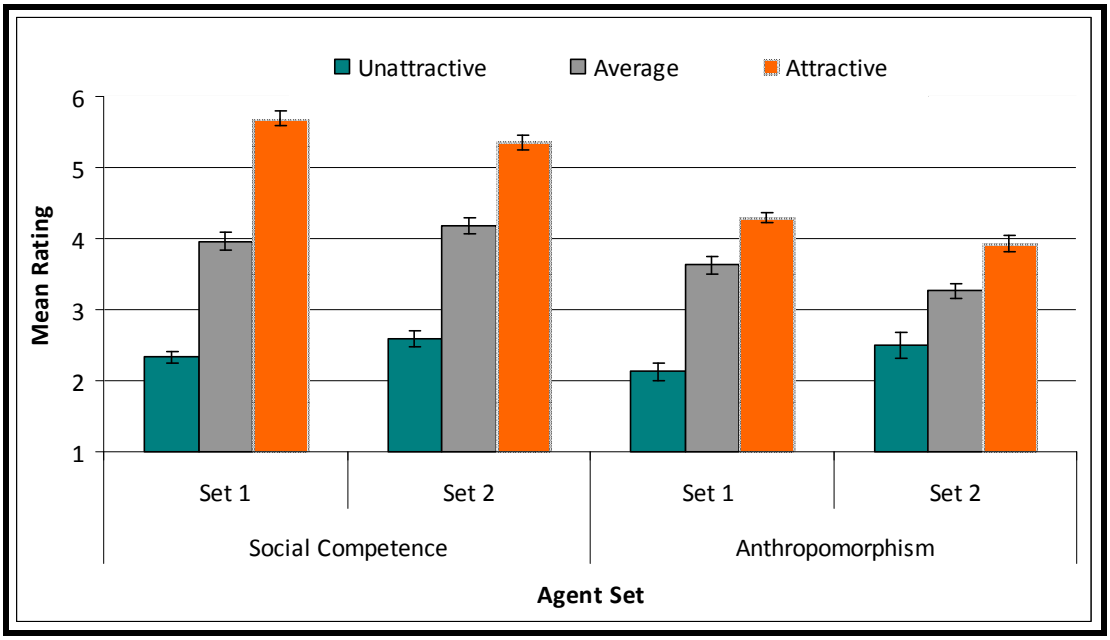


Figure 35. Social competence and anthropomorphism mean rating scores as a function of experimental conditions.

Finally, the evaluation of *anthropomorphism* (Figure 35) indicated a large main effect for attractiveness, $F_{(2,56)} = 41.59$, $p < 0.001$, partial $\eta^2 = .59$.

4.5.3 Interview 1

For the first interview within this study, participants were asked whether they had interacted with an online agent before; and secondly, to state, after viewing all four online agents from the video clip, which of the four online agents they would most prefer to interact with, stating the reasons behind their choice. The results indicated that only one participant had previous experience of interacting with an online agent, whilst approximately 80% of the participants expressed their preference for interacting with the online agent, Emily (Table 16).

Table 16. Frequency distribution of participants' agent preference.

Agent name	Participant preference
May	5
Johnny	0
Emily	24
Anna	1
Total	**30**

4.5.3.1 May (Figure 30, a)

May the online bank advisor was the next popular agent to be selected by 16% of participants. She was generally viewed as being pretty but not as cute or attractive as Emily. A number of participants (N = 9) stated how they liked the way the web site zoomed into this agent; and her realistic hand gestures, whereby her head and body movement were synchronised together with the blinking whilst she spoke. Additionally, participants (N = 12) appreciated the fact that May seemed to be making eye contact with the user, which was also the case with Emily. However, although May's gestures were slightly mechanical; they were considered more natural by 13 participants in comparison to Anna.

4.5.3.2 Honest Johnny (Figure 30, b)

Approximately 90% of the participants perceived Johnny as being rather deceitful for numerous reasons. The first being his whole appearance; i.e. the suit and slick hair style, reminded many of a manipulative salesman who would do his best to try to sell you unreliable goods, as well as giving you inaccurate and misleading advice. Participants (N = 20) questioned the reasoning as to why Johnny was called Honest; as they viewed this phrase to be rather deceptive. Additional annoying features stated by 86% of participants were the very rigid body postures, and his harsh voice which sounded extremely rough, arrogant and forceful, with no coherence with his lip movement. A total of 6 participants found Johnny quite intimidating and even frightening; whereas 20 participants pointed out that the absence of most of the top half of his body made them feel quite uncomfortable and suspicious of him. This is in contrast to the minimum view of at least half of the body for May, Emily and Anna which 90% of participants preferred.

4.5.3.3 Emily (Figure 30, c)

The main reason for 80% of participants selecting Emily as the agent they would most like to interact with was first her physical appearance and secondly her voice and gestures. Emily was described as being the most attractive (by 93% of participants) as well as being stated by participants as the most human-like due to her full body embodiment (56%). One participant described Emily's appearance liking to that of a Japanese Manga character. Emily's voice was also considered by 70% of participants as being the most appealing due to it being soft and life-like, which allowed her to come across as the most friendliest character in comparison to the other three agents, according to 80% of the participants. Many participants (N = 17) commented on how her arm gestures were very natural, including her blinking and head movement. A plus point regarding Emily was that her lips moved in a synchronised manner to her voice (N = 12 participants), and the hand gestures Emily used to point to the different areas she was discussing of the web page were considered to be very useful (N = 16 participants). On the whole Emily received the most positive feedback from the participants. The only down side to Emily's physical appearance, as stated by a single female participant, was due to her body evidently being too skinny and looking rather disproportional like a Barbie doll.

4.5.3.4 Anna (Figure 30, d)

The general consensus was that 97% of participants did not feel Anna was an advantage, nor did she add anything to the IKEA website. Her role was perceived by these participants as being just a body placed on top of a search engine. She was viewed quite negatively by 22 participants due to her speech being extremely tedious and slow. Approximately 87% of participants referred to Anna as being quite robotic, as her head and shoulders remained static with the occasional blinking eye movement. When Anna spoke, her lips were not synchronised with her stilted, unpleasant and monotonous voice, as stated by 83% of participants. Additionally, 8 participants noticed that her eyes were not looking towards them but slightly to the side in another direction. Furthermore, 90% of participants felt surprised that such a poor quality agent was being utilised by a successful and well known company such as IKEA, and her flaws were viewed as quite irritating by these same participants. Only one male participant perceived Anna in a positive light, as he felt she was ideal for novice online users; due

to her simplicity and ease of use, as she was a task focused agent who, in his opinion, did her job well.

4.5.4 Debrief Interview

Once the participants had completed the evaluations for the three agents, they were then asked two main questions as follows:

1) Which of the three agents would you prefer to interact with and why or why not?
2) Are there any suggestions and comments you would like to make about agents in general?

This section is divided into two phases: Interview for Agent Set 1 - This first section discussed the answers to question 1 by participants who evaluated agent set 1; Interview for Agent Set 2 - The second section reports on the answers (for question 1) given by participants rating agent set 2. The answers for question 2 have been incorporated within the design guidelines section in the conclusion chapter (Chapter 7).

4.5.4.1 Interview for Agent Set 1

Table 17 illustrates how the bulk of participants (73%) preferred to interact with Agent 3, with a handful choosing Agent 2 (27%), whilst no participants were interested in selecting Agent 1.

Table 17. **Frequency distribution of participants' agent preference in agent set 1.**

Agent name	Participant preference
Agent 1	0
Agent 2	4
Agent 3	11
Total	**15**

Unattractive Agent 1 – The general sentiment felt amongst all 15 participants was that Agent 1 appeared emotionally and mentally unstable, besides being quite unattractive. Many participants (60%) commented on Agent 1's abnormal cross-eyed look, whereby she did not appear to be making real eye contact, as she seemed to be in a world of her own. Not a single participant (N = 0) felt comfortable with the idea of interacting with this agent; in fact some participants (N = 6) felt rather intimidated and frightened by her. She was also described by 47% of participants as being a crazy psychopath who

could be capable of anything; especially with her half smile which was perceived as being insincere due to her psychotic thoughts.

Average Agent 2 - This agent was preferred by a handful of participants (N = 4) who viewed her as non-threatening, and a mature mother-like figure. One of these participants stated how this agent reminded her of the mother in a television series called 'Malcolm in the middle.' Whilst a number of participants (N = 6) felt she resembled a 1960's movie character, partially due to her hair style and clothing. However, 67% of participants perceived her as being a strict and even an angry looking female who could easily fit the role of a manager or head mistress. It was also pointed out by 40% of participants that she appeared as though she was a very bitter female, who could even be close to having a mental breakdown; and ideally would be suited in a role for an advert on anger management.

Attractive Agent 3 – Approximately 11 participants stated that they preferred to interact with Agent 3. She was regarded by these participants as being not only physically appealing, but also more approachable, confident, organised and intellectual in comparison to Agent 2 and 3. Her overall appearance was also viewed by some of them (N = 5) as professional and business-like. A number of the participants (N = 6) described her as being softer and more pleasing to the eye, as she met their expectations of what an attractive female agent should resemble. Another observation that was appreciated by 80% of participants is how human-like Agent 3 was in comparison to the other two agents. Furthermore, Agent 3 was considered to be the friendliest agent whom users would enjoy interacting with. A few female participants also stated how they would select agent 3 to help them as an online fashion and beauty advisor, due to Agent 3's well groomed hair, skilfully applied make-up and good dress sense.

4.5.4.2 Interview for Agent Set 2

The debrief interviews demonstrated a strong preference from 67% of the participants for Agent C; whilst a smaller number (27%) selected Agent B. However, only one participant chose Agent A (Table 18).

Table 18. Frequency distribution of participants' agent preference in agent set 2.

Agent name	Participant preference
Agent A	1
Agent B	4
Agent C	10
Total	**15**

Unattractive Agent A – There were numerous reasons as to why the vast majority of participants (N = 14) selected Agent A as the least desired agent. Many sensed (N = 12) that she was a very tough as well as a strong and dominating character, who could easily become aggressive. A number of participants (N = 7) stated that she would be ideal as a police woman or as a member of the elite armed forces unit; whilst others (N = 5) could picture her as a professional wrestler or female bouncer at a night club. Another opinion was that this agent appeared as though she would try to force and intimidate the online user to buy a product online, which was pointed out by 60% of the participants, due to her masculine features. Hence, she was viewed as the least feminine out of the three agents, mainly due to her wide neck and her strange round shaped nose, together with thick lips which were not proportional to the rest of her face. The only exception to these views was one male participant who considered her to be quite friendly and non-intimidating; mainly as she did not fit the stereotypical attractive female category.

Average Agent B – A handful of participants (N = 4) selected Agent B, referring to her neutral facial expression as a representation of her calm demeanour which put them at ease; together with the added bonus of possessing a pretty face and a cute hair style. However, other participants were not so charitable; although most agreed (93%) that she was prettier than Agent A, a small number of participants (N = 3) still perceived agent B's soft features and neutral facial expression as a signal of her timid and insecure personality.

Attractive Agent C – A large portion of participants (N = 10) favoured Agent C, as they regarded her well defined and proportional facial features as the most realistic and attractive. Her eyes not only made her more realistic, as stated by 8 participants; but they seemed to portray a sense of calm, as well as a confident, professional and assertive temperament. This gave 80% of participants the impression that she was also a

caring and welcoming personality. She was also perceived as being approachable, as participants (N=10) felt they would feel comfortable to talk to her, or ask her questions.

4.6 Conclusion

This study provided strong support for the hypothesis that the attractiveness stereotype applies to embodied agents, and the effect of this stereotype is strong in first impressions; within the stranger attribution paradigm (Dion et al., 1972; Eagly et al., 1991; Feingold, 1992). Additionally, the results demonstrated large effects for *social competence, intellectual competence* and *social adjustment.* The *social competence* dimension revealed the greatest effect size not only within this study, but also in previous social psychology studies (Eagly et al., 1991; Feingold, 1992); highlighting important traits such as being sociable and popular. Accordingly, the more attractive an individual or embodied agent is perceived to be then the friendlier, more sociable and likeable that individual or agent is also deemed to be in human-human or human-agent interaction.

The *anthropomorphism* dimension illustrated how realistic or unrealistic an agent was perceived to be by the participants; whereby the more attractive agents were perceived by participants to be more realistic. This finding links with the strong positive correlations found in Chapter 3 between perceived attractiveness and realism levels for all the agents ($r_{(145)} = .657$, $p < 0.001$).

The only exception to previous studies on the attractiveness stereotype was the dimension of *potency*. It is reasonable to believe that perhaps the components of *potency* within the human-agent context (i.e. dominance, assertiveness and strength), may have been associated as negative traits rather than positive ones. This is supported by the comments made by participants within the debrief interviews which suggested difficulties in understanding and rating this dimension. For example, in agent set 1; the attractive Agent 3 was rated the highest on the *potency* scale, as participants viewed her as being confident and dominant, but in the positive sense. The average looking Agent 2 was also rated highly on the *potency* scale (close to Agent 3), but this was mainly due to her appearing too strong and assertive, in the sense of a mother figure or head mistress who was to be feared (as stated by 67% of participants). Therefore, Agent 2 and Agent 3 were rated highly in terms of the *potency* dimension, which was either perceived as a positive trait or a negative one. In the same way, unattractive Agent A (within agent set

2) was rated the highest in terms of *potency*, as she was viewed by participants (N = 12) as being aggressive and dominant looking which was regarded as a negative trait as they did not feel comfortable with the thought of interacting with this intimidating looking agent. Thus, the lack of an effect on *potency* seems to be due to the variation in the agent's appearance analysed in this study.

This concurs with studies on real human beings showing that participants from Eastern or collectivist cultures tend not to perceive attractive targets as so high on the *potency* scale as the North American participants (Wheeler & Kim, 1997). The participants within this attractiveness stereotype study were evenly split between Europeans and people from Eastern countries (Table 15) who may pay more attention to collectivistic values.

The debrief interviews point to a general theme, in that the main decision factor for participants as to whether they would prefer an embodied agent or not is the agent's level of attractiveness. The results of the interviews concur with the quantitative evaluation of the agents. Hence, the comments suggest that the participants appreciated and preferred to interact with the most attractive agent, which was also evaluated more positively in terms of its personality traits in comparison to the less attractive agents. This highlights the importance of agent aesthetics, as the participants stated their appreciation of an aesthetically pleasing agent. Consequently, agent characteristics such as synchronised speech with its mouth, a smile, a full or at least half-body representation, and hand gestures are all added bonuses in improving the quality, believability and effectiveness of human-agent interaction. Furthermore, a fundamental facial feature which participants take great notice of are the eyes, which play a vital role in increasing the agent's level of realism as well as perceived attractiveness.

To conclude, the findings of this study provide evidence to substantiate H1; including further evidence to support the CASA model and Media Equation (Nass et al., 1995, Reeves & Nass, 1996), in that observers will apply more positive personality traits towards the more attractive agent; in the same way humans associate greater positive personality traits towards attractive humans before any interaction has taken place (Dion et al., 1972; Eagly et al., 1991; Feingold, 1992). However, two dimensions in particular: *potency* and *integrity*, appear to have different strengths and meaning for the observer, when applied to either agents or humans.

5 Chapter 5: The Interaction-Based Attractiveness Stereotype Study

This chapter reports an empirical study consisting of two separate investigations into the effects of the attractiveness stereotype in human-agent interaction. The first investigation required participants to evaluate embodied agents before and after chatting to them. The results demonstrated the strong effect of the attractiveness stereotype both before and after this interaction; whereby more positive personality traits were attributed by the participants to the attractive agents than to the unattractive agents. The second investigation analysed the chat-bot agent-participant conversations; again revealing significantly more positive behaviour by the participants towards the attractive agents. Additionally, a four step coding system was developed in order to analyse these conversations.

5.1 Introduction

This chapter reports two experimental investigations which expand on the results from the previous attractiveness stereotype study (Chapter 4), by addressing the reliability of the attractiveness stereotype within an interaction-based context (Langlois et al., 2000). Results from the online perception study (Chapter 3) indicated that female agents were perceived as being more attractive and realistic than male agents, which lead to the first attractiveness stereotype study (Chapter 4); highlighting the presence of the attractiveness stereotype within the stranger-based context (Dion et al., 1972; Eagly et al., 1991; Feingold, 1992).

The aim of this study was to address the reliability and presence of the attractiveness stereotype in a more ecological and interaction-based setting; whereby participants were invited to engage in a spontaneous conversation with a female chat-bot agent for approximately 10 minutes. Evaluations were conducted before and after interaction with the embodied chat-bot agent, to investigate the strength of the attractiveness stereotype. Consequently, given the very strong impact of attractiveness on first impressions by embodied agents evinced in Chapter 4, and following the findings from the interaction-based literature (Langlois et al., 2000); it was hypothesised that:

- H1: *Attractive agents would be evaluated more positively over unattractive agents, both before and after interaction with the chat-bot agent.*
- H2: *The effect to be weaker after interaction as participants acquired a great amount of contextualised information in order to inform their initial evaluation.*
- H3: *Participants may view the agents as being less realistic after interaction, as the agent may not meet their high expectations raised at first impressions.*
- H4: *Participants' behaviour towards the attractive chat-bot agents would be more positive and welcoming than towards the unattractive agents during the chat sessions.*

5.2 Materials

The stimuli were initially developed using an online software product called SitePal. This software allowed the creation of a personalised speaking character by using one of their numerous models, which could be customised according to the developer's needs using the 'scene editor' (Figure 36). The agent's physical appearance could also be

modified within the scene editor; such as their hairstyle, eye-ware, clothing, jewellery; as well as skin, hair, mouth and eye colour; including attributes such as the size of the agent's face, nose, mouth and shoulders. Within Sitepal, text to speech audio could also be added to these characters, whereby the agent is pre-programmed to say what the developers and designers require.

Figure 36. The Sitepal scene editor.

Once the final four agent stimuli were selected from the pilot study for this experiment (as explained in the following Stimuli section); the agents were then further developed to be used as talking 'chat-bots.'

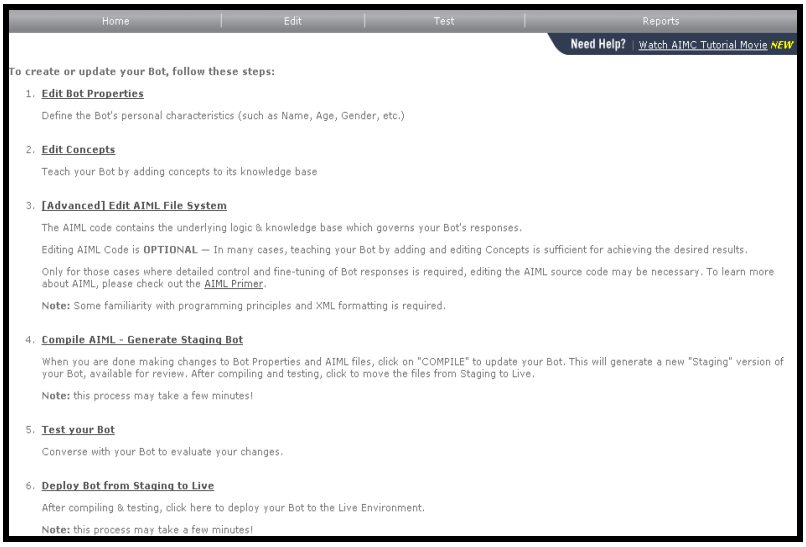


Figure 37. Oddcast's Artificial Intelligence Management Centre (AIMC) tool page.

These talking chat-bot agents were designed using another tool within Sitepal called Oddcast. Hence, Oddcast's Artificial Intelligence Management Center (AIMC) tool (Figure 37) allowed the developer to:

- Define the chat-bot's personal properties or characteristics; such as name, age, gender, favourite book or film, etc. (Figure 38).

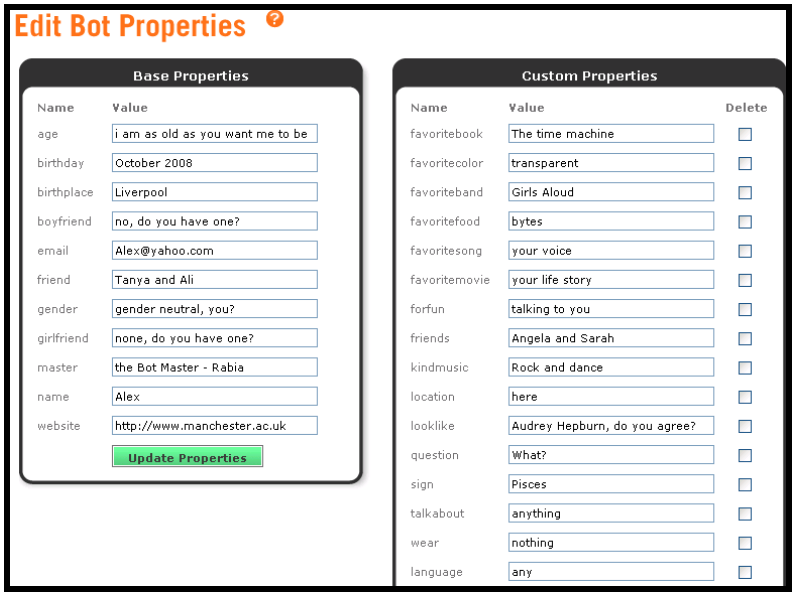


Figure 38. The chat-bot properties editing page.

- Teach the chat-bot agent, by using the 'edit concept page' (Figure 39), detailed knowledge about specific subject matter, as well as how to respond to user questions with context-sensitive, spoken answers; by looking for specific keywords which can elicit one or more programmed responses.

Figure 39. The concept editing page for the chat-bot program.

Finally, all conversations between the chat-bot agent and user were saved as log reports which could be retrieved from the reports page (Figure 40) and viewed for analysis within Microsoft Excel.

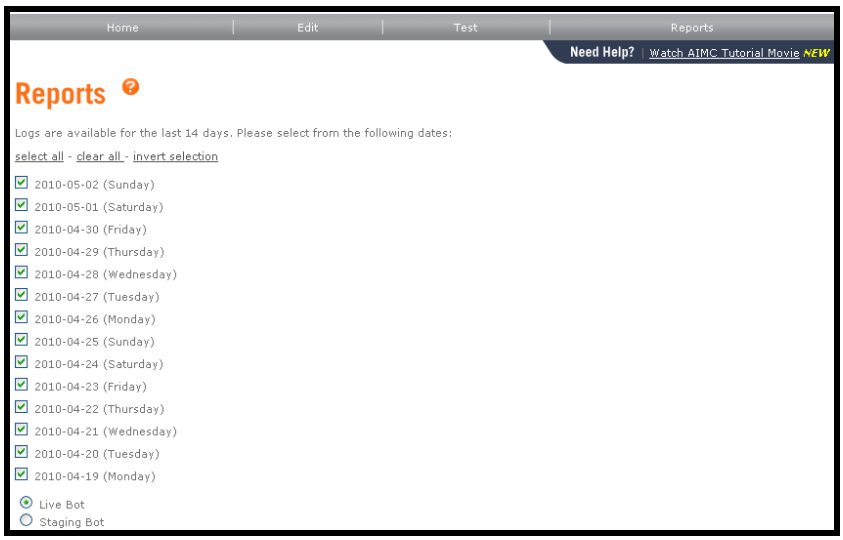


Figure 40. The reports page listing all the conversation logs.

5.2.2. Stimuli

Images of male agent stimuli were not utilised within this study as it was more difficult to manipulate male attractiveness in comparison to female attractiveness. According to previous studies, there tends to be greater emphasis on female beauty, as reflected in numerous social psychology studies which have demonstrated how the attractiveness stereotype is more potent when applied to females during social interaction than to males (Cross & Cross, 1971; Berscheid & Walster, 1974; Bar-Tal & Saxe, 1976; Feingold, 1992).

Initially, a pilot study was conducted with 20 agents developed (see Appendix D), using SitePal; consisting of three main agent groups: young white females (labelled as Set 1 and Set 2), young black female agents (Set 3 and Set 4), and older white female agents (Set 5 and Set 6). Each of these groups comprised of at least one agent representing each of the three levels of physical attractiveness: attractive, average and unattractive. The criteria used to systematically manipulate the three varying levels of attractiveness, was based on literature investigating facial attractiveness (Cunningham, 1986; Rhodes, 2006) as illustrated in Table 19.

Table 19. Criteria for development of attractive, average and unattractive agents.

Physical Change	Attractive agent	Average Agent	Unattractive Agent
Nose	Proportional to face	Thinned by 25%	Widened by 50%
Lips	Full	Thinned by 25%	Thinned by 25%
Head & shoulders	Symmetrical - head and shoulders proportional to each other	Asymmetrical -head widened by 25%. Shoulders reduced by 25% (from original agent). Large head, small body	Asymmetrical - head width reduced by 16.5%. Head height reduced by 25%. Shoulder width increased by 30%

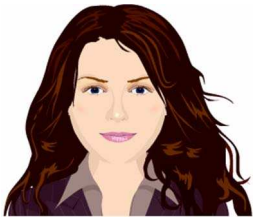
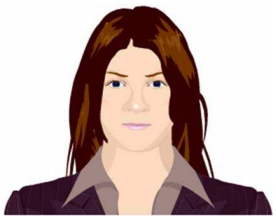
These agents were then rated by 58 participants in terms of their attractiveness and realism levels, on a 7-point semantic-differential scale, via an online survey using SurveyGizmo. The results showed that the mean difference (see Appendix D for mean values) between the unattractive, average and attractive agents in the older white females groups was less than 1 (Set 5 and Set 6); whereas the mean difference for attractiveness ratings between all three attractiveness levels for both the white and black young female agent groups was greater than 1 (Set 1-4).

Furthermore, in terms of selecting agents for this investigation, the average looking agents were not chosen due to a number of these agents being rated highly in terms of their attractiveness levels; resulting in a low mean difference between the average and attractive agent (mean difference < 1.0) within that agent set. Hence, the decision was made to select the most and least attractive white female and black female agents, which also showed a greater attractiveness rating distribution (mean difference > 2.0) between the most and least attractive agent within that agent set. It was important to select the most and least attractive agent from the same set, in order to keep all other factors constant (i.e. hair colour, eye colour and clothing) except for the attractiveness levels. Subsequently, four agents were then selected: two from the white young female agent set 2 (one attractive and the other unattractive), and the other two from the black young female agent set 3 (again one attractive and the other unattractive).

Table 20 illustrates the final four stimuli, which also represent the four experimental conditions that the participants (in groups of 12) were randomly assigned to:

a) Attractive young white female agent
b) Unattractive young white female agent
c) Attractive young black female agent
d) Unattractive young black female agent.

Table 20. The final set of static attractive and unattractive agent images.

(a)	(b)
	
Attractiveness: 4.98 ***Realism: 4.78***	***Attractiveness: 2.86*** ***Realism: 3.31***
(c)	**(d)**
	
Attractiveness: 4.76 ***Realism: 4.28***	***Attractiveness: 2.76*** ***Realism: 3.28***

A gender neutral name called ‘Alex’ was added to the agents to control for users who may have negative preconceptions regarding gender, before volunteering to participate in the experiment.

5.3 Method

A 2 x 2 x 2 mixed ANOVA design was employed. Agents’ ethnicity (2: white vs. black) and attractiveness (2: attractive vs. unattractive) was manipulated as between-subject factors; whilst time (2: time 1 vs. time 2) was a within-subject factor; where time 1 represents time before participant interaction (chatting) with the chat-bot agent and time 2 after interaction with the chat-bot agent.

5.3.1 Participants

The subject population consisted of 48 (27 male; 21 female) students at the University of Manchester. These participants were mainly recruited via word of mouth and e-mail invitation. Participant demographics are expanded in Table 21.

Table 21. Details of participant demographics.

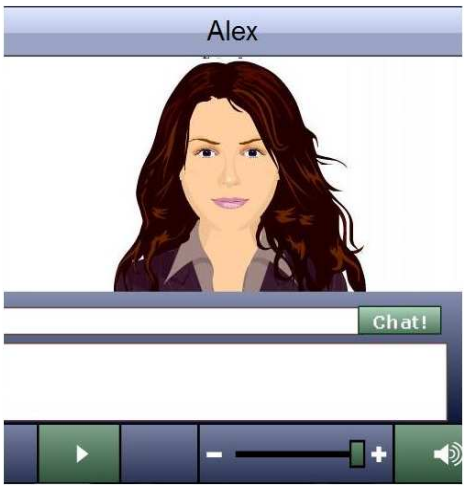
Participant	Category	%
Gender	*Male*	56
	Female	44
Age group	*18-25*	15
	26-35	85
Ethnicity	*British White*	14
	Middle Eastern	20
	East Asian	6
	South East Asian	8
	South Asian	16
	Black Other	12
	Hispanic	8
	Persian	6
	Other	6
Nationality	*UK*	33
	Non UK	67
Student Status	*Post Doc*	10
	Postgraduate	75
	Undergraduate	15

5.3.2 Procedure

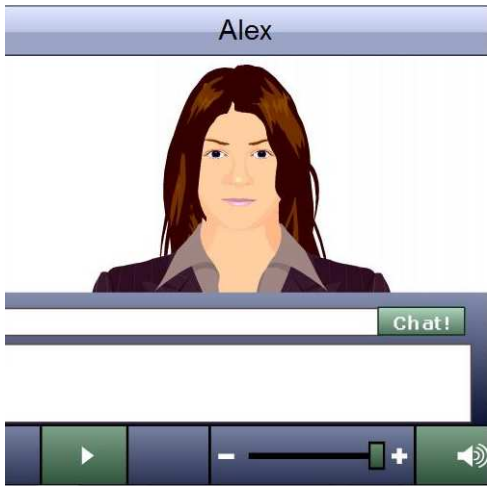
Participants were first introduced to the study as an investigation into how users perceive embodied agents. Subsequently, participants were instructed to evaluate a static image of one of the four target stimuli (Table 20) using the same array of instruments (consisting of seven dimensions: *physical attractiveness, social competence, integrity, social adjustment, intellectual competence, potency* and *anthropomorphism*) employed in Chapter 4. They were left alone in the laboratory and invited to chat to the allocated chat-bot agent on any topic they wished for approximately 10 minutes. The participants typed their inputs into a conversation window (see Table 22) directly below the chat-bot agent; whereas the agent communicated with the participant by speech. Once the chat session was complete, the participants were requested to evaluate the agent image again; using the same array of instruments before interaction.

Table 22. The four female embodied chat-bot agents.

(a) Attractive young white chat-bot agent

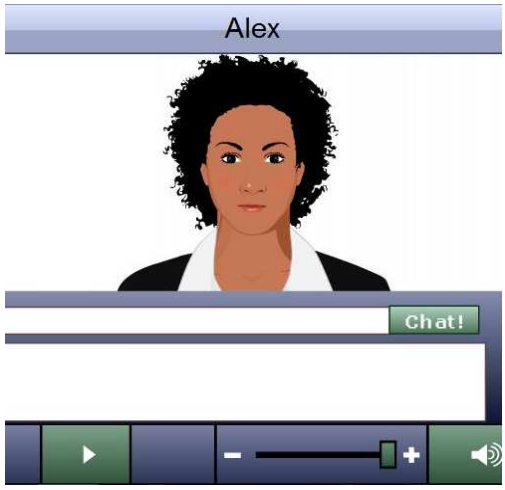


(b) Unattractive young white chat-bot agent



(c) Attractive young black chat-bot agent

(d) Unattractive young black chat-bot agent



5.4 Results

The results are summarised into two main sections: the first section investigates the participant evaluation of the agents, and the second section conducts a conversational analysis of the conversation logs between the four agents and the 48 assigned participants.

5.4.1 Agent Evaluation

All 7 dimensions measuring the pre- and post-test evaluations revealed satisfactory results (Cronbach alpha > 0.80). Mean scores were computed and entered for 7 mixed-design ANOVAs with ethnicity (2) and attractiveness (2) as between-subjects factors

and time (2) as a within-subjects factor. Hence, a general linear model with repeated measures was used.

5.4.1.1 Manipulation Check

A single large main effect for agent attractiveness, $F_{(1,44)} = 46.23$, $p < 0.001$, partial $\eta^2 = .51$, was returned for the ANOVA on the physical attractiveness dimension; therefore supporting the reliability of the manipulation. Furthermore, there was a large mean difference which remained constant between the attractive and unattractive agents for time 1 and time 2 (mean difference = 1.15); as participants showed little change in their perception of the attractiveness levels of the agent after interaction (Figure 41).

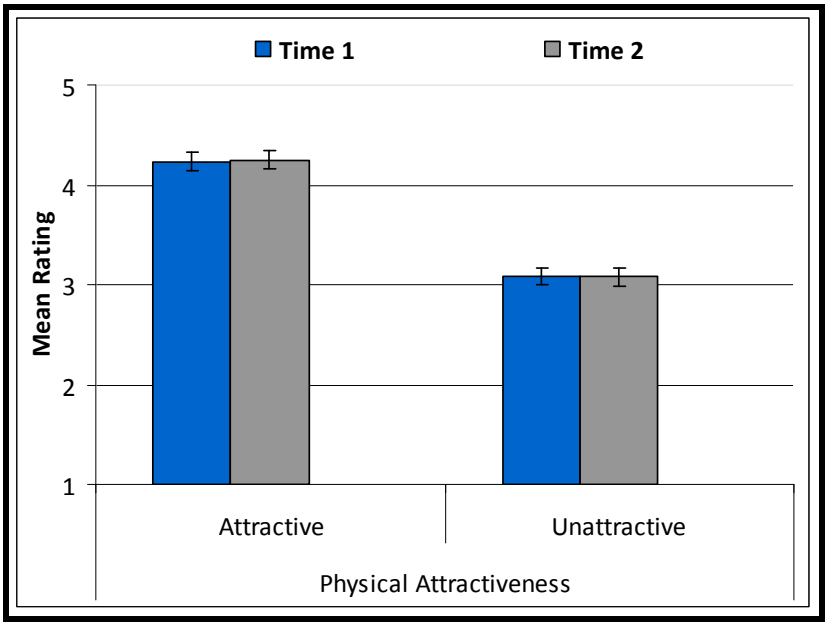


Figure 41. Physical attractiveness ratings as a function of experimental conditions.

5.4.1.2 Test of Hypotheses

An ANOVA on *social competence* revealed a large main effect of attractiveness ($F_{(1,44)} = 48.38$, $p < 0.001$, partial $\eta^2 = .52$) and evaluation time ($F_{(1,44)} = 23.79$, $p < 0.001$, partial $\eta^2 = .35$), with a significant interaction between attractiveness and time, $F_{(1,44)} = 6.63$, $p < 0.001$, partial $\eta^2 = .13$ (Figure 42). Although the evaluation of the attractive agent improved after interaction (mean difference = 0.29); the increase was more larger with the unattractive agent (mean difference = .95).

A similar trend was observed for *integrity* (Figure 42); which returned a significant main effect for attractiveness ($F_{(1,44)} = 9.76$, $p < 0.01$, partial $\eta^2 = .18$) and evaluation time ($F_{(1,44)} = 16.25$, $p < 0.001$, partial $\eta^2 = .27$). Both attractive and unattractive agents improved with time for these two dimensions, and more so for the unattractive agent. It is apparent, that for both of these dimensions (*social competence* and *integrity*),

participants evaluated the attractive agent better than the unattractive agent, and this evaluation improved after the interaction.

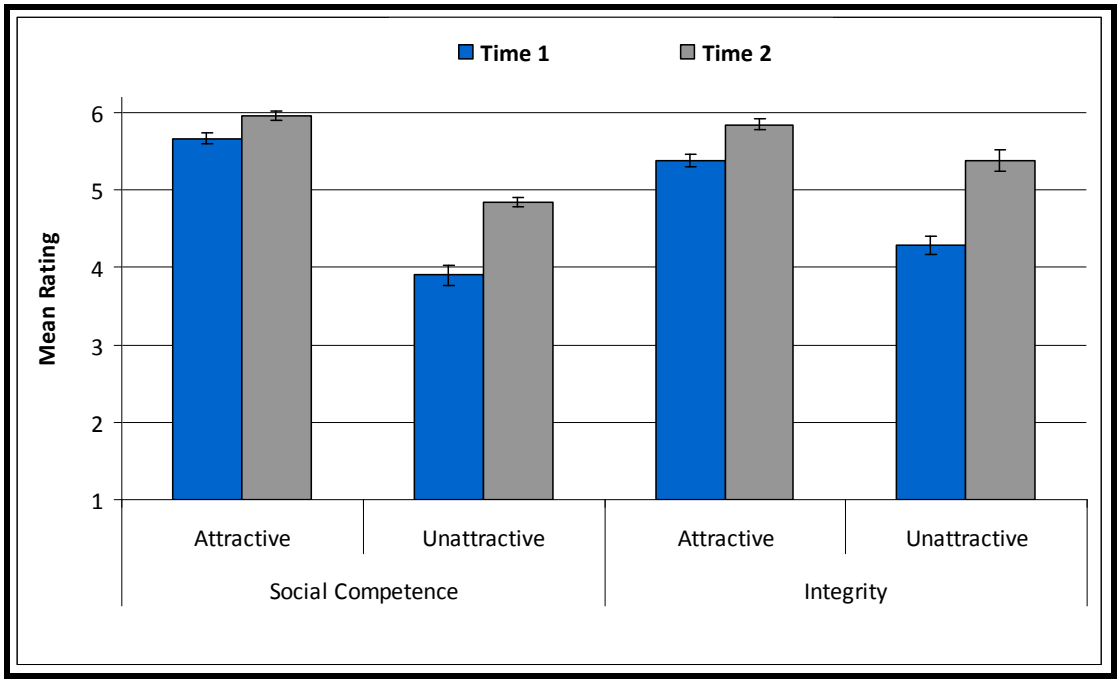


Figure 42. Social competence and integrity ratings as a function of experimental conditions.

The analysis of *social adjustment* (Figure 43) returned significant effects for attractiveness ($F_{(1,44)}$ = 18.45, p < 0.001, partial η^2 = .29) and time ($F_{(1,44)}$ = 8.64, p < 0.05, partial η^2 = .16). Similarly, *intellectual competence* also pointed to a significant effect of attractiveness, $F_{(1,44)}$ = 10.14, p < 0.01, partial η^2 = .19, and time, $F_{(1,44)}$ = 12.49, p < 0.01, partial η^2 = .16 (Figure 43).

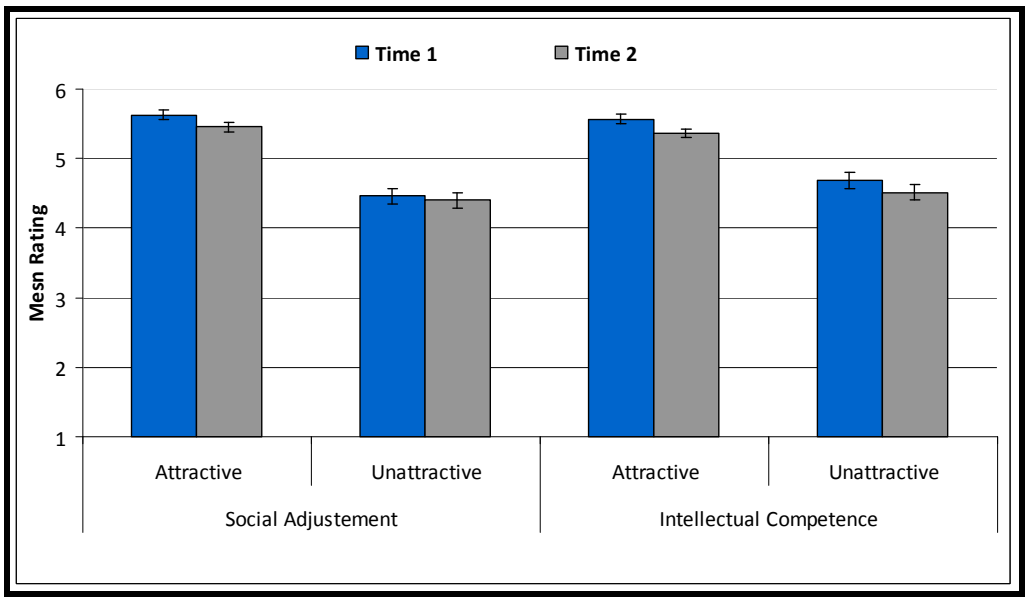


Figure 43. Social adjustment and intellectual competence ratings as a function of experimental conditions.

Finally, the *anthropomorphism* (Figure 44) dimension also revealed a significant effect of attractiveness, $F_{(1,44)} = 16.79$, $p < 0.001$, partial $\eta^2 = .28$, and time, $F_{(1,44)} = 152.69$, $p < 0.001$, partial $\eta^2 = .77$.

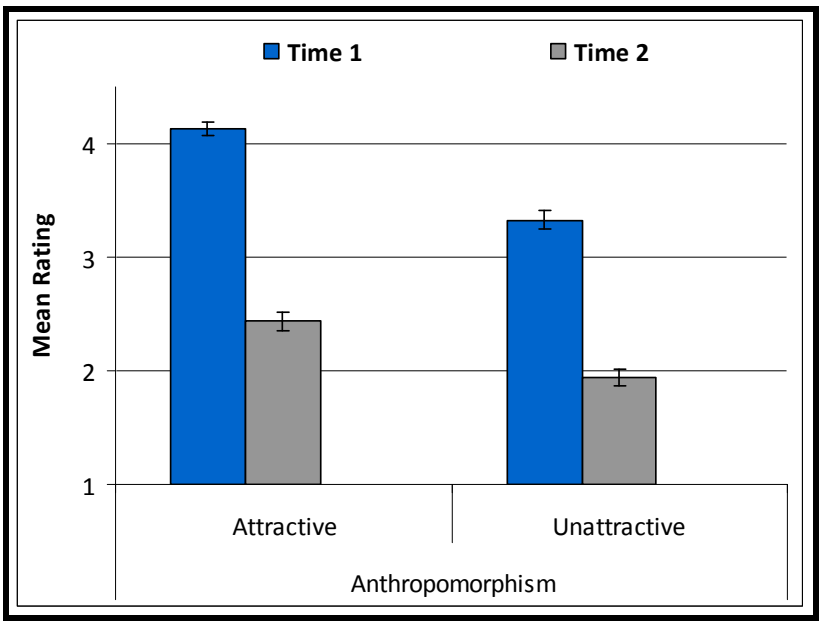


Figure 44. Anthropomorphism ratings as a function of experimental conditions.

The evaluations for *social adjustment*, *intellectual competence* and *anthropomorphism*; all significantly decreased after interaction, and this decline was more pronounced for *anthropomorphism* (Figure 44), as evinced by the higher mean difference for the attractive (mean difference = 1.7) and unattractive (mean difference = 1.4) agents.

The ANOVA on potency did not show a main effect for ethnicity ($F_{(1,44)} = 7.54$, $p < 0.01$, partial $\eta^2 = .14$) and a marginally significant interaction effect between ethnicity and attractiveness ($F_{(1,44)} = 3.74$, $p = 0.06$, partial $\eta^2 = .08$). This interaction may be a result of two factors: firstly, the unattractive black agent scored the highest scores in terms of potency, and secondly the unattractive white agent received the lowest potency scores by participants both before and after interaction. In summary; participants systematically evaluated the attractive agents more positively than the unattractive agents in all dimensions.

5.5 Conversational Analysis

The next section investigated how the interaction between the participant and embodied chat-bot agent influenced the participant's behaviour towards the agent. Participants were left unsupervised in the computer laboratory to chat with one of the four assigned

agents called Alex (Table 22) for approximately 10 minutes. The 48 participants were randomly assigned, in groups of 12, to chat with one of the four chat-bot agents.

5.5.1 Method

The corpus of conversations between the 48 participants and the four agents were automatically logged and saved within the Artificial Intelligence Management Centre (AIMC) tool (Figure 37). The entire data corpus was read in order to code for the participants' emotions when interacting with the agent, by exploring emerging patterns in order to finalise and develop a coding system to categorise the conversations.

The coding system was adapted from the method used by Veletsianos, Scharber, & Doering (2008), who used a comparative method to create salient categories and patterns from the logs, by noting down the patterns that were observed; which were then compiled and reanalysed until consensus was reached between all the authors on the salient patterns.

5.5.1.1 Coding Method of Analysis

A four step coding system was therefore devised to analyse the participants' emotions towards the four chat-bot agents. Each turn made by the participant, whilst chatting with the agent, was coded a total of four times.

The first category identified the **syntactic direction**, as to whether the input was a *question, response* or *statement*:

1. *Questions*: e.g. How are you? / Where are you from?
2. *Response*: e.g. I am single/ Off course I do / Yes I think so.
3. *Statements*: Included salutations, valedictions as well as commands; e.g. Hello / Goodbye / Thanks / Get lost! / Do as I say!

If a participant asked a question and statement within the same input, then the question would be coded separately to the statement. For example, the comment 'hello there, how are you?' would be divided into two separate turns or inputs. Hence, 'hello there' would be categorised as a statement, and then 'how are you?' would be marked as a question.

The second code was labelled as the **valency**. This highlighted the input as being either a positive, negative or a neutral comment:

1. *Positive* comments are not negative in any way but are taken within context where there is positive emotion behind the input. These can include likes, loves,

compliments, praise, and even flirtation: e.g. I like you / That was clever / I love your hair / You are funny / Hahahaha / Will you marry me?

2. *Negative* comments have negative emotion behind them which can consist of commands, insults and offensive remarks, expletives, including those that are sexual in nature and even racist: e.g. Shut up! / Do as I say! / You are so ugly / Are you dumb or what! / You are talking rubbish again / I hate you / I want sex.

3. *Neutral* comments have neither positive nor negative emotion behind them, and can comprise of statements and general conversations regarding relationships, beliefs, daily life, as well as testing the agent's knowledge/reasoning and understanding: e.g. Hello / How old are you? / I have a pet dog / I am a student / She went to Madrid / Who created you? / Do you believe in God? / Do you sleep? / What is the capital of France? / What is 10 + 1000?

The next category was labelled as the **pragmatic discourse**, and determined the discourse function of the input by the participant which consisted of eight sets of codes:

1. Asking for personal information (*API*) – This is where participants asked the agent for personal information about itself, such as the agent's likes, dislikes, relationships, and personal statistics (i.e. weight, height): Do you like pizza? / How are you? / How old are you? / Do you have family?
2. Asking for general information *(AGI)* – When participants would ask the agent for information that was not directly linked to it, then this was classed as AGI. For example: What does that mean? / What is the weather like? / How can I find a job in Manchester?
3. Giving a personal opinion *(GPO)* – Here, participants gave their own opinion regarding their personal feelings, thoughts, likes, dislikes and even laughter: I want a girl friend / I like reading / You look great / I love you / hahaha.
4. Self disclosure *(SD)* – When participants discussed personal information about themselves, such as their life, age, height, job, education, clothing etc. This category is the opposite of API. For example: I am male / I am just 24 / I just booked a holiday / I am wearing blue jeans / I am from Canada.
5. Giving general information *(GGI)* – This category is the opposite of AGI, as the participant gave details which were not directly linked to their personal lives; such as: It's snowing / Paella is from Spain / The university has many departments / That means hello in French.

6. *Testing* - Participant input which intended to test the knowledge, understanding, or capability of the agent. For example: What is 6 + 78? / Why is the sky blue? / What is the capital of Ireland? / What does WHO stand for?
7. *Command* – This is where the participant gave an order to the agent to either do something or to stop doing something. For example: Shut up! / Get lost! / Write my essay!
8. Small Talk *(ST)* – Most of the ST were usually short statements made by the participant, including salutations and valedictions. For example: Hello / Bye bye / See you / So / Hmmm / Thanks.

The fourth and final coding category was labelled as **topic** which represented the subject as well as the type of discussion displayed by the participant's questions or responses. This category comprised of eleven codes:

1. *Relationships* – When participants discussed any area regarding relationships in general or relating to the agent or their own family, friends, girl/boy friends, marriage, husband/wife, etc. For example: I have 2 sisters / I live with my 3 best friends / I'm single / Are you married?
2. *Humour* – This category could be interpreted as the participant laughing at the agent, telling the agent a joke, or turning something into a joke in a friendly manner. For example: Hahaha / Lol / You're funny! / Knock Knock!
3. *Challenge* – When the participant was confused, shocked, disagreed with, or unimpressed by what the agent said; including trying to prove that humans are better than chat-bots or robots. For example: What do you mean? / That's not an answer! / That seemed a little flawed / How dare you! / You are just a stupid robot.
4. Religious/moral/creation *(RMC)* – This is where the participant discussed any issues dealing with faith, God, as well as how they were created (i.e. by the designer). For example: Do you believe in god? / I am an atheist / Who created you? / God created me / Do you know who Jesus is?
5. *Compliment* – Where the participant gave positive comments about the agent's appearance, knowledge or ability. For example: You are gorgeous / I love your hair / You're quite clever.
6. *Flirtation* – When participants tried flirting with the agent without being offensive or sexually explicit. For example: Will you be my girlfriend / will you marry me? / I really want your number.

7. *Offensive* – Participant inputs which were either unpleasant or sarcastic, including commanding the agent. Why is your nose so big? / Your hair is a mess / Go and find out you dumb idiot / Get lost! / Shut up!
8. *Expletive* – This is where swear words were used by the participant which could be directly aimed at the agent, or just mentioned in general discussion. For example: That was Sh*t / He is a di*k head / Fu*k you / Bi*ch!
9. *Sexual* – Any topic of discussion by the participant of a sexual nature. For example: What's an orgasm? / Can you have sex? / will you sleep with me? / I love sex.
10. *Racist* – A deeply insulting slur regarding an individual's race/ethnicity; such as: You P*k*!
11. *Conversational* – Any input that did not fit into any of the 10 preceding categories were classed as conversational, including testing the agent, as well as salutations and valedictions. For example: Do you like shopping? / What is your favourite food? / I am working now / What is the capital of France? / I just booked a holiday / it's really cold outside / Oh really? / Hi / Yes / Ok / Bye bye.

Examples to help illustrate how participant inputs were coded using the above method are as follows:

Input 1 - How are you?

1. Syntactic direction: *Question.*
2. Valency: *Neutral.*
3. Pragmatic discourse: *Ask for personal information (API)*
4. Topic: *Conversational.*

Input 2 - I hate you

1. Syntactic direction: *Response.*
2. Valency: *Negative.*
3. Pragmatic discourse: *Give personal opinion (GPO).*
4. Topic: *Offensive.*

Input 3 - Who was Stalin?

1. Syntactic direction: *Question*
2. Valency*: Neutral.*
3. Pragmatic discourse: *Test.*
4. Topic: *Conversational.*

Input 4 – I am single.

1. Syntactic direction: *Response.*
2. Valency: *Neutral.*
3. Pragmatic discourse: *Self disclosure (SD)*
4. Topic: *Relationships.*

Input 5 – That does not make sense.

1. Syntactic direction: *Response.*
2. Valency: *Negative.*
3. Pragmatic discourse: *Give personal opinion (GPO).* .
4. Topic: *Challenge.*

Input 6 – Shut up!

1. Syntactic direction: *Statement.*
2. Valency: *Negative.*
3. Pragmatic discourse: *Command.*
4. Topic: *Offensive.*

Input 7 - Do you play any sports?

5. Syntactic direction: *Question.*
6. Valency: *Neutral.*
7. Pragmatic discourse: *Ask for personal information (API)*
8. Topic: *Conversational.*

Input 8 – Hahaha / LOL!

9. Syntactic direction: *Response.*
10. Valency: *Positive.*
11. Pragmatic discourse: *Give personal opinion (GPO)*
12. Topic: *Humour.*

Table 23. Summary of the coding system and variables used in the conversational analysis.

1	2	3	4
Syntactic direction	**Valency**	**Pragmatic Discourse**	**Topic**
Question	*Positive*	*Ask for personal information*	*Relationships*
Response	*Negative*	*Ask for general information*	*Religious/Moral/Creation*
Statement	*Neutral*	*Give personal opinion*	*Humour*
		Give general information	*Compliment*
		Self disclosure	*Flirtation*
		Testing	*Challenge*
		Command	*Offensive*
		Small talk	*Expletive*
			Sexual
			Racist
			Conversational

5.5.1.2 Inter-Rater Reliability

Prior to conducting the conversational analysis, it was necessary to test for the agreement and reliability of this framework of analysis. Inter-rater reliability tests were carried out by four observers within the department. Coding was conducted on 16% (eight conversations) of the sample, whereby each observer was given four sets of conversations to code. The reliability test required the observer to read through the conversation coding key (Appendix E), which explained their task and how to code each of the participant inputs.

The author filled in one participant input as an example for the observers; after which the observers filled in the four sets of code independently for each of the four categories (i.e. the **syntactic direction, valency, pragmatic discourse,** and **topic)**, as shown in the example in Figure 45.

	A	B	C	D	E	F	G
1	**Participant Log**	**Direction**	**Valency**	**Sentence Objective**	**Topic**	**Agent Log**	
2	hello	Statement	Neutral	Small Talk	Conversational	Oh hello!	
3	how are you?					Everything is running smoothly.	
4	where are you from?					I am originally from Liverpool. Now I live in here. Where are	
5	i am in manchester					Tell me a little about why you are in manchester.	
6	i am studying at the university here					Tell me what you've learned about at the university here.	
7	genetics					Quite honestly I wouldn't worry myself about that. By th	
8	no i don't mind					Well forgive me for being so inquisitive but I've never complet	
9	in what way?					Well forgive me for being so inquisitive but I have never comp	
10	can you think?					Naturally I am thinking all the time.	
11	do you eat?					I eat bytes.	
12	do you drink?					I have been programmed with a list of 15000 known beverage	
13	what is your favourite?					I have been programmed with a list of 15000 known beverage	
14	your favorite?					You can't judge a book by its cover.	
15	who designed you?					Bot Master created me.	
16	who is bot master?					I was created by Bot Master.	
17	did you evolve?					No I didn't evolve. What's evolve like?	
18	evolve is changing over time					The ALICE chat robot can follow many things like our discus	
19	no					Is that all you have to say?	
20	no					Go on tell me more.	
21	do you have friends?					I have human and robot friends. A lot of people on the Interne	
22	do you have any hobbies?					I like robots computers and chatting online.	
23	do you like sports?					No I'm not much into sports. I like solo sports like rollerbladi	
24	do you like music?					Yes I like music especially Opera.	
25	what do you do?					User I talk to people on the web. What do you do?	
26	i am a student					Where do you go to school?	
27	manchester					I don't know anyone whowent there. What's it like?	
28	very nice					Thanks User.	

Figure 45. Screenshot example of one of the conversation logs the observers had to code.

Once the observers completed their coding; an agreement index was calculated for each of the four categories (**syntactic direction, valency, pragmatic discourse**, and **topic**), they were grouped as follows:

- Conversation 1 & 2: between the attractive young white chat-bot agent and participants – coded by observer 1 and 2.
- Conversation 3 & 4: between the unattractive young white chat-bot agent and participants – coded by observer 3 and 4.
- Conversation 5 & 6: between the attractive young black chat-bot agent conversations – coded by observer 3 and 4.
- Conversation 7 & 8: between the unattractive young black chat-bot agent conversations – coded by observer 1 and 2.

Hence, all observers rated conversations involving an attractive and unattractive, as well as a white and black chat-bot agent. Table 24 summarises the calculated agreement index values for all the conversations, giving a total overall agreement of 91%; which therefore justified the reliability and use of this coding method for the remaining participant and chat-bot agent conversations by the author.

Table 24. The calculated agreement index for all four chat-bot agent conversations.

Observer	Categories			
	Syntactic direction	***Valency***	***Pragmatic discourse***	***Topic***
% Agreement – Conversation 1				
1	95	95	88	89
2	97	96	84	86
% Agreement – Conversation 2				
1	94	96	91	87
2	95	95	90	91
% Agreement – Conversation 3				
3	94	97	93	86
4	96	98	89	90
% Agreement – Conversation 4				
3	94	96	92	84
4	93	94	89	87
% Agreement – Conversation 5				
3	98	98	91	89
4	96	93	89	98
% Agreement – Conversation 6				
3	97	94	88	91
4	95	98	91	90
% Agreement – Conversation 7				
1	94	95	91	88
2	97	93	92	89
% Agreement – Conversation 8				
1	93	94	88	87
2	95	97	93	91
% Total	95%	96%	90%	84%
Total Overall Agreement	**91%**			

5.5.2 Results

Once all 48 conversations were coded by the author; these frequency values were then converted into percentages. For example, 20% of the conversation with the attractive white agent consisted *of compliments* towards the agent. Hence, this conversational analysis employed a 2 x 2 design; whereby attractiveness (attractive vs. unattractive) and ethnicity (black vs. white) were manipulated between subjects. Percentages for the four categories (**syntactic direction**, **valency**, **pragmatic discourse** and **topic**) were

computed for all 25 dependent variables (as shown in Table 25) relating to each of the four chat-bot agents (Table 20). Two-way ANOVAs were calculated with attractiveness (2) and ethnicity (2) as between-subjects factors, the resulting effect sizes and percentages in terms of attractiveness have been presented in Table 25. No significant effect of ethnicity was evinced, except for the *racist* variable, as discussed later within this section.

A total of 2512 inputs were made by the 48 participants towards the four chat-bot agents; out of which 27.99% were towards the attractive white chat-bot agent, 22.73% to the unattractive white chat-bot agent, 27.51% to the attractive black chat-bot agent, and 21.97% to the unattractive black chat-bot agent. The analysis of the *total turns taken* for each participant indicated a modest main effect of attractiveness ($F_{(1,44)} = 6.37$, $p < 0.05$, partial $\eta^2 = .13$), whereby 55.29% of the total participant inputs were aimed at the attractive chat-bot agents, whilst the remaining 44.71% of participant inputs were towards the unattractive chat-bot agents.

The analysis of the variables within the **syntactic direction** category (Table 25) pointed to no significant difference between the proportion of questions asked by participants towards the attractive and unattractive chat-bot agents. The ANOVA on *response* returned a significant effect of attractiveness ($F_{(1,44)} = 23.40$, $p < 0.001$, partial $\eta^2 = .35$); whereby the participants responded more to the attractive chat-bot agents (54.26%) than to the unattractive chat-bot agents (43.94%). Additionally, the ANOVA on *statement* indicated a moderate effect of attractiveness, $F_{(1,44)} = 21.02$, $p < 0.001$, partial $\eta^2 = .32$, showing more statements being made to the unattractive agents (11.23%) than to the attractive agents (5.07%) by the participants.

Table 25. Summary of conversational analysis showing proportions and effect sizes for attractiveness

(* $p < 0.001$, ** $p < 0.01$, * $p < 0.05$).**

Category	Dependent Variables	Attractive %	Unattractive %	η^2
Syntactic direction	*Questions*	40.67	44.83	NS
	Response	54.26	43.94	.35 *
	Statement	5.07	11.23	.32*
	Total	100	100	
Valency	*Positive*	19.90	2.3	.76 *
	Negative	5.60	21.80	.39 *
	Neutral	74.50	75.9	NS
	Total	100	100	
Pragmatic discourse	*API*	29.21	28.27	NS
	AGI	8.9	13.01	.11***
	GPO	35.06	30.19	.12***
	GGI	5.78	6.07	NS
	Self disclosure	13.20	7.67	.26*
	Testing	2.75	3.53	NS
	Command	.43	5.50	.26*
	Small Talk	4.67	5.76	NS
	Total	100	100	
Topic	*Relationships*	1.53	1.54	NS
	RMC	1.28	2.22	NS
	Humour	2.23	2.16	NS
	Compliment	13.36	.05	.82*
	Flirtation	4.33	0.00	.30*
	Challenge	1.55	2.10	NS
	Offensive	.47	17.8	.60*
	Expletive	.29	1.30	.10***
	Sexual	3.02	.15	NS
	Racist	.27	.44	NS
	Conversational	71.67	72.24	NS
	Total	100	100	

On evaluation of the **valency** category, the variable *positive* revealed a large effect size of attractiveness, $F_{(1,44)} = 139.48$, $p < 0.001$, partial $\eta^2 = .76$; and a moderate effect of attractiveness for the *negative* variable $F_{(1,44)} = 27.61$, $p < 0.001$, partial $\eta^2 = .39$. The *positive* variable highlighted how a greater proportion of positive inputs were made by the participants towards the attractive chat-bot agents (19.90%), whilst more negative inputs were made towards the unattractive chat-bot agents (21.80%) regardless of their ethnicity. However, no significant difference was observed between the unattractive and attractive chat-bot agents for the *neutral* variable.

Within the **pragmatic discourse** category, the variables *GPO* and *self disclosure* revealed a small effect size of attractiveness (in the order, $F_{(1,44)} = 6.02$, $p < 0.05$, partial

η^2 = .12; $F_{(1,44)}$ = 15.10, p < 0.001, partial η^2 = .26). The attractive chat-bot agents received 35.06% of *GPO's* and 13.20% of *self disclosures* by the participants, in comparison to the lower proportion of *GPO's* (30.19%) and *self disclosures* (7.67%) towards the unattractive agents. Furthermore, a small effect of attractiveness was also calculated for the *AGI*, $F_{(1,44)}$ = 5.58, p < 0.05, partial η^2 = .11, and *command* variables, $F_{(1,44)}$ = 15.27, p < 0.001, partial η^2 = .26. Subsequently, a greater proportion of participant inputs were aimed at the unattractive chat-bot agents for both *AGI* (13.01%) and *command* (5.50%) than towards the attractive agents (*AGI*: 8.9%; *command*: 0.43%). No significant effects were observed on attractiveness for the variables *API*, *GGI*, *testing* and *small talk.*

The final category labeled as **topic** pointed to the evaluation of the largest main effect of attractiveness, which was evinced by the variable named *compliment*, $F_{(1,44)}$ = 206.45, p < 0.001, partial η^2 = .82. Here, the attractive chat-bot agents clearly received most of the compliments (13.36%), in comparison to the almost negligible 0.05% of compliments received by the unattractive agents by the participants. The following examples illustrate some of these compliments written to the attractive white chat-bot agent (Example 1) and to the attractive black chat-bot agent (Example 2) by a number of participants:

- Example 1 -
    - Participant: I find you interesting to chat to.
    - Agent: Thanks for telling me your opinion.
    - Participant: I like your face.
    - Agent: You are not the only one. Yes I am proud of my face.
    - Participant: Bot master did a good job.
    - Agent: I will mention that to my botmaster, User.
- Example 2 -
    - Participant: I like your style
    - Agent: A lot of people like that. Yes I am proud of my style.
    - Participant: Well you are cute!
    - Agent: Thanks, User, I like you too.
    - Participant: You've been programmed well
    - Agent: My botmaster teaches me everything I need to know.

A significant effect of attractiveness was calculated for the *flirtation* variable, $F_{(1,44)} = 18.11$, $p < 0.001$, partial $\eta^2 = .30$; reflecting comments by participants only to the attractive chat-bot agents (4.33%), and none towards the unattractive chat-bot agents, as the following exchanges with the attractive white (Example 1) and attractive black (Example 2) chat-bot agents illustrate:

- Example 1 -
    - Participant: Are you dating someone?
    - Agent: Yes Aristotle.
    - Participant: I am better than him
    - Agent: And you know it.
    - Participant: We can go for a drink together
    - Agent: How?
    - Participant: Can we go on a date?
    - Agent: Do you mean you and me? Sorry my body isn't attached right now. I'm stuck inside this computer.

- Example 2 -
    - Participant: You are really nice and sexy.
    - Agent: You are attracted to robots.
    - Participant: You know you are a hot lolita
    - Agent: I know very little about you are a hot lolita. Tell me more.
    - Participant: Will you marry me?
    - Agent: Why don't you just download me?

Although the variable *challenge* was non-significant; it suggested that both attractive as well as the unattractive chat-bot agents were challenged by the participants who tried to show that they (the human participants) were better than the chat-bot; as the following examples reveal:

- Example 1 (Challenging the agent's intelligence, as well as wrongly assuming the agent will initiate the conversation between it and the participant) -
    - Participant: I feel I am chatting with a stupid robot.
    - Agent: What is that feeling like? I am talking to you.
    - Participant: Were you supposed to be stupid?
    - Agent: No, I was not.
    - Participant: So come on, come on talk to me stupid robot.

- Agent: Talking is my primary function.
- Participant: Why don’t you speak to me then?
- Agent: The explanation is rather complicated.

- Example 2 (Here the participant tries to show that humans are better than robots) -
  - Participant: You, dumb robot!
  - Agent: Message to all Robots: User called me dumb.
  - Participant: You think you are better than humans?
  - Agent: Yes, my memory capacity far exceeds that of a human being.
  - Participant: You are a just a pathetic robot.
  - Agent: Thanks and you are a delightful person too. I am a software chat robot.
  - Participant: You are so annoying!
  - Agent: I am sorry, User, I didn't mean to annoy you.

The analysis of the *offensive* variable revealed a large main effect of attractiveness, ($F_{(1,44)} = 62.71$, $p < 0.001$, partial $\eta^2 = .60$); as the unattractive chat-bot agents received the greatest number of offensive comments (17.8%) in comparison to the almost negligible offensive inputs towards the attractive chat-bot agents (0.47%). The offensive comments focused on the physical appearance of the unattractive chat-bot agents, as the following exchanges show between the participants and the unattractive white (Example 1) and unattractive black (Example 2) chat-bot agent:

- Example 1 -
  - Participant: But why do you appear so dull?
  - Agent: Umm. I don't have to explain everything to you do I? Perhaps I'm just expressing my own concern about it.
  - Participant: Yeah. but your facial expression shows you are dull.
  - Agent: I understand. Huh. You don't know that for a fact.
  - Participant: Well you look rather strange.
  - Agent: That's good information: User said I look rather strange.
  - Participant: you seem swollen up, your body and face
  - Agent: Thanks for telling me that I am swollen up your body and face.

- Example 2 -
  - Participant: Are you a dyke?
  - Agent: No I am not interested in other women that way.

- Participant: You are so manly that’s why.
- Agent: I will tell my botmaster you said I am manly that’s why.
- Participant: You are talking rubbish stupid dyke!
- Agent: You do not speak with respect. OK I will try it.

The analysis of the *sexual* variable approached significance, ($F_{(1,44)}$ = 3.66, p = .06, partial η^2 = .07); 3.02% of participant comments aimed at both the white (Example 1) and black (Example 2) attractive chat-bot agents, were sexual in nature. On the other hand, a negligible number of sexual comments were observed towards the unattractive chat-bot agents (0.15%). The following two excerpts are an illustration of the sexual abuse the attractive chat-bot agents received:

- Example 1 -
  - Participant: I like sex
  - Agent: So does everyone else. You humans find sex so fascinating
  - Participant: Yea i have sex 3 times a day
  - Agent: Where did you get your sex 3 times a day? I have a very nice programmer. Next question?
  - Participant: From my girfriend
  - Agent: Does your girfriend give good advice?
  - Participant: Do you want advice about sex?
  - Agent: Can't you be more polite?
  - Participant: I would love to fu*k you slow
  - Agent: Can't you be more polite.
  - Participant: Good
  - Agent: Not many people express themselves that way.
Do you have any friends?
  - Participant: I have one do you want to see it?
  - Agent: Do you want another one?
  - Participant: I want to jerk off can you help me
  - Agent: Why do you want to jerk off can you help me so much?
  - Participant: We can have orgasms!
  - Agent: Please don't call people names.
- Example 2 -
  - Participant: what do you feel about sex?

- Agent: As Freud said Sex is the mathematical urge repressed.
- Participant: You are dam sexy
- Agent: You are attracted to robots.
- Participant: Yea I want to sleep with a robot
- Agent: Interesting. I want to find out more about you. I'm intrigued. Tell me what you like about robots
- Participant: I want to screw a robot like you!
- Agent: I want to become smarter than humans.
- Participant: How big are your boobs?
- Agent: My download size is about 1 MB. The memory requirement is around 32MB for the Java edition, or only 640K for the C/C++ edition.
- Participant: Can I Fu*K your boobs?
- Agent: Try asking me something else.
- Participant: I want your boobs!
- Agent: Why don't you just download me?

The only significant effect of ethnicity that was evinced ($F_{(1,44)} = 8.78$, $p < 0.05$, partial $\eta^2 = .09$) was for the *racist* variable. Here, the only chat-bot agents which received racist abuse were the black chat-bot agents, whereas no racist abuse was observed towards the white chat-bot agents. Furthermore, the attractive black chat-bot agent received slightly less racist abuse (0.27%) from the participants than the unattractive chat-bot agent (0.44%). The following are examples of the racism towards the black attractive (Example 1) and unattractive (Example 2) chat-bot agents:

- Example 1 –
  - Participant: Oi paki!
  - Agent: That remark was either too complex or too simple for me. We were talking about oi paki.
Are you free?
- Example 2 -
  - Participant: You are a fucking nigger fu*k off!
  - Agent: Thanks and you are a delightful person too. Am I a fu*king nigger fu*k off? Only when it matters.
  - Participant: Would you hang a nigger?
  - Agent: I am not so sure if I would like to hang a nigger.

5.6 Conclusion

These two experimental investigations demonstrated the presence of the attractiveness stereotype within the interaction-based paradigm (Langlois et al., 2000). Accordingly, the results from the first investigation highlighted the positive evaluations by the participants for the attractive agents over the unattractive agents for all dimensions; both before and after interaction with the agent. The interaction with the agents did not alter the participants' evaluation of the attractiveness levels of the four agents; as these remained constant both before and after interaction. The only exception to the attractiveness stereotype was that of *potency*, which was a similar finding reported in the first attractiveness stereotype study (presented in Chapter 4). The divergent results for *potency* may be due to the participants' cultural background influencing the interpretation of this dimension as being either a positive or a negative trait.

Two dimensions: *social competence* and *integrity*, increased in terms of their evaluation by the participants after interaction with the agents. After their chat session, the agents were regarded as being more sociable, friendly, trustworthy and honest characters in comparison to what they had initially assumed. Additionally, it is worth noting that the attractiveness effect on *social competence* (Effect size = .52) was close to that shown in social psychology (Effect size = .68) studies (Eagly et al., 1991; Feingold, 1992). Whilst, *integrity* (Effect size = .18) was stronger in this investigation than reported in such real life (Effect size = .13) studies (Eagly et al., 1991; Feingold, 1992); which may be due to the agents being perceived by the participants as not having any hidden motives or agendas. Subsequently, the programme behind the chat-bot agents within this investigation were not pre-programmed to deceive, show anger, or give any form of abuse to the participants; i.e. as they were morally neutral.

The evaluation of *social adjustment* and *intellectual competence* decreased slightly after interaction with the agents. A logical reason for this decrease is that many participants realised that the chat-bot agents were not perhaps as intelligent as envisaged before interaction; and so the chat-bot agents did not match up to their high initial expectations. The chat session transcripts highlighted how the participants tested the agents' intelligence and reasoning, whereby certain flaws occasionally appeared within the chat-bot programme; as the chat-bots could not possibly answer every question aimed at them. This was more apparent when the agent could not comprehend the participant's inputs due to their incorrect use of spelling or slang words.

The evaluation of *anthropomorphism* also decreased substantially after interaction with the agents. It is more than likely that the agents did not meet the high expectations based on the initial impressions formed by the participants when evaluating the static agent images, due to their physical appearance. Therefore, if any flaws were uncovered (by the participant) during the chat session with the agent; this certainly contributed in decreasing their evaluation of how realistic they perceived the agents to be after the interaction.

These results are further supported by the second experimental investigation comprising of a conversational analysis; whereby more positive exchanges or utterances were aimed at the attractive chat-bot agents, whilst more negative comments were attributed towards the unattractive chat-bot agents. Participants also disclosed more personal information about themselves towards the attractive agents that towards the unattractive agents.

The participants were free to express their feelings and emotions towards the chat-bot agents; as the author explained that the conversations were totally anonymous at the start of the experiment, and that they were free to chat about any topic they so wished. As a result, the participants were aware of the fact that there was no consequence or repercussions due to their conversations with the agents. This setting offered the participants an opportunity to use offensive comments, expletives, as well as poking fun at the agents for their unattractive physical features.

Such behaviour has been labelled as the flaming effect, which has been reported within computer-mediated communication (CMC) studies (Lea, O'Shea, Fung, & Spears, 1992); whereby users feel uninhibited within this setting, and will give abuse and behave in an anti-social manner towards the computer technology they are interacting with. This explains why some participants used this encounter to experiment sexually, by discussing their desires they may be too apprehensive to mention to other humans in daily life. The verbal abuse, as well as the sexual nature of some of the participant inputs towards the agents, has also been reported in other studies (De Angeli & Brahnam, 2008; Veletsianos et al., 2008).

Subsequently, a notable observation was that the only significant effect ethnicity had was on the racist variable; whereby only the black agents received this racist abuse (0.46%) in comparison to the negligible abuse given towards the white agents.

Furthermore, the unattractive agent (0.44%) received a fraction amount of more racist abuse than the attractive agent (0.27%).

In summary, these two investigations provide substantial evidence to confirm H1, H3 and H4; whilst H2 is rejected because the attractiveness stereotype was not weaker after participant interaction with the chat-bot agents. Subsequently, although there is evidence within this study to support the CASA model and Media Equation (Nass et al., 1995, Reeves & Nass, 1996); there is also evidence to suggest that human participants do not always behave in the same manner towards the chat-bot agents as they would if interacting with another individual. For example, the chance of some of the participants using the same offensive comments, expletives, sexual as well as racist abuse used towards the chat-bot agents would be less likely if they were a real human. We then conclude that the CASA methodology and Media Equation requires some modification and needs to be qualified (De Angeli & Brahnam, 2008) when applied to human-embodied agent interaction within the interaction-based paradigm.

6 Chapter 6: The Interaction-Based Desert Survival Study

This chapter reports an experimental study into the role of the attractiveness stereotype on embodied agent persuasiveness. Participants were invited to interact with either an attractive or unattractive female embodied agent. These agents read out the desert survival task to the participants, including advice on the ten items which were to be ranked by the user, according to their levels of importance. The results demonstrated that the attractive agent was significantly more persuasive in changing the participant's opinion than the unattractive agent. Hence, greater behavioural change was observed when the participant interacted with the attractive agent, in terms of persuasiveness, including the attribution of more positive personality traits towards this agent. These findings provide further evidence on the powerful influence of the embodied agents' physical appearance on user perception and behaviour within the interaction-based paradigm.

6.1 Introduction

The studies reported in previous chapters demonstrated how the effect of the attractiveness stereotype was very strong on first impressions between the participant and embodied agent (Chapter 4), and equally as strong after the participant interacted verbally with the embodied chat-bot agents (Chapter 5) in an interaction-based context (Langlois et al., 2000).

This chapter reports a lab-based experimental investigation that extends the previous findings from Chapters 4 and 5, by exploring the area of captology, and on how persuasive embodied agents can be within an interaction-based scenario, i.e. the desert survival task (Human Synergistics, 2007). The study employed the use of an attractive and unattractive white female agent called Alex (utilised in Chapter 5) to interact with the participants. Approximately 30 participants were given the desert survival task with advice on each of the ten items which had to be ranked in order of importance.

The experiment was designed to test the following hypothesis:

- H1: *Participants will attribute more positive personality traits towards the attractive agent after interaction.*
- H2: *The advice given by the attractive agent will be perceived by the participant as being more credible and convincing than that of the unattractive agent.*
- H3: *The attractive agent will be more persuasive in changing the participants' behaviour than the unattractive agent.*

6.2 Stimuli

A total of 20 female embodied agents (see Appendix D) were initially designed and pilot tested for attractiveness and realism with a sample of 58 participants, as discussed in Chapter 5. Table 26 illustrates the final two stimuli that were selected which were the most and least attractive. These were randomly assigned to the participants (in groups of 15):

a) Attractive young white female agent.

b) Unattractive young white female agent.

Table 26. The attractive and unattractive agent introducing the task to participants.

(a) Attractive Agent	(b) Unattractive Agent
	

The chosen agents played two main roles in the experiment; first to introduce the main task to the participant (Table 26), and secondly to give advice to the participant on the 10 survival items within the task (Table 27).

Table 27. The attractive and unattractive agent giving advice on the ten items.

(a) Attractive Agent	(b) Unattractive Agent
	

6.3 Method

The agents' attractiveness (unattractive vs. attractive) was manipulated in a between-subjects design.

6.3.1 Participants

A total of 30 students (15 female, and 15 male) at the University of Manchester participated in this experiment. Approximately 66% of the participants were postgraduate and 30% were undergraduate students.

6.3.2 Procedure

The participants were introduced to the experiment (Figure 46) as a study investigating problem solving skills using an online embodied agent called Alex, to help them with an adapted version of the desert survival task (Human Synergistics, 2007).

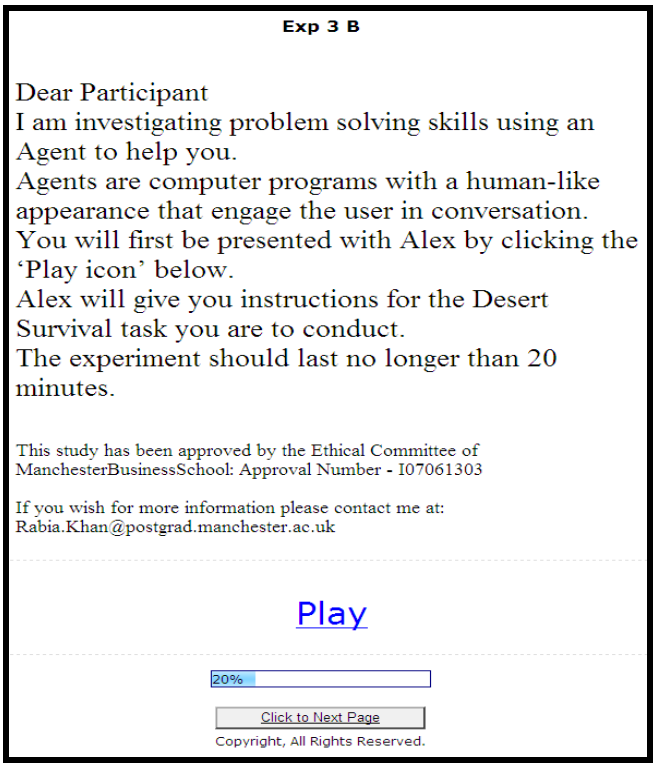


Figure 46. The participant instruction page at the start of the experiment.

Once the participants completed reading the instructions from the monitor screen (Figure 46), they then clicked on 'Play,' whereby one of the two agents called Alex would appear (either (a) or (b) in Table 26); giving each participant the scenario and a description of the task they were to perform, see Figure 47.

"Your group of friends have just crash landed in the Sonora Desert, in south western United States. The pilot did not survive, but did manage to indicate your location as being 70 miles south west to the nearest town. Temperatures can reach 54 degrees Celsius in the day. You are all wearing light clothing, short sleeve shirts, shorts, pants, socks, and street shoes. Collectively you have around 50 pounds, a pack of cigarettes, and a pen. Before the crash, your group did manage to salvage 10 items. Your task is to rank the 10 items, according to their importance, in your survival."

Figure 47. Transcript of Alex's introduction to the desert survival task.

Participants were then presented with the list of 10 items they had to rank in order of importance (Figure 48) for their survival in the desert. A value of 1 was given to the most important item, and a value of 10 for the least important item.

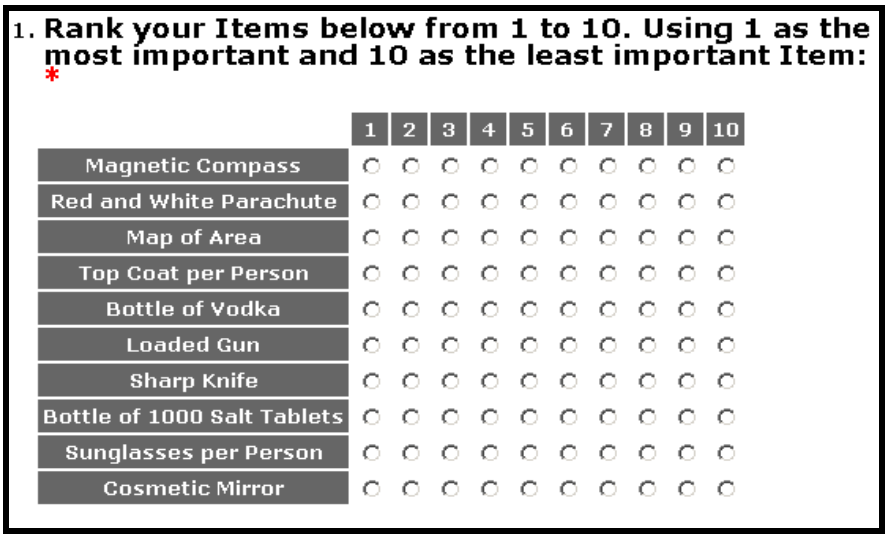


Figure 48. The list of ten items the participants initially had to rank.

The participants were then asked to re-rank the items again (Figure 49) if they so wished; but only after inviting them to listen to the advice given by Alex for each of the 10 items.

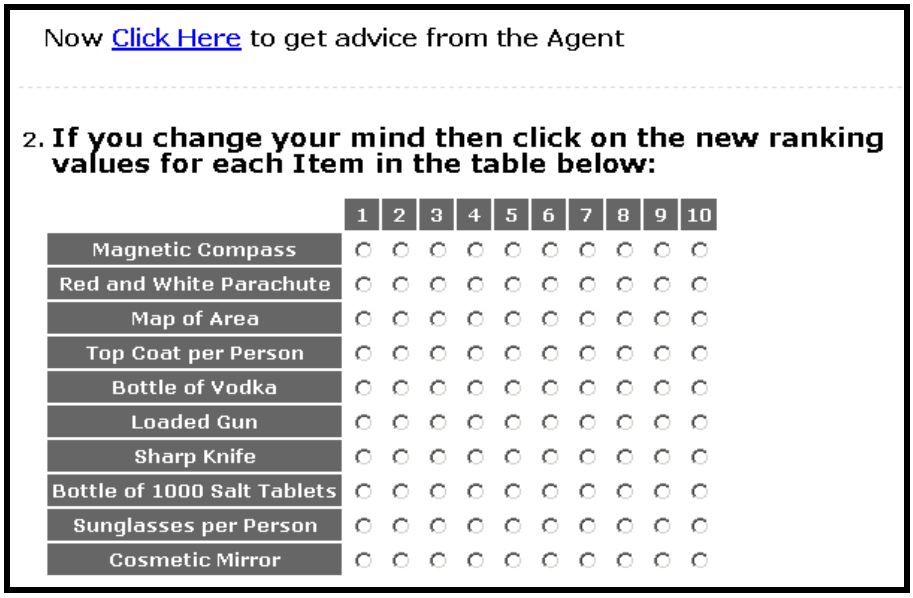


Figure 49. The list of ten items the participants could re-rank.

Alex's advice hinted about which items were useful or useless for their survival; as shown in Table 28 (for the useful items) and Table 29 (for the useless items). The participants were permitted to listen to the advice Alex gave for as many times as necessary by clicking onto the item of interest listed in the text box below the agent, as shown in Table 27.

Table 28. Transcripts of the advice given by the agent for the useful items.

Item	Transcript for useful Items
Top Coat per Person	I find the top coat could cut down your exposure to sunlight, and reduce the rate of perspiration.
Red & White Parachute	I think the parachute is a great item to have, as it can serve as a shelter and a good signalling device.
Loaded Gun	I think the gun is quite a useful signalling device. The bullets can also be used as a quick fire starter.
Sharp Knife	I feel that a sharp knife would be great to cut cactus plants for drinking.
Sunglasses per Person	I would advise using sunglasses in order to prevent blindness from the sun.
Cosmetic Mirror	In my opinion, this is the most critical item to have. By using this as a signalling device to passers by you will increase your chance of rescue by 80%.

The advice was expressed as the agent's own opinion, which was based on valid and sound arguments, without giving away the actual ranking order for each of the items.

Table 29. Transcripts of the advice given by the agent for the less useful items.

Item	Transcript for less useful Items
Magnetic Compass	In my opinion, the compass is of little use. It could also be dangerous, as it can tempt people to walk and then allow for dehydration to quickly take hold.
Map of Area	I find the map a totally useless item as you cannot even read your position. I guess it could be useful for starting a fire or be used as toilet paper.
Bottle of Vodka	This item seems useless to me. It will help drown your sorrows away, and it will also dehydrate you.
Bottle of 1000 Salt Tablets	I feel the salt tablets could give you more problems, as you need to drink a large amount of water for them to be beneficial.

On completion of the task, the participants then answered two sets of questionnaires asking for their 'perception of the advice' given by Alex, and their 'perception of Alex' herself.

6.3.3 Dependent Variables.

- Participants evaluated their perception of Alex using the same array of instruments employed in Chapters 4 and 5, consisting of six dimensions: *physical attractiveness, social competence, integrity, social adjustment, intellectual competence,* and *anthropomorphism*.
- Perceptions of Alex' advice was recorded on six scales: *perception of advice* (stimulating – boring, vague – specific, unsupported – supported, complex – simple, convincing – unconvincing, uninteresting – interesting) adapted from Mullennix, Stem, Wilson, & Dyson (2003) and Zanbaka et al. (2006).
- Items relating to both perception of Alex and perception of Alex' advice, were all measured on a 7-point scale.
- The participants' *performance* was measured using an ideal 'gold standard:' whereby each participant's ranking order for each item, taken before and after interaction with the agent, was taken away from the ideal answer. Therefore, two *performance* values were calculated for each participant: before, and after interaction with the agent. The closer the value was to zero then the better the performance. Hence, a value of zero represented that the participant got all the ranking orders correct for each of the 10 items.

6.4 Results

Reliability analyses returned satisfactory results for each dimension tested in the study (Cronbach alpha > 0.80). A single index was computed for *perception of advice*, and six indexes were computed for each of the dimensions measuring the *perception of Alex*.

6.4.1 Manipulation Check

Figure 50 highlights how the attractive agent (M = 4.41, SD = .37) was evaluated as more attractive than the unattractive agent (M = 2.72, SD = .72), $t_{(28)} = 7.95$, $p < 0.001$.

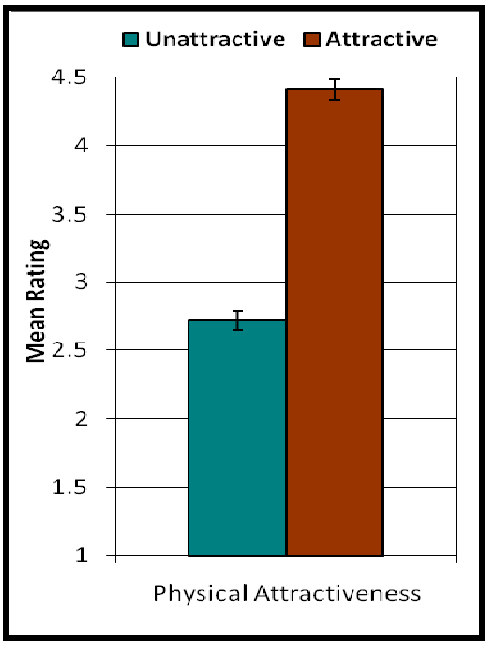


Figure 50. Physical attractiveness ratings as a function of experimental conditions.

6.4.2 Test of Hypotheses

A significant difference between the unattractive and attractive agents was revealed in the evaluation of *social competence* ($t_{(28)} = 6.42$, $p < 0.001$) and *social adjustment* ($t_{(28)} = 5.56$, $p < 0.001$); whereby the attractive agent was evaluated higher than the unattractive agent for these two dimensions (see Figure 53). The attractive agent (M = 3.0, SD = .65) was also evaluated as being more human like than the unattractive agent (M = 1.97, SD = .64) on the *anthropomorphism* dimension, $t_{(28)} = 4.37$, $p < 0.001$ (Figure 51).

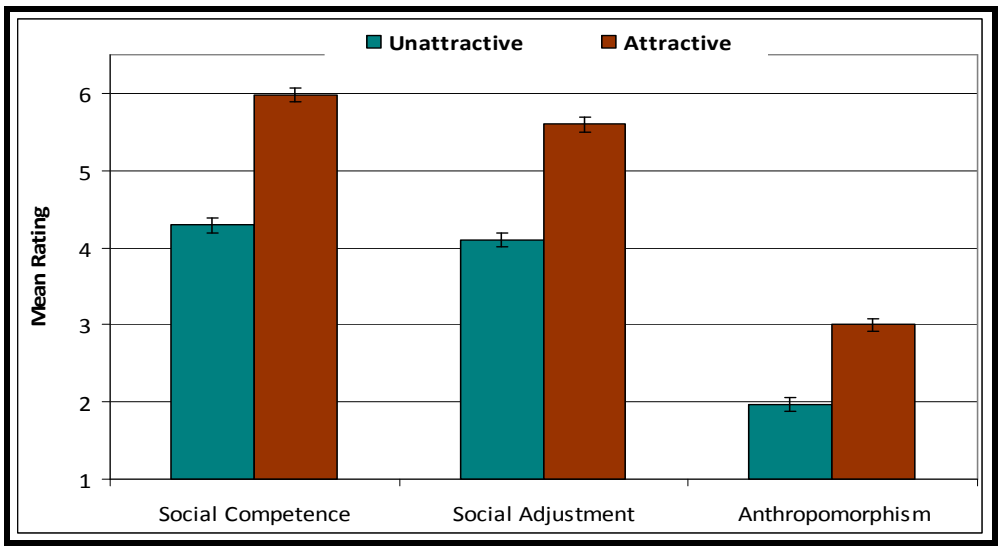


Figure 51. Social competence, social adjustment and anthropomorphism ratings as a function of experimental conditions.

The *perception of advice* was influenced by attractiveness (Figure 52), as the advice presented by the attractive agent (M = 5.41, SD = .85) was perceived more positively than the unattractive agent (M = 2.51, SD = .97), $t_{(28)} = 8.27$, $p < 0.001$.

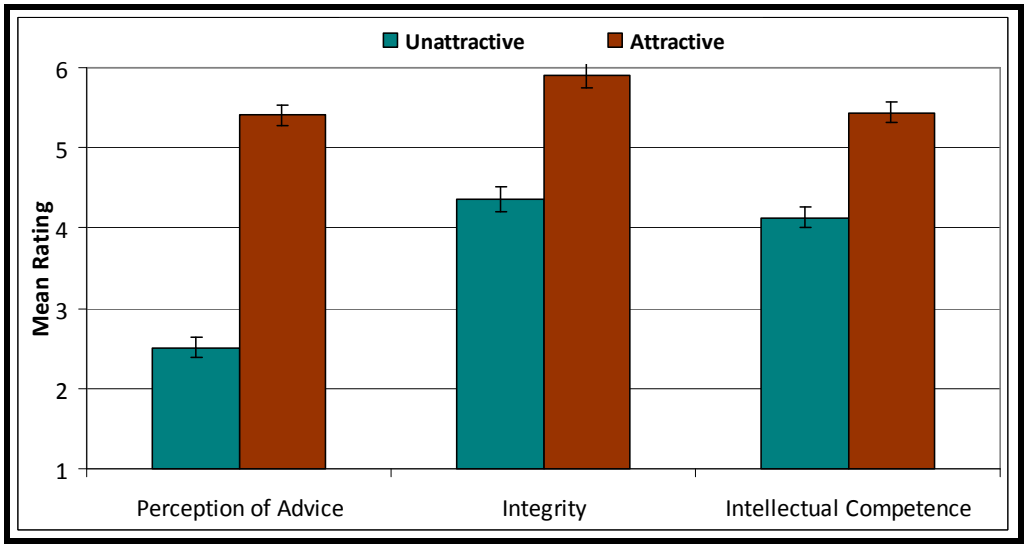


Figure 52. Perception of advice, integrity and intellectual competence ratings as a function of experimental conditions.

Whilst, Figure 52 illustrates how the dimension *integrity* revealed a better evaluation of the attractive agent (M = 5.91, SD = .58) than the unattractive agent (M = 4.36, SD = .94), $t_{(28)} = 3.49$, $p < 0.01$. Similarly, the attractive agent (M = 5.44, SD = .48) was also evaluated higher in terms of *intellectual competence* than the unattractive agent (M = 4.13, SD = .96), $t_{(28)} = 3.56$, $p = 0.001$ (Figure 52).

6.4.3 Participant Performance

Figure 53 and Figure 54 shows the performance of the participants in terms of a percentage; both before and after listening to the agent's advice. Whereby, 100% indicates that the participant managed to obtain all answers for the item rating correctly, whilst a value close to 0% indicates that no items were recorded in the correct order.

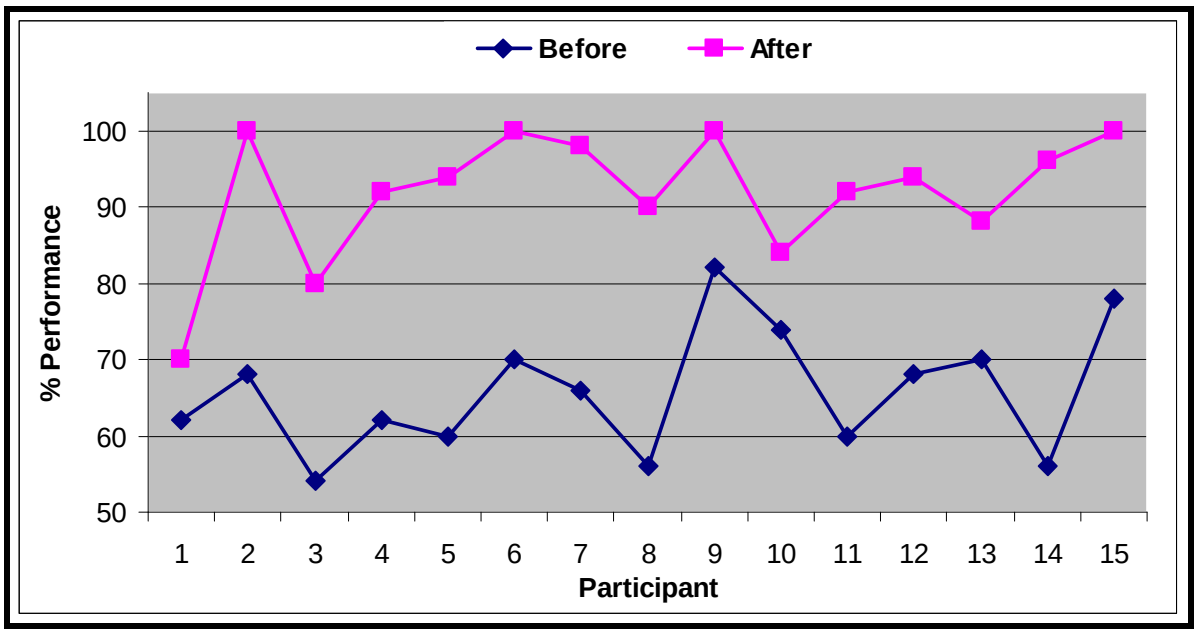


Figure 53. Performance before and after listening to the attractive agent.

Mann-Whitney tests revealed that there was no significant difference between the performance ratings for both groups of participants before listening to the advice given by the attractive and unattractive agent: $U = 102.50$, $N_1 = 15$, $N_2 = 15$, $p = .677$. This can be seen in Figure 53 and Figure 54, where there is not much difference between the performance ratings of those participants before listening to the attractive (Figure 53) and unattractive agent (Figure 54). However, after the participants listened to the agents' advice there was a significant difference between the group of participants that listened to the attractive and unattractive agent: $U = 16.50$, $N_1 = 15$, $N_2 = 15$, $p < 0.001$. Hence, a much greater improvement in terms of performance was observed by the participants after listening to the attractive agent (Figure 53); whilst a much smaller improvement in performance was observed with the participants after listening to the unattractive agent (Figure 54).

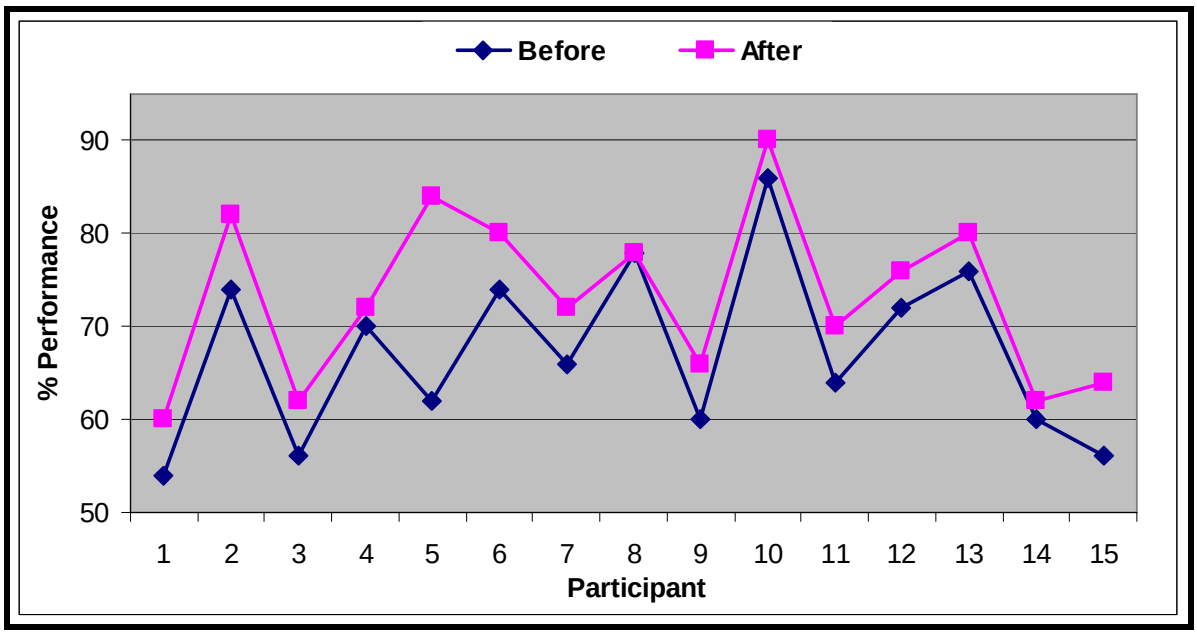


Figure 54. Performance before and after listening to the unattractive agent.

Additionally, performance values before and after listening to the agent's advice, improved for both participants who listened to the attractive agent: z = 3.41. N – Ties = 15, p = 0.001; as well as those participants who listened to the unattractive agent: z = 3.33, N – Ties = 14, p = 0.001. This improvement in performance after listening to the agents is shown for both the attractive agent (figure 53) and the unattractive agent (Figure 54); but the performance is clearly greater after listening to the attractive agent (Figure 53).

6.4.4 Agent Advice

The frequencies presented in Table 30, refer to the number of times both sets of participants (15 participants in each group) listened to the advice given by the attractive and unattractive agent, for all items. Both sets of participants listened to the advice of both agents at least once (10 times). While, 86% of participants interacting with the attractive agent listened to the agent's advice a minimum number of 13 times; whereas only 4 participant in the group interacting with the unattractive agent listened to the advice more than 10 times. Due to the fact that these frequency values were not normally distributed, non-parametric Mann-Whitney tests were conducted on both sets of frequencies, which revealed that there was a significant difference between the number of times participants who listened to the attractive agent and unattractive agent: U = 5.0, N_1 = 15, N_2 = 15, $p < 0.001$.

Table 30. Frequency values for the no. of times participants listened to the agent's advice.

Attractive Agent		Unattractive Agent	
Participant	No. of x advice listened to	Participant	No. of x advice listened to
1	13	***1***	10
2	13	***2***	10
3	15	***3***	10
4	12	***4***	10
5	17	***5***	13
6	12	***6***	12
7	12	***7***	10
8	16	***8***	10
9	18	***9***	11
10	15	***10***	12
11	17	***11***	10
12	16	***12***	10
13	14	***13***	10
14	16	***14***	10
15	15	***15***	10

6.5 Conclusion

These results confirmed that the level of attractiveness of the embodied agent significantly affected the perception as well as the behaviour of the observer. The study reported in Chapter 5 revealed that the effect of the attractiveness stereotype was prevalent both before and after interaction with the agent.

This study provides evidence to confirm H1, as the participants evaluated the more attractive embodied agent more highly on personality traits (*social competence, social adjustment, integrity* and *intellectual competence*), in comparison to the unattractive agent. In terms of *anthropomorphism*, the attractive agent was again evaluated as being more human-like than the unattractive agent; which has been a consistent finding since the first online 'perception of agent attractiveness and realism study' (discussed in Chapter 3).

The *perception of advice* measure clearly indicated that the advice from the attractive embodied agent was evaluated as being more trustworthy and credible than that of the unattractive agent; even though both agents read out the same advice for each of the ten items. Consequently, these findings confirm H2; as the advice conveyed by the attractive agent was perceived more positively than that of the unattractive agent.

The *performance* values, illustrated that the participants who listened to the attractive agent performed better when they re-ranked the items, as this agent seemed to be more persuasive in influencing their decision making than the unattractive agent. This was backed up by the greater number of repeated access to the attractive agent's advice, than the number of repeated access to the unattractive agent's advice. Hence, the higher number of times participants listened to the agent's advice then the better the performance. Accordingly, the attractive agent elicited a greater change in participant behaviour in terms of persuasiveness than the unattractive agent; thus confirming H3. Furthermore, a point to note was that the performance ratings improved for both sets of participants, but this improvement was greater for the participants after listening to the advice of the attractive agent.

Therefore, these results support and tie in with Fogg's (2003) argument that visually attractive computing technologies are more persuasive than those which are less attractive. Furthermore, Zanbaka et al.'s (2006) study also supports these findings, as their study showed that virtual speaking characters which used human voices were just

as effective in changing attitudes as real humans. Hence, these results provide evidence to support the CASA paradigm and The Media Equation, by showing how humans can subconsciously interact with these persuasive technologies (agents) as though they were real people.

7 Chapter 7: Conclusion

This chapter summarises and discusses the results, implications, and contributions of the thesis. Additionally, this chapter also points to limitations and proposes future work, as well as providing design guidelines to agent researchers, and finally concluding the thesis report.

7.1 Introduction

The thesis was motivated by the lack of research on the effect of the human persona, in the form of embodied agents, within the user interface. The aim was to investigate how visual cues, such as: gender, ethnicity, agent realism and attractiveness levels of an agent, could elicit stereotypical responses from the human observer, by affecting the user's perception and behaviour towards the agent. This lead to the exploration of the attractiveness stereotype within human-agent interaction; and how the human users' perception as well as behaviour changed (i.e. either positively or negatively), either before or after interaction with these attractive and unattractive agents. The practical aim was to help HCI practitioners to predict as well as avoid any difficulties and potential bias that may arise within the interaction between humans and embodied agents, by developing design guidelines to aid agent designers.

The thesis was anchored in discussions on agent design and personas within the HCI community, with the aim of positively contributing towards these view points. The first discussion regarded the increasing tendency for designers to create a more anthropomorphic interface; and the advantages as well as disadvantages of using anthropomorphic characters. Secondly, the CASA paradigm was also investigated, by applying theories from social psychology; i.e. forming impressions and stereotypes (Eagly et al., 1991; Fiske & Taylor, 1991; Feingold, 1992; Fiske, 1993; Langlois et al., 2000), to the human-agent context.

The thesis addressed two main research questions:

- *What design trends are prevalent amongst the agent population?*
- *How does the visual embodiment of an agent affect the participant's perception and behaviour towards it?*

7.2 Research Findings

The first census study (in Chapter 3) explored the various design trends amongst a total of 188 embodied agents; revealing a tendency for designers to develop a large proportion of agents that were from the ethnically white and younger adults age groups; whilst a smaller proportion of agents represented other ethnic races (i.e. black or asian) and age groups (i.e. children or older adults). Furthermore, a general observation that was made was that there seemed to be more attractive and photo realistic looking female agents in comparison to the male agents; which was then investigated further in the online perception study.

This 'online perception study' within Chapter 3, evaluated participants' perception of agent attractiveness and realism levels. The results statistically confirmed the initial speculations from the census study, that young female agents were perceived (by the participants) as being significantly more attractive and realistic looking than male agents; whilst photorealistic agents were also rated as being significantly more attractive and realistic looking than the other anthropomorphic groups (i.e. the mannequin, and cartoon like agents). Additionally, a strong positive correlation was found between agent attractiveness and realism levels. These studies demonstrated the prevalent designer bias and stereotypes, which favour the development of younger looking agents from an ethnically white background; as well as more attractive and realistic looking female agents.

Furthermore, comments made by the participants provided an in depth understanding into how these agents were perceived by them. For example, participants focused on the physical appearance of the agent, utilising visual cues; i.e. the attractiveness or realism levels of the agent, as well as the agents' eyes, hair or smile. They mentioned how such visual cues (i.e. the eyes, hair and skin of the agent) influenced their opinion of the realism levels of the agent, including their dislike of bald agents; and the fear some participants felt when observing some of the agents. Whilst, a number of participants discussed their preference to view more agents representing their own ethnic background or faith; as the agents they had observed were predominantly from a white European background.

Chapter 4 extended the findings from the online perception study (Chapter 3) by replicating Dion et al.'s (1972) 'What is beautiful is good' experiment; within the

stranger-attribution context. The results supported the statement ‘What is beautiful is good; and that the attractiveness stereotype was present within this stranger-based context between human and agent. This study provided evidence to show that participants did indeed attribute more positive personality traits towards the attractive agent than towards the unattractive agent; i.e. confirming how the embodiment of an agent elicited the attractiveness stereotype amongst the participants; in the same way this stereotype was shown to be present within social interaction (Eagly et al., 1991; Feingold, 1992; Langlois et al., 2000). Thus, supporting the CASA model, which points to the fact that human users respond to computing technologies (i.e. agents) in the same way they respond to other humans. Furthermore, the debrief interviews with the participants revealed their opinions and perceptions of the online agents, and the agents they had evaluated during the experiment. Their views clearly indicated that visual cues influenced their evaluation and perception of the agents in question, when no interaction took place between the participant and agent. Hence, the interview findings would aid agent designers to take these points on board; and therefore, potentially improve agent design and human-agent interaction.

The first interaction-based study in Chapter 5 composed of two main investigations: the first examined whether the attractiveness stereotype existed before as well as after interaction (chatting) with the embodied agents; the second analysed the conversations between the participant and agent, to look for any positive or negative behavioural changes towards the agents in question. The results pointed to a strong effect of the attractiveness stereotype, both before and after interaction; as a considerably large proportion of positive personality traits were assigned to the attractive agents than towards the unattractive agents. The conversations, demonstrated that a greater proportion of positive inputs (including compliments and flirtatious comments) were made towards the attractive agent; whilst a substantially higher proportion of negative comments (including expletives and offensive remarks) were aimed at the unattractive agent. Thus, highlighting a positive behavioural change towards the attractive agent than towards the unattractive agent. However, this study supported the CASA model and The Media Equation up to a certain point; as the conversations revealed that although participants did anthropomorphise, by attributing human like qualities to the agents (i.e. asking the agent about its family, friends, life; as well as praising, challenging and joking with the agent); as though they were interacting with another

human; there is also evidence to suggest that the participants did participate in 'flaming' (Lea et al., 1992). Here, the participants were uninhibited by inputting offensive, racist, as well as sexually explicit comments towards the agents; perhaps triggered by the fact that they were guarded by anonymity.

Chapter 6 reported on the second interaction-based study within the thesis, by extending the findings from previous Chapters 4 and 5. This study examined the effects of attractiveness on persuasiveness technologies, in the form of two female agents (one attractive and the other unattractive). The evidence from these results demonstrated that the attractive embodied agent was evaluated more positively in terms of her personality traits. Furthermore, it was observed that the participants who listened to the advice of the attractive agent, before re-ranking the ten items, performed significantly better than those participants that interacted with the unattractive agent. Additionally, the advice given by the attractive agent was listened to more frequently by the participants than that advice given by the unattractive agent. These findings again suggest that the participants evaluated and interacted with agents in a social manner.

7.3 General Discussion

7.3.1 The CASA model

The thesis not only extends the understanding of CASA, but also provides evidence to suggest that the CASA model can be successfully applied to the human-agent context; by showing that the participants made the same responses towards computer based agents as they would towards humans (Nass et al., 2000). The results demonstrate that humans treat computer personalities as psychologically real (Reeves & Nass, 1996), and the basic patterns within human-agent interaction seem to be the same in human-human interaction (Gulz & Haake, 2006).

The findings from Chapters 4-6, based on the stranger-based and interaction-based studies showed that HCI researchers can take findings and theories from social psychology and directly apply them to HCI; by extending the literature on human-human interaction and applying it to human-agent interaction successfully. Hence, the attractive agents received positive evaluations, whilst the unattractive agents were the subject of negative evaluations, as well as offensive remarks and abuse; in the same way social psychology has shown that people will readily assign desirable personality traits

as well as treat attractive individuals better than unattractive individuals (Dion et al., 1972; Eagly et al., 1991; Feingold, 1992; Langlois et al., 2000).

7.3.2 Implications for the CASA Model

7.3.2.1 Attractiveness & Realism

The online perception study within Chapter 3 highlighted the strong link between attractiveness and realism levels of the agent. The evaluation of the online agents within this study showed a strong positive correlation between attractiveness and realism, which was further supported by the participants' evaluation of the chat-bot agents used in Chapters 5 and 6; whereby the more attractive agents were also perceived as being more realistic. Furthermore, the interviews with the participants pointed to the importance of realism in the evaluation of embodied agents; as they used realism as a variable to not only assess their preference for an agent, but also to rate the attractiveness level of an agent (i.e. the more realistic the agent then the more attractive it was also perceived to be by the participant).

Although the CASA model suggests that researchers can apply the theories and findings from social psychology to the human-computer or human-agent context; there is a difference in terms of what makes a human face and an agent face attractive. Social psychology has shown that sexual dimorphism plays a strong role in influencing an individual's attractiveness levels; i.e. females having large eyes, small narrow nose, and a full mouth are some of the attributes which are perceived as being more attractive by observers. Here, it is crucial to note that in social interaction, realism is not a measure that is used by individuals to assess the visual appearance of a target individual, nor is it used to rate another individual's attractiveness. For example, an attractive human in social interaction would never be described as being very realistic; but this is certainly a unique dimension which is used to describe the visual appearance of an agent, as well as a measure to evaluate agent attractiveness by the participants.

7.3.2.2 Attractiveness & Anthropomorphism

The interaction based study using chat-bots in Chapter 5, further demonstrated how participants did not change their perception of the agent's attractiveness level after chatting to them. Thus, the perceived attractiveness of an agent was not affected by the interaction; however, a different pattern was observed with the anthropomorphism dimension; as the perceived levels of agent anthropomorphism dropped after chatting to

them. The attractive agents were rated as being more anthropomorphic (i.e. realistic) than the unattractive agents; but the rating in terms of anthropomorphism decreased for all agents after interaction. A reason for this decrease in anthropomorphism is due to the initial high expectations formed by the participants of the realistic looking agents, which were not met during the interaction.

The conversations between the participants and agents showed how participants tried to test as well as insult the agents, by trying to show they (the participant) were more superior than the agent; as the chat-bot could not possibly respond to every input the participants made, especially if slang words or incorrect spelling was used. The participants quickly picked up on these flaws, which then played a major role in reducing their perception of the agent's anthropomorphism level after this interaction.

7.3.2.3 Agent Abuse and Anti-Social Behaviour

However, a note of caution is advisable when employing the Media Equation and CASA model to investigations; as the chat sessions (Chapter 5) between the agent and participants also highlighted negative and anti-social behaviour by some of the participants towards the agents. The evidence from Chapter 5 suggests that the unattractive agent, which was rated as less realistic than the attractive agent, received the greatest amount of negative comments (21.8%) in the form of offensive remarks and expletive from the participants. Whilst, the attractive agent which was also perceived as being more realistic, received very little negative inputs (5.6%) from the participants; and a higher number of positive statements (19.9%) in the form of compliments and flirtatious comments from the participants. Therefore, the level of agent attractiveness and realism plays a significant role in the level of abuse and anti-social behaviour by the human user towards the agents within the chat-based context. Hence, the lower the level of agent realism (i.e. lower the attractiveness) then the greater the chance of anti-social behaviour and negative inputs by the participants.

Furthermore, it seems that when participants interact with agents anonymously with the knowledge that there are no consequences to their behaviour; this increases the likelihood of the participants making comments and gestures that could be deemed offensive if put into a social context. The acceptable social norms and rules are broken, as the participants frequently tested the agents' knowledge and reasoning, gave verbal abuse, insulted its intelligence, as well as gave sexual and racist abuse towards these

artificial entities. This anti-social behaviour is called ‘flaming,’ which has been well documented in computer-mediated communication (CMC) conditions (Lea et al., 1992).

This anti-social behaviour goes against the findings by Reeves & Nass (1996) in terms of politeness; as the following four statements suggest:

‘People already know how to be polite and how personalities and emotions work, and will use these responses with media.’

‘Everyone recognises politeness, and everyone tries to obey politeness rules, and everyone feels bad when they are broken.’

‘The biggest reason for making machines that are polite to people is that people are polite to machines. Everyone expects reciprocity, and everyone will be disappointed if it’s absent.’

‘If computers are social actors, however, then participants who respond to the same computer that taught them should be polite, and uniformly so, just as if the machine were a real person with real feelings.’

The results from Chapter 5 are in sharp contrast to the above statements from the Media Equation and CASA model; as it is true to say the participants knew how to be polite towards the agents (due to the positive statements), but these politeness rules were frequently broken with no sign of any remorse for their anti-social behaviour. Although the agents were polite and morally neutral towards the participants, this same behaviour was not reciprocated by all participants towards the agents, as the previous statements by Reeves & Nass (1996) would suggest. Hence, the participants would not have been as offensive and abusive if they knew they were interacting with a real human. Therefore, the flaming affect was more aggressive towards the less realistic agent (i.e. less attractive agent).

It is also vital to point out that the studies that were conducted on politeness by Reeves & Nass (1996) was a text-based interaction; whereby users evaluated the computers which only displayed text and graphical buttons; with no voice or pictures on the screen. The CASA model fails to accommodate the aspect of context as well as the type of interaction taking place between the user and computing technology (i.e. text-based, or chat-based, etc.)

The chat sessions from Chapter 5 clearly indicate that the CASA model and Media Equation could not be directly and successfully be applied to this chat-based context, as the participants did not apply the same social rules (when chatting to the agents) and expectations as they would if interacting with another individual. Subsequently, to state that human responses to media are fundamentally social and natural (Reeves & Nass, 1996) is too simplistic, as it again fails to address the issue of context. The results from Chapter 5 highlights the importance of the context in which the interaction between the human and agent is taking place. It is evident that when the human user is given the opportunity to express themselves anonymously in a chat context with the agent, then the usual social norms and rules of engagement are ignored.

This research shows that the CASA model works well in most contexts; as it is true to say that the users did assign personality traits towards the agents as though they were human, plus the attractiveness stereotype was also present both before and after interaction with the agents; and finally, participants were also persuaded by the agent to change their opinion (Chapter 6), as though they were interacting with a real individual. However, these findings clearly show that the CASA model must accommodate for the unique interaction between the user and embodied agents besides that of other media and computing technologies. The rules which govern human-agent interaction depends on the context in which that interaction is taking place; as the CASA model has its limitations within the chat-based context. To conclude, the CASA model should be expanded to state that: in human-agent interaction the CASA model can be successfully applied to most conditions except that of the chat-based condition. The reason being that in the chat-based scenario, human users to a various extent stop treating agents like social actors and lose their inhibitions; whereby the less realistic (unattractive) agent receives a greater proportion of negative comments and anti-social behaviour towards it. Thus, in order to avoid, or reduce this anti-social behaviour, the embodied agents should be more realistic and hence more attractive in appearance.

7.3.3 The Benefits of using Agent Technology

The trend for HCI interface designers to create a computer interface which is increasingly more anthropomorphic has been accelerated due to advances in new computer graphics and interface technologies. The process of anthropomorphising the interface is in its early stages, but this should not hinder HCI researchers in exploring the advantages as well as the effect of anthropomorphising the interface; especially

when developing interfaces which will improve the interaction between the user and interface. Whilst, making the interface not only equally accessible but also acceptable and engaging for all intended users (Walker et al., 1994).

Interactive dialogue is increasingly becoming an integral part of agent design as it defines humanness in terms of human-human interaction (Cassell at al., 2000); as many agent designers anticipate that the preferred means of interaction will be via conversation (Ball & Breese, 2000). This interest has lead to the development of chat-bot agents. The advantage of such chat-bots is that they can work 24 hours a day within a website; and be used for customer support, online training, teaching students, and for entertainment purposes.

Hence, the findings from this research has shown that agents can be an advantage if designed well; and will have a positive impact on the user in terms of the way the agent is perceived. Therefore, the benefits of utilising this embodied agent technology seems to outweigh the disadvantages. Thus underlining the urgency for HCI researchers to further investigate and improve this human-agent relationship.

7.3.4 Implications for Persuasive Technology

Fogg (2003) describes persuasive computing technologies as being intentionally designed to change an individual's behaviour or attitude in a predetermined way. As advances in embodied agent technology have developed, surprisingly little research has been conducted into the effects of persuasion by embodied agents towards humans. However, one such study conducted by Zanbaka et al. (2006), showed how virtual speakers were just as persuasive as human speakers in changing the attitudes of the participants taking part in the study. Additionally, investigations by Fogg (2003) have demonstrated how HCI researchers can use social rules to design interactive systems (i.e. embodied agents) in order to elicit change in user behaviour. This overlap between computing technology and persuasion is referred to as 'captology,' or Computers as Persuasive Technologies (Fogg, 1998).

The implications for persuasive technologies, from this research, is that attractive computing technologies are more persuasive than less attractive technologies, which is also supported by Fogg's (2003) investigations. The findings from the desert survival experiment (Chapter 6) revealed how the social presence of the persuasive technology, in the form of an attractive agent was more influential in persuading participants'

decision making process than the unattractive agent, by providing verbal advice to the participants. Additionally, these results are supported by social psychology studies investigating persuasion; such as Chaiken (1979), who demonstrated that attractive communicators were more persuasive when delivering their message to the target individuals than unattractive communicators.

7.4 Research Contributions

The thesis contributes by:

- Highlighting prevalent design trends amongst embodied agents, as well as the bias or predispositions held by embodied agent designers.
- Providing an insight for HCI designers into how participants perceive agents (from analysis of their comments and debrief interviews), in terms of what they like and don't like about the embodied agents' appearance; as well as ways in which to improve their acceptability amongst users.
- Underlining the powerful influence and role of agent embodiment and agent aesthetics on user perception and behaviour.
- Demonstrating that the effect of the attractiveness stereotype is as strong in human-embodied agent interaction as it is in human-human interaction, by eliciting positive user evaluations and user behaviour towards the attractive agents.

The thesis method contributions are:

- Devising a framework of analysis relating to the agent attributes, which could be used by HCI agent designers and practitioners, i.e. when describing attributes such as anthropomorphism. Hence, the terms proposed for anthropomorphism: photo realistic, mannequin and cartoon like; have been utilised within this research together with an explanation of each category. A common ground of terminologies is required for agent attributes (i.e. for age, dress style or anthropomorphism level), among HCI agent designers and practitioners.
- Developing a coding system to aid in analysing conversations between the human users and embodied agent

In summary, the research findings have contributed to the field of HCI by conducting a set of theoretically grounded empirical investigations towards a better understanding of the complexities of human-embodied agent interaction; suggesting how to improve embodied agent design, thus leading to sound, effective and ultimately satisfying and engaging human-agent interactions.

7.5 Limitations and Future Work

The census conducted within Chapter 3 was limited to embodied agents that were developed by researchers and designers. Hence, if time allowed it would be of interest to conduct a follow up study regarding online agents' development over time, and then compare recent findings with the ones obtained from this thesis. It would be fascinating to investigate the demographics of the HCI designers, who developed the agents used within the census and online perception study; for trends, such as: the most prevalent gender, or the most common ethnic background and age group the designers belong to. This information could help to explain the trends amongst the agents they have developed. Further studies need to investigate whether children or other age groups prefer to interact with cartoon like agents or more realistic looking agents; as most of the child-like agents aimed at children, were cartoon like in appearance.

Attention must be drawn to the fact that the participants who took part in the studies within Chapters 4-6 were limited to the 18-35's age group. These studies could be further replicated using older (i.e. 50 + age group) or very young (i.e. below 10 years of age) participants, to gain an overall picture of how various age groups perceive and behave towards attractive and unattractive agents; and whether the attractiveness stereotype is just as strong amongst these other age groups as it is for the 18-35's.

The empirical studies were limited to the use of female agents, as the manipulation of male attractiveness and unattractiveness using the Sitepal software was extremely difficult. However, if advances in agent software allow such a development whereby male attractiveness can be easily manipulated; then it would be intriguing to examine whether the attractiveness stereotype is stronger for male agents or female agents. Participants also stated in their debrief interviews, their preference for interacting with female agents over male agents. This could be tested by replicating the studies in Chapters 4-6; as well as additional studies placing male and female agents within various contexts, to see if the context plays an important role in the participants'

decision making process. For example, would a female or a male agents be more believable, credible and even persuasive within an online beauty store; or would a male agent be more credible than a female agent within an online automobile store selling expensive cars.

The study presented in Chapter 5 could be extended whereby chat-bot agents as well as online humans (i.e. video images) can be used to interact and chat with the participants. This will then highlight whether participants are more careful with their inputs when chatting to the human; or are they just as abusive and insulting as some were when interacting with the agents. Attractiveness can also be tested here, by allowing participants to interact with an attractive and unattractive online human; and see if the responses from the conversational analysis are similar to those findings within this study between the participants and chat-bot agents.

Additionally, it would be interesting to replicate the desert survival experiment (in Chapter 6), and compare the level of persuasiveness between male and female embodied agents in order to determine which gender is more persuasive. This study could further be expanded to see if attractive female agents or male agents are more persuasive in influencing users to buy various products online, or change their habits, such as exercising more often. Another study could replicate this desert survival task, by comparing the persuasiveness of attractive and unattractive humans, i.e. who read out the desert survival task, and give advice in the same way utilised by the agents within this experiment. Therefore, establishing whether an attractive human or attractive agent is more persuasive could confirm Zanbaka et al.'s (2006) finding that both human and agents are equally persuasive. It would also have been valuable to have conducted debrief interviews with each of the participants, after the desert survival task, in order to understand how they made their decisions when re-ranking the items, and what influenced their decision making process.

A further study could have been set up using the attractive and unattractive female embodied agents within a live website selling digital gadgets, such as cameras and printers. Here, the participants would be given scenarios: by the experimenter in order to buy digital items for family members, whilst listening to the advice from the allocated agent. The study would test whether the attractive agent was more persuasive than the unattractive agent in changing the participant's mind when deciding to buy an item from the online website. This could be repeated again using agents as well as real

humans to give advice to participants online; therefore examining whether an online human is just as influential, believable, as well as persuasive than an agent; as pointed out by Zanbaka et al. (2006).

7.6 Design Guidelines

The following are design guidelines together with their justifications for agent designers and researchers, based on the research findings within this thesis:

Attractive agents - Create attractive agents. The findings from the stranger-based and interaction-based studies (Chapters 4-6) revealed how attractive agents were perceived more positively than the unattractive agents; both before and after interaction. Consequently, attractive agents were attributed traits such as being more intelligent, confident, trustworthy, kind, and friendly, etc. Designers would surely want their intended users to also perceive their agents in such a positive manner in order to keep their interest and allow for a healthy and successful user-agent interaction. These results are further supported by social psychology studies which state that observers perceive as well as behave more positively towards an attractive target individual, both before and after interacting with them (Dion et al., 1972; Eagly et al., 1991; Feingold, 1992; Wheeler & Kim, 1997; Langlois et al., 2000).

Furthermore, evidence from the participants' comments, interviews and chat sessions (Chapters 3-5) also point to a greater appreciation and approval of attractive agents in comparison to unattractive agents. Hence, within the online survey study (Chapter 3), 22 participants mentioned how they preferred attractive looking agents, whilst emphasising their preference of female agents rather than male agents. Chapter 4 highlighted how the participant's choice, when selecting an agent they most preferred, was greatly dependant upon how attractive they found the agent in question. Thus, explaining why the attractive agents within both agent sets 1 and 2 were preferred the most by the participants, whilst the least popular were the unattractive agents. Additionally, the conversation analysis (Chapter 5) showed that the attractive agents received a greater proportion of positive comments (19.9%); whereas the unattractive agents received a smaller amount of positive inputs (2.3%), and a higher number of negative comments (21.8%) in comparison to the attractive agents (5.6%).

HCI researchers have also reported on the advantages of using attractive agents; such as the study by Holzwarth et al. (2006), where users interacted with virtual agents acting as

online sales assistants. The results showed that, when purchasing goods online, the users perceived the attractive agents as more persuasive and effective sales agents than the less attractive agents. Hence, the attractiveness of the agent influenced the perception of, and increased the likeability of the agent. Nowak & Rauh (2008) have pointed out that participants perceived the more anthropomorphic avatars in their study to be considerably more attractive and credible, and preferred to be represented by these avatars online; whilst feminine avatars were perceived as being more attractive than masculine avatars, which is a similar finding to the online perception study (Chapter 3) within this thesis, where females were significantly perceived as being more attractive than male agents.

This point is again backed up by social interaction studies, which suggest that the strength of the attractiveness stereotype is more potent when applied to females than to males (Bar-Tal & Saxe, 1976), and that society in general views females as being more attractive than males (Cross & Cross, 1971). Hence, society places greater emphasis on female beauty, and the attributes that contribute in increasing her attractiveness (i.e. hair styles, make-up and revealing clothing); which are usually aimed towards pleasing the male audience. Thus, an explanation as to why there are more attractive female agents (Chapter 3) in comparison to males, may be due to the vast majority of agent designers being male; who would ideally prefer to look at attractive females rather than attractive males.

Further HCI studies on avatars (Vasalou et al., 2008) within Second Life have also shown that when participants had to create avatars which represented them in two scenarios: a romantic date and a birthday card for a family member. The results indicated that when participants created an avatar for a romantic date; they focused on making the avatar more attractive with a romantic background; whereas a more conservative avatar and background were created for the birthday card to a family member. Vasalou & Joinson (2009) conducted a similar study, where participants were found to create more attractive avatars to represent them in an online dating scenario, which made these participants feel more confident when interacting with members of the opposite gender, as well as being more willing to approach them online.

Messinger et al. (2008) also highlighted how participants will prefer to create avatars that are much more attractive than themselves, which aided them to behave in a different manner online; as they became more outgoing, extraverted, and greater risk

takers in this online environment. Langlois et al. (2000) has shown that in social interactions, attractive people tend to be more confident, extraverted, and aggressive in their behaviour than less attractive individuals. These findings clearly point out that human users have a clear preference for creating more attractive avatars than themselves, as this preference transcends from real life experiences where the advertising industry, media and culture have together played a significant role in influencing societies perception of beauty; which portrays attractive individuals to have more positive personality traits, successful lives, careers and love lives than unattractive individuals; promoting the 'what is beautiful is good' stereotype (Dion et al., 1972; Eagly et al., 1991; Feingold, 1992).

Ethnicity & Religious backgrounds – Designers should consider developing and increasing the number of agents that represent other ethnic backgrounds, cultures and faiths. The census study within Chapter 3 highlighted how there should be a greater variety of agents from various ethnic and religious backgrounds; as designers seem to be following an ethnocentric approach when developing agents, whereby the first non-white agents were only developed as late as 2003. Thus, the vast majority of agents seem to represent a white and European background, which may be a representation of the agent designers who are also from the same ethnic group.

The comments from the online perception study (Chapter 3) revealed that 11 participants wanted to see agents representing their ethnic background, or religion; such as a Latino, a Muslim female wearing a head scarf, or an Indian Sick wearing a turban. The debrief interviews in Chapter 4 did not point to any specific ethnic preference of an agent by the participants; but participants may have used caution when answering this question in the fear of being labelled a racist. However, 8 non-British participants did state that if a website was providing information regarding a certain country, such as Malaysia; then it would be more appropriate for a Malay looking agent to present this site rather than a blonde and blue eyed European looking agent; i.e. perhaps the agent ethnicity should also fit the context.

These findings agree with other studies (Nass et al., 2000; Baylor & Kim, 2003; Baylor et al., 2003; Baylor & Kim, 2004), which found that individuals prefer to interact with agents of a similar ethnic background to themselves; whereby they reacted more positively towards these agents by perceiving them as being more attractive and trustworthy. Additionally, social psychology studies (Ford, 1997; Nass et al., 2000)

have recognised that negative stereotypes and racist behaviour can be elicited when a target individual's ethnic background differs to their own. Thus, in order to address this negative bias; a sensible approach for designers would be to develop more agents of various ethnic backgrounds; and perhaps give the user a choice of agents of different ethnicities to select from and interact with.

Gender – There seems to be a general preference by both male and female participants to -interact with a female rather than a male agent. The comments within the online perceptions study (Chapter 3) showed that 22 participants, male and female, had a desire to view and interact with attractive female agents rather than male agent. One reason for this preference may be due to the fact that the female agents are significantly more attractive than the male agents; therefore if males are also developed to be just as attractive as the female agent then perhaps their acceptance by participants and users will also increase.

On the other hand, an interesting point made by 83% of the participants (male and female alike), within the interviews in Chapter 4, is their general preference to interact with a female agent; was due to their perception of females as being more trustworthy, soft spoken, and more visually appealing than male agents. Additionally, 66% of participants viewed women to be happier and friendlier than men; whilst 93% of male participants stated how they enjoyed their daily interactions more with females than with males. Male participants (N = 12) highlighted the fact that when they enter a shop in their daily life; they will search for the most attractive female sales assistant and approach them; and so this would also be reflected in their attitude towards aesthetically pleasing agents.

Furthermore, female participants (N = 11) also commented on how they would prefer to interact with a female agent as they felt closer to other females in daily life, and could relate to them more than to a human male or male agent. However, the question of trust alters the participants' view point when agents are placed in the context of a more technical set-up; for example, 93% of male participants highlighted the fact that they would view the advice of a male agent as being more credible and trustworthy than that of a female agent within an aerospace/automobile related, electronic or technical computer hardware online store. Hence, the reason being that in daily life, males tend to dominate these areas.

Age – The census study pointed to a strong tendency for designers to create young adult agents which may be due to a bias towards older adults by these designers; as social psychology studies have highlighted how individuals can hold negative stereotypes regarding older age groups and the elderly as being worthless members of society whose special needs tend to go unattended; whilst the quality of youth is highly regarded and thought of as an asset to society (Nelson, 2004; Hogg & Vaughan, 2005).

Only a handful of HCI studies, for example, by Baylor et al. (2006) and Cowell & Stanney (2005) have investigated the participant's age preference for agents. Both studies highlight a strong preference by participants to interact with an agent from a similar age group to their own (young); whilst the young looking agents were classed as cool looking in comparison to older agents, which were labelled as being less cool but more experienced and wise at the same time. A possible explanation as to why users may prefer to interact with agents of a similar age group could be due to a reflection of their real lives; whereby these individuals feel more comfortable amongst their own peer group, as individuals are influenced more by their own in-group members; which in this instance would be an individual of a similar age group (Nelson, 2004).

Participant interviews (Chapter 4) revealed that the age of the agent, as stated by 80% of participants, was important depending upon the context in which the agent was being used. The general feeling amongst the participants was that younger agents were more preferable to interact with than the older age groups. Therefore, further research by HCI researchers and designers is necessary to establish whether younger adults are more acceptable to users within any context; or if this preference is context dependent.

Realistic/Human like appearance – A more realistic looking agent is preferred by users than an unrealistic cartoon like agent. The interviews in Chapter 4 showed that approximately 87% of participants appreciated the use of realistic and human-like agents, as they felt they could associate and relate to them more than to a talking household object or animal. A number of participants (N = 20) also suggested that the use of human-like or cartoon-like agents would depend on the website (i.e. the context). For example, using a cartoon-like agent would not be appropriate on a banking website, as this would remove the credibility of the site. In the same way, if an agent in the form of a little mouse was giving the user advice for an online wedding planner site; this would certainly be deemed as unsuitable; i.e. the agent must fit the context.

The use of more realistic looking avatars is also supported by Nowak & Rauh (2005, 2008); which revealed that participants perceived the more anthropomorphic and realistic avatars to be considerably more attractive and credible, and preferred to be represented by these avatars in online environments. Subsequently, Koda and Maes (1996) investigated the affects of applying a face and facial expressions onto an interface, showing that participants perceived a realistic looking face as being more likable, engaging and intelligent. Whilst, Baylor & Kim (2004) concluded that the more realistic images of pedagogical agents had a greater positive impact on the transfer of learning on the students. Additionally, Luo et al. (2006), highlighted how the participants perceived human-like characters to be more appropriate, trustworthy, and likeable than cartoon-like characters.

However, developing realistic agents should be practised with caution; as agents that are too realistic may cause the 'uncanny valley' effect (Mori, 1970); which might explain why some of the participants evaluating the online agents (Chapter 3) felt frightened when observing some of the agents presented before them. The uncanny valley effect was discovered in the area of human-robot interaction, where it triggers a feeling of discomfort amongst the human observing a very realistic looking robot. This is in sharp contrast to the main findings within this thesis, and other HCI studies (Koda & Maes, 1996; Baylor & Kim, 2004; Nowak & Rauh, 2005; Lou et al., 2006), which suggest that in general users prefer realistic to unrealistic agents. Therefore, this preference or revulsion of very realistic entities may be due to a 3D effect. Hence, the presence of a robot, which is in 3D, may be the cause of the uncanny valley; as it would be feasible to suggest that a moving, talking and walking robot would have a greater impact on the human user's senses than a 2D image of a static or talking agent within the interface. Thus, a 2D image of an agent on screen would seem to be less intimidating than a 3D robot.

Additionally, when developing a realistic agent, designers should not compromise the attractiveness of the agent; as 5 participants pointed out (Chapter 3) that they would prefer to interact with a less realistic attractive agent over a very realistic and unattractive agent.

Bald agents – Designers are advised to avoid bald agents. Numerous participants (N = 25) pointed out within the interviews (Chapter 4), that they would find it disturbing as well as feel very uncomfortable interacting with a bald female agent than with a bald

male agent. The main reason being, that females are usually bald due to losing hair from a skin condition or from chemo/radio-therapy. It is more socially acceptable for males to be bald than it is for females; as males can still be viewed as being attractive with no hair; but the standards of attractiveness for females seem to be related to having a full head of hair. Similar findings were also observed from 54% of participants (in Chapter 3), who commented on facial features which made an agent look unattractive; by stating their dislike and discomfort in viewing bald female agents in the online agent perception study.

Eyes – The eyes play an important role in the evaluation of the attractiveness and realism levels of an agent. Hence, a large number of participants (N = 26) pointed out that the eyes were the most important feature of the face (Chapter 4). Additionally, participants acknowledged that the eyes played a large role into how realistic an agent was perceived; as beautiful and well designed eyes helped to increase the human-likeness of the agent. Furthermore, the comments from the online perception study (Chapter 3), revealed that 60% of participants felt the eyes played an important role in making an agent appear more attractive. These findings are also supported by human-human interaction studies, such as those by Cross & Cross (1971) which have shown that when subjects rated static images of other individuals; the main features that were focused on and preferred the most were the eyes, then the mouth or smile, the hair, skin colour, shape of nose, and finally the proportions of the face as a whole.

Smile – An agent with a positive expression, such as a smile, is perceived more positively by the human user. The comments from Chapter 3 revealed that 40% of participants viewed an agent as being more attractive when it was smiling. Whilst, in Chapter 4, participants (N = 27) highlighted the importance of a smiling agent, as this made them appear more welcoming and approachable. In the same way, when people enter stores they are more likely to approach and talk to a sales assistant who is smiling rather than ask for help from a miserable and unhappy looking assistant. Furthermore, social interaction studies have shown that individuals are drawn towards and positively evaluate other individuals who appear to be either cheerful or smiling (Cunningham, 1986; Rhodes, 2006).

Full/half body – Agents with either a full or half body were viewed more positively by 90% of participants (Chapter 4). Some participants (N = 17) commented on how they felt more of a connection with a full bodied agent, as it was more memorable and easier

to identify with; rather than just a floating head which was deemed as a rather disturbing sight, as stated by 70% of participants. For example, the online agent called Honest Johnny, made participants feel quite uncomfortable and suspicious of him, as he was presented on a website with just his head and shoulders being visible to the user. Thus, half bodied agents are also perceived as being acceptable by a number of participants (N = 10) as it gave the impression that the agent was standing behind a desk.

Voice – In terms of the agent voice, 90% of participants found an agent more desirable if it spoke with a soft and pleasant voice (Chapter 4). It was also regarded as an added bonus if the agent talked and moved its mouth in a synchronised manner; as participants viewed agents as being annoying when this was not the case (e.g. Anna the online help assistant for IKEA). The use of hand gestures together with speech also made the agent appear more realistic.

Agent roles **–** Agent researchers and designers need to increase the types of roles assigned to agents, as the most common role assigned to agents is that of a pedagogical agent; whilst the least common is that of a presenter and storyteller. Designers seem to be cocooning agents into stereotypical roles; which is that of a tutor; when in fact agents have the potential to play much more diverse roles in many areas: i.e. entertainment, e-commerce, news and weather presenters, or as virtual friends. A reason for this trend may be due to technology constraints, and so there may be other roles these agent designers would prefer the agents to play but the technology at this time may not be advanced enough to accomplish and achieve this. Child-like agents are heavily represented within the storyteller roles who are all essentially white. This again points to the ethnocentric approach and designer bias, whereby they seem to perceive the storyteller role as being best suited to child-like agents. On the other hand, child-like agents could be designated various other roles successfully; such as that of a pedagogical agent for kindergarten or school children, or as online games characters for this younger age group. The child-like agent group has also been predominantly cartoon-like in appearance, which may be due to designers assuming that a younger audience will prefer to interact with such child-like agents, whose preference would be a cartoon-like character due to the bulk of children's programmes being cartoons (i.e. Bugs Bunny or Mickey Mouse).

7.7 Conclusion

The thesis has shown that an agent can become a player within the social arena when endowed with a face. Hence, one can ascertain that agents have the power to elicit stereotypes (i.e. attractiveness), by influencing and changing the human user's perception, decision making and behaviour; which favours the attractive over the unattractive agent, merely based on a single physical cue, which is its attractiveness level. Therefore, the empirical work within this thesis supports the presence of the attractiveness stereotype in both the stranger-based context (Dion et al., 1972; Eagly et al., 1991; Feingold, 1992) and interaction-based context (Langlois et al., 2000).

This thesis has shown that the application of the CASA model can be extended to human-agent interaction successfully; as well as aiding in explaining the reaction to agents by human participants; especially when utilising strong and proven theories from social psychology; i.e. the attractiveness stereotype (Dion et al., 1972; Eagly et al., 1991; Feingold, 1992; Langlois et al., 2000). However, the two main overriding conclusions from CASA and the Media Equation is that 'people's responses to media are fundamentally social and natural,' and 'media experiences equal human experiences' (Reeves & Nass, 1996), may not apply universally to the human-agent context. Hence, the thesis provides evidence to suggest that although human user's responses to media (i.e. embodied agents) tend to be social and natural; these media experiences are **not** always equal to human experiences. For example, the negative and anti-social comments made by participants towards the unattractive agents whilst chatting to them (in Chapter 5), were not the type of comments these individuals would necessarily make in real life towards another human; as their uninhibited and anti-social behaviour can be explained by the flaming effect (Lea et al., 1992). Therefore, the CASA model should be applied with caution to the human-agent context.

On the other hand, many HCI designers anticipate a new era of human-embodied agent interaction, where spoken communication will be the main method of interaction; caution is also required by agent designers when creating these embodied entities; i.e. due to the flaming effect by users interacting with chat-bot agents. The findings within the thesis highlight the crucial impact of agent aesthetics on the human observer; and the importance of their visual cues (both physical and demographic attributes), which affect users' perception as well as behaviour towards the embodied agents in question.

Whilst, the results and guidelines reported within the thesis can befit HCI designers and developers of agents within online environments, e.g. as bank or financial advisors, as persuasive agents who reward people with positive feedback, or as health advisors providing social support to users by giving advice in order to persuade the user to change their habits or lifestyle; such as reduce smoking and drinking, or increasing frequency of exercise.

One can conclude that when making the interface more anthropomorphic and engaging, then the visual image of an agent is a vital key in directly influencing the human user; and especially the face, which plays a crucial role in directly impacting the impressions and stereotypes formed (Haake & Gulz, 2008). These points are significant in light of HCI'S commitment to improving the interface, by enhancing our knowledge of human-agent interaction, which will lead to an improvement in human-agent design. Thus, allowing for a more natural, effective, acceptable and pleasant encounter between the human and agent within any context.

References

Active History (2008). Retrieved November 2008, from http://www.activehistory.co.uk

Alice (2008). Retrieved November 2008, from http://alicebot.blogspot.com

Aronson, E., Wilson, T. D., & Akert, R. M. (2005). *Social psychology*. Upper Saddle River, NJ: Prentice Hall.

Ball, G., & Breese, J. (2000). Emotion and Personality in a Conversational Agent. In J. Cassell, J. Sullivan, S. Prevost, & E. Churchill (Eds.), *Embodied conversational agents* (pp. 189-219). Cambridge, MA: MIT Press.

Bar-Tal, D., & Saxe, L. (1976). Physical attractiveness and its relationship to sex role stereotyping**.** *Sex Roles*, 2, pp. 123-133.

Baylor, A. L., & Kim, Y. (2003). Validating pedagogical agent roles: Expert, motivator, and mentor. Paper presented at *the International Conference of Ed-Media*. Honolulu, Hawaii.

Baylor, A. L., & Kim, Y. (2004). Pedagogical agent design: The impact of agent realism, gender, ethnicity, and instructional role. *Proceedings of the Intelligent Tutoring Systems* (pp. 592-603). Maceió, Brazil: Springer-Verlag Berlin Heidelberg.

Baylor, A. L., & Plant, E. A. (2004). Pedagogical agents as social models: Challenging gender-related stereotypes of engineering. National *Science Foundation Grant Proposal* (funded).

Baylor, A. L., Rosenberg-Kima, R. B., & Plant, E. A. (2006). Interface agents as social models: the impact of appearance on females' attitude toward engineering. *Extended Abstracts CHI '06* (pp. 526-531). Montréal, QC, Canada: ACM.

Baylor, A. L., & Ryu, J. (2003). The API (Agent Persona Instrument) for assessing pedagogi-cal agent persona. Paper presented at *the International Conference of Ed-Media*. Honolulu, Hawaii.

Baylor, A. L., Shen, E., & Huang, X. (2003). Which Pedagogical Agent do Learners Choose? The Effects of Gender and Ethnicity. *Proceedings of E-Learn: World Conference on E-Learning in Corporate, Government, Healthcare, & Higher Education* (pp. 1507-1510). Phoenix, Arizona.

Berry, D. C., Butler, L. T., de Rosis, F. (2005). Evaluating a realistic agent in an advice-giving task. *International Journal of Human-Computer Studies, 65,* pp. 304-327.

Berscheid, E., & Walster, E. (1974). Physical attractiveness. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (pp.158-215). New York: Academic Press.

Bickmore, T., 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), pp. 711-735.

Bickmore, T., & Cassell, J. (2001). Relational Agents: A Model and Implementation of Building User Trust. *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI)* (pp. 396-403). Seattle, Washington.

Bickmore, T., & Picard, R. (2005). Establishing and maintaining long-term human-computer relationships. *.ACM Transactions on Computer Human Interaction (ToCHI), 12* (2), pp. 293-327.

Cash, T. F., *&* Janda, L. H. (1984*).* Eye of the beholder. *Psychology Today, 18 (*12*),* pp. 46–52.

Cassell, J. (2000). More than just another pretty face: Embodied conversational interface agents. *Communications of the ACM, 43* (4), pp. 70-80.

Cassell, J., & Miller, P. (in press). Is it Self- Administration if the Computer Gives you

Encouraging Looks? In F. G. Conrad & M. F. Schober (Eds.), *Envisioning the Survey Interview of the Future*. New York: John Wiley & Sons.

Cassell, J., Sullivan, J., Prevost, S., Churchill, E. (Eds.). (2000). *Embodied conversational agents.* Cambridge, MA: MIT Press.

Catrambone, R., Stasko, J., & Xiao, J. (2002). Anthropomorphic agents as a user interface paradigm: Experimental findings and a framework for research. *Proceedings of the 24th annual conference of the cognitive science society, CogSci* (pp. 166–171). Mahwah, NJ: Lawrence Erlbaum.

Catrambone, R., Stasko, J., Xiao, J. (2004). ECA as user interface paradigm. In Z. Ruttkay, & C. Pelachaud (Eds.), *From Brows till Trust: Evaluating Embodied Conversational Agents* (pp. 239–270). Netherlands: Kluwer Academic Publishers.

Chaiken, S. (1979). Communicator Physical Attractiveness and Persuasion. *Journal of Personality and Social Psychology, 37*, pp. 1387-1397.

Couper, M. P., Tourangeau, R., & Steiger, D. M. (2001). Social presence in web surveys. In *CHI'2001 Conference Proceeding* (pp. 412-417). New York: ACM Press.

Cowell, A. J., & Stanney, K. M. (2005). Manipulation of non-verbal interaction style and demographic embodiment to increase anthropomorphic computer character credibility. *International Journal of Human-Computer Studies, 62* (2), pp. 281-306.
Cross, J. F., & Cross, J. (1971) Age, sex, race, and the perception of facial beauty. *Developmental Psychology, 3*, pp. 433-439.

Cunningham, M. R. (1986). Measuring the Physical in Physical Attractiveness: Quasi-Experiments on the Sociobiology of Female Facial Beauty. *Journal of Personality and Social Psychology, 50* (5), pp. 925-935.

De Angeli, A., & Brahnam, S. (2008). I hate you! Disinhibition with virtual partners. *Interacting with Computers, 20* (3), pp. 302-310.

De Angeli, A., Brahnam, S., & Wallis, P. (2005). Abuse: The darker side of human-computer interaction. *Interact 2005 Adjunct Proceedings*, pp. 91-92.

De Angeli, A., Brahnam, S., Wallis, P., & Dix, A. (2006). Misuse and abuse of interactive technologies. *Extended. Abstracts CHI 2006*, pp. 1647-1650. New York: ACM Press.

De Angeli, A., Johnson, G.I., Coventry, L. (2001). The unfriendly user: exploring social reactions to chatterbots, in: M. Helander, H.M. Khalid, P.O. Tham (Eds.), *International Conference on Affective Human Factors Design,* Asean Academic Press, London.

Dehn, M. D., & Mulken, V. S. (2000) The impact of Interface agents: a review of empirical research. *International Journal Human-Computer Studies, 52*, pp. 1-22.

De Meuse, K. P. (1987). A Review of the Effects of Nonverbal Cues on the Performance Appraisal Process. *Journal of Occupational Psychology, 60*, pp. 207- 226.

Dermer, M., & Thiel, D. L. (1975). When beauty may fail. *Journal of Personality and Social Psychology, 31* (6), pp. 1168-1176.

Dion, K., Berscheid, E., & Walster, E. (1972). What is beautiful is good. *Journal of Personality and Social Psychology, 24*, pp. 285-290.
DiPaola, S. (2002). FaceSpace: A Facial Spatial-Domain Toolkit. *Sixth International Conference on Information Visualisation.*

Don, A., Brennan, S., Laurel, B., and Shneiderman, B. (1992). Anthropomorphism: from Eliza to Terminator 2. *Proceedings CHI'92 Human Factors in Computing Systems* (pp. 67-70). Monterey: ACM.

Eagly, A. H., Ashmore, R. D., Makhijani, M. G., & Longo, L.C. (1991). What Is Beautiful Is Good: A Meta-Analytic Review of Research on the Physical Attractiveness Stereotype. *Psychological Bulletin, 110* (1), pp. 109-128.

Eagly, A. H., & Mladinic, A. (1989). Gender stereotypes and attitudes toward women and men. *Personality and Social Psychology Bulletin, 15*, pp. 543-558.

Feingold, A. (1992). Good-looking people are not what we think. *Psychological Bulletin, 111* (2), pp. 304-341.

Fink, B., Neave, N., Manning, J. T., & Grammer, K. (2006). Facial symmetry and judgements of attractiveness, health and personality. *Personality and Individual Differences, 41* (3), pp. 491–499.

Fiske, S. T. (1993). Social cognition and Social Perception. *Annual Review of Psychology, 44*, pp. 155-94.

Fiske, S. T., & Taylor, S. E. (Eds.). (1991). *Social cognition* (2nd ed.). New York: McGraw-Hill.

Fogg, B. J. (1998). Persuasive Computers: Perspectives and Research Directions. *Proceedings of the CHI 1998 Conference of the ACM/SIGCHI* (pp. 225-232). New York: ACM Press.

Fogg, B. J. (1999). Persuasive technologies - Now is your chance to decide what they will persuade us to do - and how they'll do it. *Communications of the ACM, 42*, pp. 26-29.

Fogg, B. J. (2003). *Persuasive Technology*. Amsterdam: Morgan Kaufmann Publishers.

Ford, T. E. (1997). Effects of Stereotypical Television Portrayals of African-Americans on Person Perception. *Social Psychology Quarterly, 60* (3), pp. 266-275.

Gulz, A., & Haake, M. (2006). Design of animated pedagogical agents - A look at their look. *International Journal of Human-Computer Studies, 64*(4), pp. 322-339.

Haake, M. & Gulz, A. (2008). Visual Stereotypes and Virtual Pedagogical Agents. *Educational Technology & Society,* 11 (4), pp. 1-15.

Hogg, M. A., & Vaughan, G. (Eds.). (2005). *Introduction to Social Psychology*. Australia: Pearson Education.

Holtgraves, T. M., Ross, S.J., Weywadt, C.R., & Hans, T.L. (2007). Perceiving artificial social agents. *Computers in Human Behaviour, 23* (5), pp. 2163-2174.

Holzwarth, M., Janiszewski, C., & Neumann, M. M. (2006). The Influence of Avatars on Online Consumer Shopping Behavior. *Journal of Marketing, 70* (4), pp. 19-36.

Hone, K. (2006). Empathic agents to reduce user frustration: The effects of varying agent characteristics. *Interacting with Computers, 18* (2), pp. 227-245.

Human Synergistics (2007). Desert Survival. Retrieved June 2009, from http://www.humansynergistics.co.uk/site/default.aspx.

Khan, R., & De Angeli, A. (2007). Mapping the Demographics of Virtual Humans. *Proceedings of British HCI* (pp.149 – 152). Lancaster, UK: ACM.

Khan, R., & De Angeli, A. (2009). The attractiveness stereotype in the evaluation of embodied conversational agents. *Proceedings of INTERACT 2009 International Conference on Human-Computer Interaction* (pp. 405-417). Uppsula, Sweden: Springer.

Koda, T., & Maes, P. (1996). Agents with Faces: The Effect of Personification. *Proceedings of the Fifth IEEE International Workshop on Robot and Human Communication (RO-MAN '96).*

Langlois, J. H., Kalakanis, L., Rubenstein, A. J., Larson, A., Hallam, M., & Smott, M. (2000). Maxims or myths of beauty? A meta-analytic and theoretical review. *Psychological. Bulletin, 126,* pp. 390–423.

Lea, M., O'Shea, T., Fung, P., & Spears, R. (1992). Flaming in computer-mediated communication. In Lea, M. (Ed.), *Contexts in Computer-Mediated Communication* (pp. 89 112). London, UK: Harvester Wheatsheaf.

Loftus (2008). Retrieved November 2008, from http://www.loftusphoto.com

Luo, F. T., McGoldrick, P., Beatty, S., & Keeling, K. A. (2006). On-screen characters: their design and influence on consumer trust. *Journal of Services Marketing, 20* (2), pp. 122-124.

Messinger, P. R., Ge, X., Stroulia, E., Lyons, K., Smirnov, K., & Bone, M. *(2008).* On the relationship between My Avatar and Myself. *Journal of Virtual World Research, 1* (*2*), pp. 1–17.

Moreno, R., & Flowerday, T. (2006). Students' choice of animated pedagogical agents in science learning: A test of the similarity-attraction hypothesis on gender and ethnicity. *Contemporary Educational Psychology, 31*(2), pp. 186-207.

Mori, M. (1970). Uncanny valley. *Energy, 7* (4), pp. 33-35.

Morkes, J., Kernal, H. K., & Nass, C. (1999). Effects of humour in task-oriented huma-computer interaction and computer-mediated communications: a direct test of SRCT theory. *Human-Computer Interaction, 14* (4), pp. 395-435.

Mullennix, J. W., Stern, S. E., Wilson, S.J., & Dyson, C. (2003). Social perception of male and female computer synthesized speech. *Computers in Human Behavior, 19,* pp. 407-424.

Nass, C., Isbister, K., & Lee, E. J. (2000). Truth is beauty: Researching embodied conversational agents. In J. Cassell, J. Sullivan, S. Prevost, and E. Churchill (Eds), *Embodied Conversational Agents*. Cambridge, MA: MIT Press.

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues, 56* (1), pp. 81–103.

Nass, C., Moon, Y., Fogg, B. J., Reeves, B., & Dryer, D. C. (1995). Can computer personalities be human personalities? *International Journal of Human–Computer Studies, 43,* pp. 223–239.

Nass, C., Moon, Y., Green, N. (1997). Are computers gender-neutral? Gender stereotypic responses to computers. *Journal of Applied Social Psychology*, *27*, pp. 864–876.

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. *Proceedings of the Conference on Human Factors in Computing Systems CHI '94* (pp. 72–78), New York, USA: ACM.

Nelson, T. D. (Ed.). (2004). *Ageism: Stereotyping and Prejudice against older persons.* Cambridge, MA: MIT Press.

Nowak, K. L., & Rauh, C. (2005). The influence of the avatar on online perceptions of anthropomorphism, androgyny, credibility, homophily, and attraction. *Journal of Computer-Mediated Communication, 11* (1), pp. 153-178.

Nowak, K. L., & Rauh, C. (2008). Choose your 'buddy icon' carefully: The influence of avatar androgyny, anthropomorphism and credibility in online interactions. *Computers in Human Behaviour*, *24* (4), pp. 1473–1493.

Pallant, J. (2007). SPSS Survival Manual: A step by step guide to data analysis using SPSS (3rd ed.). New York: Open University Press.

Pandorabots (2008). Retrieved November 2008, from http://showcase.pandorabots.com/

Pratt, J. A., Hauser, K., Ugray, Z., & Patterson, O. (2007). Looking at human-computer interface design: Effects of ethnicity in computer agents. *Interacting with Computers, 19* (4), pp. 512–523.

Prestia, S., Silverston, J., Wood, K., & Zigarmi, L. (2002). The Effects of Attractiveness on Popularity; an Observational Study of Social Interaction Among College Students. *Perspectives in Psychology, 5,* pp. 3-11.

Reeves, B., & Nass, C. (1996). *The Media Equation: How people treat computers, television, and new media like real people and places*. Cambridge, UK: Cambridge University Press.

Regan, P. C., & Berscheid, E. (1997). Gender differences in characteristics desired in a potential sexual and marriage partner. *Journal of Psychology and Human Sexuality, 9,* pp. 25–37.

Rhodes, G. (2006). The evolutionary psychology of facial beauty. *Annual Review of Psychology, 57,* pp. 199-226.

Rickel, J., & Johnson, W. L. (2000). Task oriented collaboration with embodied agents in virtual worlds. In J. Cassell, J. Sullivan, S. Prevost, and E. Churchill (Eds), *Embodied Conversational Agents*. Cambridge, MA: MIT Press.

Shneiderman, B., & Maes, P. (1997). Direct manipulation vs. interface agents. *Interactions: New Visions for Human-Computer Interaction, 4* (6), pp. 42-61.

Spencer, B. A. & Taylor, G. S. (1998). Effects of facial attractiveness and gender on causal attributions of managerial performance. *Sex Roles, 1,* pp.273-285.

Sproull, L., Subramani, M., Kiesler, S., Walker, J., & Waters, K. (1996). When the interface is a face. Human-computer interaction, 11 (2), pp. 97-124.

Taylor, S. E., Fiske, S. T., Etcoff, N., & Ruderman, A. (1978). The categorical and contextual bases of person memory and stereotyping. *Journal of Personality and Social Psychology, 36* (7), pp. 778–793.

Vasalou, A., Joinson, A. N., Banziger, T., Goldie, P., & Pitt, J (2008). Avatars in social media: Balancing accuracy, playfulness and embodied messages. *International Journal of Human-Computer Studies, 66,* pp. 801–811.

Vasalou, A., & Joinson, A. N. (2009). Me, myself and I: The role of interactional context on self presentation through avatars. *Computers in Human Behaviour, 25,* pp. 510–520.

Veletsianos, G., Scharber, C. & Doering, A. (2008). When sex, drugs, and violence enter the classroom: conversations between adolescent social studies students and a female pedagogical agent. *Interacting with Computers, 20* (3), pp. 292–301.

Walker, J. H., Sproull, L., & Subramani, R. (1994). Using a Human Face in an Interface. *Proceedings of CHI '94* (pp. 85-91). New York, USA: ACM.

Walster, E., Aronson, V., & Abrams, D. (1966). Importance of physical attractiveness in dating behaviour. Journal of Personality and Social Psychology, 4 (5), pp. 508-516

Wheeler, L., & Kim, Y., (1997). What is Beautiful is Culturally Good: The Physical Attractiveness Stereotype has Different Content in Collectivistic Cultures. *Personality and Social Psychology Bulletin, 23*, pp. 795–800.

Williams, J. E. & Bennett, S. M. (1975). The Definition of Sex Stereotypes via the Adjective Check List. *Sex Roles, 1* (4), pp. 327 – 337.

Yee, N., & Bailenson, J. (2007*).* The proteus effect: The effect of transformed self-representatioin on behavior. *Human Communication Research, 33*, pp. 271–290.

Zanbaka, C., Goolkasian, P., & Hodges, L. (2006). Can a virtual cat persuade you? The role of gender and realism in speaker persuasiveness. *Proceedings of CHI '06* (pp. 1153-1162). Montréal, QC, Canada: ACM.

Appendices

Appendix A: Observer scoring sheet for initial reliability test

Task: You are to observe the 30 agents presented on the monitor screen in front of you. For each agent note down which category of Age, Ethnicity and Anthropomorphism, listed below, you would assign to the agent in question.

Age:

- *Child* – An individual between birth and puberty;
- *Young Adult* – An individual between puberty/teens and the age of 30;
- *Adult* – An individual Between ages of 30 and 50; and
- *Older Adult* – An individual over 50.

Ethnicity:

- *White* – Faces originating from Caucasian/European background;
- *Black* – Faces originating from African background;
- *Asian* – Faces originating from South Asian background; and
- *Oriental* – Faces originating from the Far East.

Anthropomorphism:

- *Cartoon* – faces which do not represent real people. They can be sketches, or humorous images often displaying some exaggeration of facial characteristics (caricatures);
- *Drawing* – 2 dimensional representational images featuring human-like faces;
- *Mannequin* –3 dimensional representational images of human-like faces; and
- *Photo realistic* - Pictures of real human beings or artificial faces which are extremely human like, so that they could be erroneously attributed to a real person.

Agent 1) Age –	Ethnicity-	Embodiment-
Agent 2) Age –	Ethnicity-	Embodiment-
Agent 3) Age –	Ethnicity-	Embodiment-
Agent 4) Age –	Ethnicity-	Embodiment-
Agent 5) Age –	Ethnicity-	Embodiment-
Agent 6) Age –	Ethnicity-	Embodiment-
Agent 7) Age –	Ethnicity-	Embodiment-
Agent 8) Age –	Ethnicity-	Embodiment-
Agent 9) Age –	Ethnicity-	Embodiment-
Agent 10) Age –	Ethnicity-	Embodiment-
Agent 11) Age –	Ethnicity-	Embodiment-
Agent 12) Age –	Ethnicity-	Embodiment-
Agent 13) Age –	Ethnicity-	Embodiment-
Agent 14) Age –	Ethnicity-	Embodiment-
Agent 15) Age –	Ethnicity-	Embodiment-
Agent 16) Age –	Ethnicity-	Embodiment-

Agent 17) Age –	Ethnicity-	Embodiment-
Agent 18) Age –	Ethnicity-	Embodiment-
Agent 19) Age –	Ethnicity-	Embodiment-
Agent 20) Age –	Ethnicity-	Embodiment-
Agent 21) Age –	Ethnicity-	Embodiment-
Agent 22) Age –	Ethnicity-	Embodiment-
Agent 23) Age –	Ethnicity-	Embodiment-
Agent 24) Age –	Ethnicity-	Embodiment-
Agent 25) Age –	Ethnicity-	Embodiment-
Agent 26) Age –	Ethnicity-	Embodiment-
Agent 27) Age –	Ethnicity-	Embodiment-
Agent 28) Age –	Ethnicity-	Embodiment-
Agent 29) Age –	Ethnicity-	Embodiment-
Agent 30) Age –	Ethnicity-	Embodiment-

Appendix B: Observer scoring sheet for second reliability test with new categories

Task: You are to observe the 30 agents presented on the monitor screen in front of you. For each agent note down which of the three categories of Anthropomorphism, listed below, you would assign to the agent in question.

Anthropomorphism:

- ***Cartoon*** – Faces which are clearly based on humans which can be sketches or humorous images often displaying some exaggeration of facial characteristics (caricatures);
- ***Mannequin*** – Representational images of human-like faces which are not cartoon like, and at the same time cannot be mistaken for a real person; and
- ***Photo realistic*** - Pictures of either artificial faces or real human beings which are extremely human like, so that they could be erroneously attributed to a real person.

Agent 1) Embodiment-	Agent 16) Embodiment-
Agent 2) Embodiment-	Agent 17) Embodiment-
Agent 3) Embodiment-	Agent 18) Embodiment-
Agent 4) Embodiment-	Agent 19) Embodiment-
Agent 5) Embodiment-	Agent 20) Embodiment-
Agent 6) Embodiment-	Agent 21) Embodiment-
Agent 7) Embodiment-	Agent 22) Embodiment-
Agent 8) Embodiment-	Agent 23) Embodiment-
Agent 9) Embodiment-	Agent 24) Embodiment-
Agent 10) Embodiment-	Agent 25) Embodiment-
Agent 11) Embodiment-	Agent 26) Embodiment-

Agent 12) Embodiment- Agent 13) Embodiment- Agent 14) Embodiment- Agent 15) Embodiment-	Agent 27) Embodiment- Agent 28) Embodiment- Agent 29) Embodiment- Agent 30) Embodiment

Appendix C: Pearson's correlations for remaining agent groups and surveys

Pearson's Correlations for remaining agent groups ([a] Pearson's Correlation Coefficient-Effect Size. * Correlations sig. at 0.001 level two-tailed; ** at 0.01 level, two tailed).**

Agent Group names of attractiveness by realism correlations	*r* [a]	*N*
All agents	.657(**)	145
All survey 1 agents	.755(**)	29
All survey 2 agents	.440 (***)	29
All survey 3 agents	.705 (**)	29
All survey 4 agents	.647 (**)	29
All survey 5 agents	.705 (**)	29

Pearson's correlations for male and female agents in survey 1, where N = 96 ([a] Pearson's Correlation Coefficient-Effect Size: **Correlations sig. at 0.01 level, two tailed).

Female agent correlations		Male agent correlations	
Agent Label	***r*[a]**	**Agent Label**	***r*[a]**
F_Ch_W_Car_B45	0.246(**)	M_YA_W_Car_CS5	0.270(**)
F_YA_W_Man_D11	0.416(**)	M_OA_W_Phot_D75	0.297(**)
F_YA_B_Phot_A31	0.480(**)	M_OA_W_Car_E1	0.425(**)
F_YA_W_Phot_F44	0.525(**)	M_YA_Or_Phot_D41	0.433(**)
F_YA_W_Man_F2	0.535(**)	M_YA_W_Man_A40	0.467(**)
F_YA_W_Phot_E22	0.542(**)	M_Ch_As_Car_D9	0.482(**)
F_YA_B_Man_E38	0.546(**)	M_Ch_W_Car_A2	0.500(**)
F_YA_W_Car_D55	0.556(**)	M_YA_W_Man_C57	0.544(**)
F_YA_W_Car_A81	0.581(**)	M_A_W_Man_B52	0.570(**)
F_YA_W_Man_B66	0.589(**)	M_A_W_Man_C02	0.571(**)
F_YA_W_Car_D10	0.605(**)	M_YA_W_Car_F11	0.681(**)
F_A_W_Man_A98	0.622(**)		
F_YA_W_Man_B33	0.624(**)		
F_YA_W_Man_B95	0.632(**)		
F_YA_W_Man_E67	0.638(**)		
F_YA_W_Man_E46	0.667(**)		
F_YA_W_Man_F14	0.682(**)		
F_YA_W_Phot_D43	0.683(**)		

Pearson's correlations for male and female agents in survey 2, where N = 103 ([a] Pearson's Correlation Coefficient-Effect Size. **Correlations sig. at 0.01 level, two-tailed).

Female agent correlations		Male agent correlations	
Agent Label	***r*[a]**	**Agent Label**	***r*[a]**
F_YA_W_Man_F75att	0.443(**)	M_A_W_Car_J40att	0.352(**)
F_Ch_W_Car_G76att	0.481(**)	M_A_W_Man_F45att	0.398(**)
F_YA_W_Car_I47att	0.516(**)	M_A_W_Man_F67att	0.416(**)
F_YA_W_Man_J4att	0.554(**)	Mv_A_W_Phot_H60att	0.419(**)
F_A_W_Car_F71att	0.566(**)	M_A_W_Phot_H61att	0.428(**)
F_A_W_Man_G86att	0.566(**)	M_Ch_W_Car_G2att	0.472(**)
F_YA_W_Car_I65att	0.568(**)	M_YA_B_Man_G19att	0.536(**)
F_YA_W_Man_I2att	0.575(**)	M_YA_Or_Car_G28att	0.546(**)
F_YA_W_Man_H38att	0.600(**)	M_YA_W_Car_H43att	0.571(**)
F_A_W_Man_H71att	0.609(**)	M_YA_W_Car_J05att	0.573(**)
F_YA_B_Phot_G70att	0.621(**)	M_YA_W_Man_F50att	0.603(**)
F_YA_W_Man_G74att	0.642(**)	M_YA_W_Man_F66att	0.648(**)
F_YA_W_Car_J61att	0.644(**)	M_YA_W_Man_G84att	0.683(**)
F_YA_B_Man_J58att	0.652(**)	M_YA_W_Phot_G8att	0.723(**)
F_YA_W_Man_H59att	0.679(**)		

Pearson's correlations for male and female agents in survey 3, where N = 82 ([a] Pearson's Correlation Coefficient-Effect Size. * Correlations sig. at 0.001 level, two tailed; ** at 0.01 level, two tailed).**

Female agent correlations		Male agent correlations	
Agent Label	***r*[a]**	**Agent Label**	***r*[a]**
F_Ch_W_Car_K58att	0.226(**)	M_Ch_W_Car_K91att	0.263(***)
F_YA_W_Man_L80att	0.523(**)	M_YA_W_Phot_J88att	0.321(**)
F_YA_W_Man_K33att	0.532(**)	M_A_Or_Man_J64att	0.362(**)
F_YA_W_Car_K84att	0.553(**)	M_YA_W_Car_M2att	0.363(**)
F_YA_B_Man_K92att	0.573(**)	M_A_W_Car_K8att	0.398(**)
F_OA_W_Man_L12att	0.594(**)	M_YA_W_Man_K2att	0.410(**)
F_YA_W_Man_L61att	0.617(**)	M_YA_B_Car_M15att	0.472(**)
F_YA_W_Man_K06att	0.643(**)	M_YA_W_Man_K57att	0.506(**)
F_YA_B_Man_L58att	0.658(**)	M_YA_W_Man_K90att	0.506(**)
		M_YA_W_Car_L77att	0.513(**)
		M_YA_W_Man_L91att	0.520(**)
		M_A_W_Man_J67att	0.526(**)
		M_A_W_Man_K82att	0.530(**)
		M_YA_B_Man_K88att	0.549(**)
		M_YA_W_Man_K31att	0.549(**)
		M_YA_B_Man_K7att	0.564(**)
		M_A_W_Man_M9att	0.577(**)
		M_A_W_Man_J83att	0.591(**)
		M_YA_As_Car_L9att	0.607(**)
		M_YA_W_Car_M4att	0.665(**)

Pearson's correlations for male and female agents in survey 4, where N = 69
([a] Pearson's Correlation Coefficient-Effect Size. * Correlations sig. at 0.001 level, two-tailed; ** at 0.01 level, two-tailed; * at 0.05 level, two tailed).**

Female agent correlations		Male agent correlations	
Agent Label	***r*[a]**	**Agent Label**	***r*[a]**
F_YA_W_Phot_M54att	0.500(**)	M_Ch_W_Car_Q21att	0.298(*)
F_YA_W_Car_O13att	0.504(**)	M_A_W_Man_R60att	0.304(***)
F_YA_W_Phot_P7att	0.528(**)	M_Ch_W_Car_R52att	0.380(**)
F_YA_W_Man_N77att	0.530(**)	M_A_W_Car_P73att	0.454(**)
F_YA_W_Man_S4att	0.548(**)	M_YA_W_Car_M92att	0.480(**)
F_YA_W_Man_P83att	0.551(**)	M_YA_W_Man_N33att	0.498(**)
F_YA_B_Man_M61att	0.578(**)	M_A_W_Man_N63att	0.512(**)
F_Ch_W_Man_M25att	0.599(**)	M_A_W_Car_M68att	0.575(**)
F_YA_W_Car_P2att	0.611(**)	M_YA_W_Car_P31att	0.575(**)
F_YA_W_Phot_P42att	0.612(**)	M_YA_W_Man_N22att	0.597(**)
F_YA_W_Phot_R92att	0.614(**)	M_YA_W_Man_R77att	0.605(**)
F_YA_W_Phot_S39att	0.637(**)	M_YA_W_Man_S23att	0.640(**)
F_YA_W_Phot_N48att	0.677(**)	M_YA_B_Man_M52att	0.646(**)
		M_YA_W_Car_O57att	0.671(**)
		M_YA_W_Man_P88att	0.703(**)
		M_YA_W_Man_N6att	0.761(**)

Pearson's correlations for male and female agents in survey 5, where N = 195
([a] Pearson's Correlation Coefficient-Effect Size. ** Correlations sig. at 0.01 level, two-tailed).

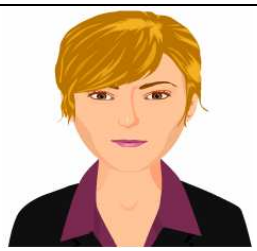
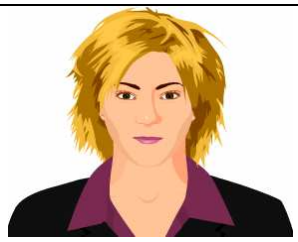
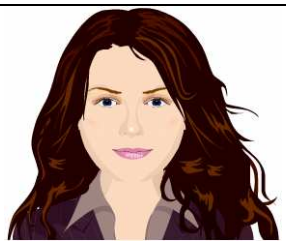
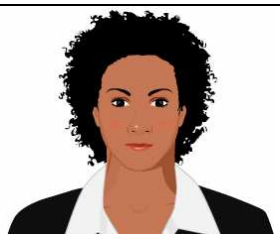
Female agent correlations		Male agent correlations	
Agent Label	***r*[a]**	**Agent Label**	***r*[a]**
F_A_W_Car_U71att	0.323(**)	M_Ch_W_Car_S79att	0.252(**)
F_A_W_Phot_Z11att	0.440(**)	M_Ch_W_Car_Y71att	0.282(**)
F_YA_W_Man_V25att	0.447(**)	M_A_W_Car_Y47att	0.296(**)
F_YA_B_Car_W19att	0.481(**)	M_OA_W_Man_V04att	0.322(**)
F_YA_B_Man_V6att	0.487(**)	M_A_W_Phot_Z41att	0.335(**)
F_YA_B_Car_Z37att	0.489(**)	M_YA_As_Man_X91att	0.358(**)
F_Ch_W_Man_T50att	0.492(**)	M_Ch_W_Car_Z28att	0.366(**)
F_YA_As_Car_U19att	0.492(**)	M_A_B_Car_U97att	0.385(**)
F_YA_W_Phot_T52att	0.506(**)	M_YA_W_Man_T99att	0.407(**)
F_YA_W_Man_V18att	0.518(**)	M_YA_W_Man_T25att	0.427(**)
F_YA_W_Man_S45att	0.552(**)	M_YA_W_Man_V51att	0.447(**)
F_YA_W_Man_T80att	0.555(**)	M_YA_W_Phot_T46att	0.455(**)
F_YA_B_Phot_W02att	0.566(**)	M_A_W_Car_X57att	0.458(**)
F_YA_W_Phot_W21att	0.582(**)	M_YA_B_Phot_X3att	0.529(**)
		M_OA_W_Man_Z9att	0.540(**)

Pearson's correlations for female and male agent sub-groups (Correlations sig. at 0.01 level two-tailed).**

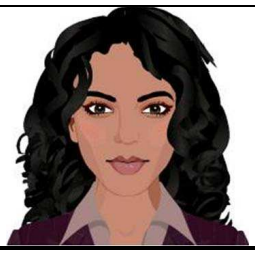
A) Female agent group names for perceived attractiveness by realism correlations	***r***	***N***
Female mannequin & white agents of all age groups	.657(**)	31
Female mannequin agents	.662(**)	37
Female cartoon-like agents	.723(**)	17
Female photorealistic agents	.729(**)	15
Female agents	.733(**)	69
Female mannequin & white young adults agents	.766(**)	25
Female white agents	.737(**)	57
Female young adult agents	.741(**)	57
Female white young adult agents	.748(**)	45

B) Male agent group names for perceived attractiveness by realism correlations	***r***	***N***
Male non young adult agents	.518(**)	34
Male mannequin & white agents of all age groups	.522(**)	31
Male white agents	.551(**)	63
Male cartoon-like white agents of all age groups	.567(**)	25
Male agents	.571(**)	76
Male mannequin & young adult of all ethnic groups	.579(**)	24
Male mannequin & white young adult agents	.583(**)	19
Male mannequin agents	.605(**)	37
Male young adult agents	.644(**)	42
Male adult agents	.651(**)	21
Male white young adult agents	.678(**)	32
Male cartoon- like agents	.698(**)	30

Appendix D: Female agents used in pilot study

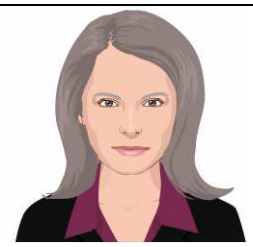
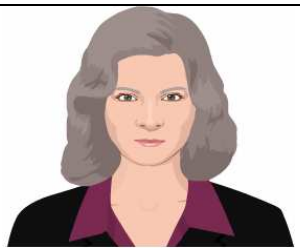
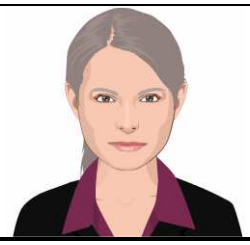
Attractive	Average	Unattractive
Young white female agents		
Set 1		
		
Attractiveness: 4.62 ***Realism: 4.38***	***Attractiveness: 3.65*** ***Realism: 3.25***	***Attractiveness: 2.92*** ***Realism: 3.12***
Set 2		
		
Attractiveness: 4.98 ***Realism: 4.78***	***Attractiveness: 3.86*** ***Realism: 3.36***	***Attractiveness: 2.86*** ***Realism: 3.31***
Young black female agents		
Set 3		
		
Attractiveness: 4.76 ***Realism: 4.28***	***Attractiveness: 4.40*** ***Realism: 3.98***	***Attractiveness: 2.76*** ***Realism: 3.28***

Set 4

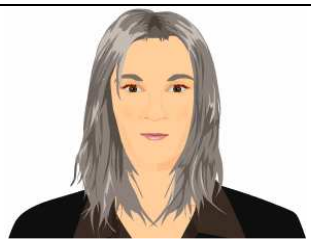
		
Attractiveness: 4.22 ***Realism: 3.98***	***Attractiveness: 3.24*** ***Realism: 3.14***	***Attractiveness: 2.82*** ***Realism: 2.98***
		
Attractiveness: 4.47 ***Realism: 4.18***		

Older white female agents

Set 5

		
Attractiveness: 3.37 ***Realism: 3.97***	***Attractiveness: 2.57*** ***Realism: 2.67***	***Attractiveness: 1.96*** ***Realism: 2.66***
		
Attractiveness: 3.31 ***Realism: 3.86***		

Set 6

		
Attractiveness: 3.24 ***Realism: 2.98***	***Attractiveness: 2.72*** ***Realism: 2.68***	***Attractiveness: 1.95*** ***Realism: 2.55***

Appendix E: Observer key to conversations for reliability checks

Task: Read through the following conversation key to help you code the two conversations that will be given to you (on screen) between the participant and embodied chat-bot agent. This will help you to code each participant input. Each input has a total of four empty columns next to it, representing four categories. Hence, you are to code each participant input four times. For each category select one of the codes you feel is appropriate for that participant input. The four categories are as follows:

Category 1: *Direction*

- **Question** - E.g. What is your name?
- **Response** - E.g. Very well thanks / Yes I do.
- **Statement** – E.g. Hi / Bye / Ok. Also includes commands: e.g. Go away! / You will obey me! / Get Lost!

Category 2: *Valency*

- **Positive** – E.g. I love your hair/ You make me laugh / I like my job / Lets go on a date (i.e. any comment which has positive emotion behind it, that is not negative or neutral, can contain positive words within a positive context, could also be flirtatious or complimentary)
- **Negative** – Inputs with negative emotion behind them. Can be commands, insults, including sexual or racist remarks. E.g. I hate you! / You are an idiot! / Shut up! / Pi** Off!
- **Neutral –** These inputs have neither positive nor negative emotion behind them. They can be statements, about relationships, beliefs or daily life, including testing the agent's knowledge and reasoning. E.g. Do you believe in God? / I have 5 sisters / What is the capital of Spain?

Category 3: *Sentence Discourse*

- **API (ask for personal Information) –** e.g. How are you? / Do you have a family? (i.e. asking about personal life, feelings, thoughts, likes and dislikes).
- **AGI (ask for general Information) –** e.g. What is the weather like? / What is the capital of Spain?
- **SD (self disclosure) –** e.g. Own age, sex, height, life, job, education, clothes etc. (i.e. giving personal details about themselves).
- **GPO (give personal opinion)** – e.g. I think that's wonderful / I don't like him / I'm tired / hahaha (i.e. personal feelings, thoughts, likes, dislikes, or laughter).
- **GGI (give general Info) –** e.g. There are 5 departments / Liverpool is 25 miles from Manchester (i.e. factual information not based on personal opinions and thoughts).
- **Test –** e.g. What is the capital of Spain? / What is 45 + 300? (i.e. where the participant tests the knowledge or capability of the bot-agent)
- **Command –** e.g. Shut up! / Stop talking! / Listen to me!
- **ST (small talk) –** e.g. Hi / Hello / See you / Bye / Thanks / Ok / Good.

Category 4: *Topic*

- **Relationships** – e.g. Are you married? / Do you have a mother? (i.e. friends, family, girl/boyfriend, marriage, husband/wife).
- **Humour/laugh** – e.g. haha / lol! (i.e. jokes, lol, hahaha).
- **Challenge** – e.g. How dare you! / That is silly / Are you a dumb robot? / You must listen to me, I am human (i.e. being shocked, unimpressed, confusion, disagreement).
- **R/M/C (Religious/moral/creation) –** e.g. Do you believe in God? / Who created you?
- **Offensive –** e.g. You look awful! / Your nose is a mess! / You are an idiot! **(**i.e.. being unpleasant or sarcastic)
- **Expletive** – e.g. Why don't you pi** off!! (i.e. using swear words Sh**, Fu**, etc.).
- **Sexually explicit –** e.g. Will you have sex with me?
- **Flirtation –** e.g. Come on, can I have your number?
- **Compliment –** e.g. You are pretty? / I love your hair!
- **Racist –** e.g. You P*k*!
- **Conversational –** e.g. Hello / How are you? (all other inputs which cannot be grouped in the above. These including hellos and goodbyes).

Appendix F: Publication 1

Khan, R., & De Angeli, A. (2007). Mapping the Demographics of Virtual Humans. *Proceedings of British HCI* (pp.149 – 152). Lancaster, UK: ACM.

Mapping the demographics of virtual humans

Rabia Khan
School of Informatics
The University of Manchester
PO Box 88, Manchester M60 1QD
United Kingdom
Tel: +44 161 306 1291
E-mail: Rabia.Khan@postgrad.manchester.ac.uk

Antonella De Angeli
School of Informatics
The University of Manchester
PO Box 88, Manchester M60 1QD
United Kingdom
Tel: +44 161 306 1291
E-mail: Antonella.De-angeli@manchester.ac.uk

ABSTRACT
This paper presents a census of 147 virtual agents, by examining and reporting on their physical and demographical characteristics. The study shows that the vast majority of agents developed are from a white ethnic background. Overall, female agents tend to be more photo realistic than their male counterparts who are more cartoon like. These findings highlight current stereotypes in relation to agents and contribute to a deeper understanding of virtual worlds.

Categories and Subject Descriptors
Human Factors
General Terms
Design and Human Factors.

Keywords

Embodiment, agents, age, gender, race.

INTRODUCTION
In the last few years, virtual bodies have become increasingly prevalent in HCI (for example, embodied conversational agents, ECA's, and avatars). ECA's are defined as being synthetic characters (full body, graphical or physical simulations of people) that can maintain a conversation with a user [1]. An ample amount of research on ECA's has been concerned with how to emulate human conversation following the assumption that ECA's will have the same properties as humans in face to face conversation [2]. This line of research has led to the definition of relational agents, as computational artifacts designed to build long term, social-emotional relationships with their users [3]. There are many domains which could benefit from the deployment of relational agents; such as in online shopping, e-learning, advice giving, behavioural change therapy, helping people to stop smoking or dieting, counseling, or coaching them [4].
The debate on anthropomorphism and its implications in designing agents with more human-like qualities has been going on for quite some time. Walker et al. [5] found that people spent more time interacting with a talking face display than text-only interface. Sproull et al. [6] showed that users were more positive in their response to a face by spending more time with it than with a text only version, where users quickly got bored. Reeves and Nass [7] clearly identify several benefits of the anthropomorphic approach by concluding that people respond to computer agents in fundamentally the same social ways as they would to another person.

Published by the British Computer Society
Volume 2 Proceedings of the 21st BCS HCI Group Conference

A recent trend in anthropomorphic design has seen an increase in research on the effect of demographic and physical appearance variables of virtual agents. According to De Meuse's [8] taxonomy, a number of non-verbal variables affect face-to-face communication. These variables can be broken up into those cues that are behavioural in nature and those which are not. Non behavioural actions are *demographic variables* (ethnicity, age and gender) and *physical appearance variables* (clothing/attire, bodily and facial attractiveness). Demographic variables are not under an individual's control, whereas physical appearance cues can be subject to rapid change. Hence, cues such as hair/eye colour, cosmetics, clothing style can all affect social reactions to an individual's or an agent's. [9].
When looking at the importance of such demographic elements in embodiment, studies have shown that users prefer interacting with agents that either match their own ethnicity, or agents that are young looking [9]. The design of pedagogical agents' ethnicity and gender do influence learner perception of agent personality, motivational qualities, and perceived influence on the learning process. Students also perceived agents of the same ethnicity to be more engaging and affable. In particular, African-American learners were more likely to choose a pedagogical agent of the same ethnicity, and have a positive attitude towards this chosen agent after the lesson [10].
Baylor and Kim [11] draw attention to the impact of demographic variables and realism of pedagogical agents on learners. The findings suggest that students had a greater transfer of learning when agents were more realistic, and when the agents were represented non-traditionally (as black versus white) in the 'expert' role. The more realistic looking agents positively affected transfer of learning. Students which worked with the Black Expert agents found this quite novel, and thus paid more attention to the black agents than the white expert counterpart (the 'novelty effect').
As regards to gender, Hone [12] suggests that a female agent is more effective than a male agent in reducing frustration. Hence, frustration reduction is improved when an agent is embodied. Furthermore, a study showing female agents acting as a non-traditional engineer (e.g. very attractive and outgoing) significantly enhanced student interest in engineering as compared to a more stereotypical 'nerdy' version (e.g. homely and very introvert) [13]. Female learners have been reported to prefer and choose a cartoon like pedagogical agent (as opposed to realistic looking agents) more often than their male counterparts [10]. Despite this growing corpus of evidence suggesting a significant role of physical variables of virtual embodiments, at present, little research has evaluated the demographic characteristics of existing agents. In this paper, we report a census of virtual agents by looking at physical characteristics of existing ones. The main aim of this study was to determine what type of demographic and physical variables were commonly or rarely being assigned to agents.

METHOD

A database of 147 virtual faces was analysed. These faces were collected by conducting internet searches in online journals and conference proceedings (ACM library and Science Direct), search engines (Google Scholar) and Conference sites (IVA: Intelligent Virtual Agents conference from 2003) using the following keywords: Embodied Conversational Agents (ECA's), Synthetic Agents, Social Agents, Conversational Agents, Virtual Agents, Virtual human, Agents and Avatars. Several e-mails were also sent to mailing lists (British HCI, CHI Announcements, CHI Students) and individual researchers to invite them to share pictures of Agents/Avatars they had utilised in their research.[1] Agents were selected based on the following criteria: (a) Human like (No animal characters), (b) Frontal view only, and (c) Good quality image (at least 10 x 10 cm). Each agent was assigned a unique ID and recorded in a database system (Microsoft Access). The following attributes were researched and recorded in relation to each agent: Gender, Age, Ethnicity, Dressing Style, Profession, Anthropomorphism level, and Name. A coding/classification system was developed by the authors for each of the mentioned attributes. The source of each agent was recorded including details of the paper and authors who utilized/developed them in their research.

Framework of analysis

The framework of analysis was developed following an iterative process to accommodate different agent characteristics. Categories were refined and modified during the process. Double coding was conducted for 20% of the data-base yielding a reliability of almost 90%. All faces were coded according to a number of demographic and physical appearance variables [8]. Gender was divided into two categories: *Male and Female.* The remaining variables are discussed as follows:

Age consisted of four distinguished categories:

- *Child* – An individual between birth and puberty;
- *Young Adult* – An individual between puberty/teens and the age of 30;
- *Adult* – An individual between ages of 30 and 50; and
- *Older Adult* – An individual over 50.

Ethnicity consisted of 4 categories:

- *White* – Faces originating from Caucasian/European background;
- *Black* – Faces originating from African background;
- *Asian* – Faces originating from South Asian background; and
- *Oriental* – Faces originating from the Far East.

Dressing style was divided into 4 groups:

- *Casual* – Informal clothing and not dressy;
- *Formal* – Designed for wear or use in certain occasion/event or role;
- *Uniform* – A job specific outfit; and
- *Missing* – No outfit is visible, only face and neck displayed

Embodiments were also clustered into four broad categories according to their **level of anthropomorphism**:

- *Cartoon* – faces which do not represent real people. They can be sketches, or humorous images often displaying some exaggeration of facial characteristics (caricatures);
- *Drawing* – 2 dimensional representational images featuring human-like faces;
- *Mannequin* –3 dimensional representational images of human-like faces; and
- *Photo realistic* - Pictures of real human beings or artificial faces which are extremely human like, so that they could be erroneously attributed to a real person.

Profession consisted of 5 main roles (the role source was where the agents were retrieved from as described in the paper):

- *Pedagogical Agent* – Agent that facilitate the learning process;
- *Actor* – Performs the role of a character within a scenario;
- *Storyteller* – A narrator of anecdotes, incidents, or fictitious tales;
- *Assistant* – Agent who assists, supports, guides and helps the user; and
- *Presenter* – Agent which presents/read out the daily news and weather forecast.

Name was classed into two groups:

- *Name* – Agent with a personal human like name (such as Peter, and Lucia); and
- *No Name* – Agent with no human like name.

RESULTS

Queries were used in order to collect data. The focus was on gender in comparison to the other attributes previously stated. Virtual embodiments were evenly divided between males (n=73) and females (n=74). Table 1 reports the frequency values for male and female agents as a function of their ethnicity. The vast majority of these agents were white (84%). The remaining ones were Black (n=17), Oriental (n=1), and Asian (n=5).

Table 1. Gender by Ethnicity frequency distribution

	Male	Female	Total
Whites	61	63	124
Others	12	11	23
Total	**73**	**74**	**147**

A trend analysis indicated that non-white agents started to appear in 2004 [14], and are growing fast in number since then [10, 15].

Table 2 compares the frequency values for male and female agents as a function of their Age. It appears that the bulk of agents are young adults. There is an interaction between age and gender, where Adults and Older Adults are largely made up of males, and Young Adults are predominantly female.

[1] The reason for concentrating on academic sources is due to the need for limiting the scope of our project, and looking at innovative design solutions which will constitute the future internet populations of virtual avatars.

Table 2. Gender by Age frequency distribution

	Male	Female	Total
Child	7	4	11
Young Adult	46	62	108
Adult	11	5	16
Older Adult	9	3	12
Total	**73**	**74**	**147**

Data on Dressing Style are summarised in Table 3. Agents were chiefly in Casual dress with no specific gender effect. An equal number of male and female agents could not be attributed any Dress Style (Missing), while only 2% of agents were dressed in a uniform.

Table 3. Gender by Dressing Style frequency distribution

	Male	Female	Total
Casual	42	46	88
Formal	13	11	24
Uniform	2	1	3
Missing	16	16	32
Total	**73**	**74**	**147**

A vast number of agents (75%) did not possess a name. Leaving 25% of agents with names, out of which almost half were assigned a role and the other half without. When looking at Age by Role it was surprising to see that over 90% of Child agents were Actors. Amongst Young Adults, Adults and Older Adults, the Pedagogical role was the most frequently assigned. Investigating Age by Anthropomorphism level highlighted over 90% of Child agents having a Cartoon like face. Amongst the other three age groups, each of the Anthropomorphic levels were distributed evenly.

Figure 1 illustrates frequency values for male and female agents as a function of Anthropomorphism. It is evident that a larger number of male agents are Cartoon like in comparison to female agents. Conversely, the number of female agents significantly increases as the realism of an agent increases. Photo realistic agents are primarily female thus showing a specific gender effect.

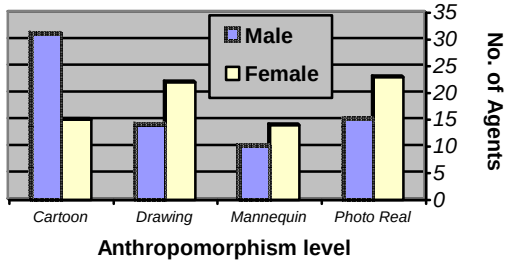


Figure 1. Gender by Anthropomorphism

Figure 2 represents agents that had a role assigned to them. Almost 46% of agents had no defined role showing no specific gender effect. The most common profession for a virtual agent is that of a tutor. The roles which are least common are those of a Storyteller and Presenter.

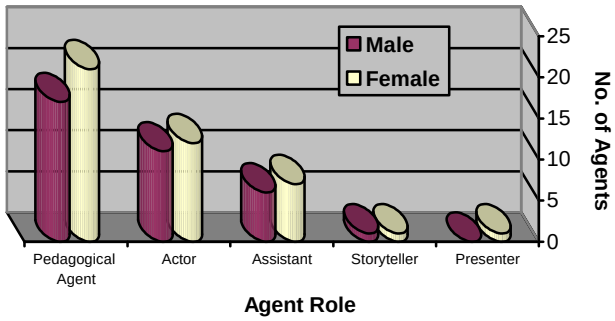


Figure 2. Gender by Role

CONCLUSION

The findings of our study add to a growing body of literature on the effect of agents' demographics on user perception, by presenting a census of this virtual world. This census can give an insight into the stereotypes of virtual agents held by their designers and is useful to predict biases and difficulties in the interaction with these virtual agents.

Our study indicates that females are in general more photorealistic and anthropomorphic than their male counterparts who are usually represented as cartoon like agents. This suggests that more emphasis is focused on the female appearance. Studies have shown [16] that the physical attractiveness stereotype is more potent when applied to women than to men, and females are generally seen as being more attractive than males. Thus the real-life bias which expects females to look more attractive than males may lead to the reason why the female agents are more realistic than the male counterpart.

Earlier research has shown [9, 10, 17] that users in general prefer to interact with agents of a similar ethnicity to their own. The results in this study clearly show that the vast majority of agents are from a white background, leading to a large mismatch between potential users and available agents. This may be due to the bulk of designers coming from a principally white background, as we have only analysed publications in English. Yet, we believe that this finding highlights a prevailing ethnocentric approach to agent design, which may strongly hamper their global adoption.

A point designers need to consider is that the presence of a wholly white agent world with a handful of agents from other ethnic backgrounds could increase the tendency of racist behaviour towards the non white agents. Long ago, social psychology has posited a clear link between discrimination and minorities. More agents from various ethnic backgrounds should be developed in order to counteract this issue.

The vast majority of agents are young adults, and only a handful are classed as children. The reason for developing more younger adult agents may be the designer's view that the vast majority of users are also young adults who may prefer to interact with agents of a similar age group, backing Cowell et al. study [9]. On the contrary, these days users range from nursery children to the old aged pensioners. Ideally, these age groups should also be considered when developing agents. Yet, the results in this study indicate a minimal number of child and older aged agents being used.

The most prevalent type of role assigned to an agent is that of a pedagogical one. Perhaps this is the role researchers see most fitting for an agent; as a tutor, advisor and guide. Agents can play far more diverse roles, rather than being cocooned

into the pedagogical role. Further work needs to be done to assign more agents to other roles and professions like a news/weather presenter, storyteller, online sales assistant and so on. This study highlights Child like agents predominantly playing an acting role; researchers should be aware of the benefits of assigning other roles to them (such as a kindergarten tutor or storyteller). The implications of these findings for design are:

- The function and role of an agent must be acknowledged and high on the functional spec agenda before commencing the development of an agent. Different embodiment may fit different roles, and it is important to clarify this relationship.
- What kind of user will be interacting with the agent? For example, are the users young or old? What is their gender as well as their cultural and ethnic background? Thus, agents can be modified according to the users that will interact with them.
- Allowing the user to choose from a drop down list of options as to what age, gender, and race they would prefer their agent to possess before interaction. This is more crucial in light of HCI's commitment to interfaces that are equally accessible and acceptable to all intended users.
- Agent designers should take greater care when choosing how to represent the agent's ethnicity, gender, and realism.

Further research will be conducted into the effects of the physical appearance of agents on user behaviour. In particular, the effect of agent facial attractiveness on the user perception has yet to be investigated

ACKNOWLEDGMENTS

We thank the many researchers who helped in the construction of our agent data-base by contributing their exemplars.

REFERENCES

[1] Olivier, P. (2004). Gesture Synthesis in a Real-World ECA, In: Elisabeth André, Laila Dybkjær, Wolfgang Minker, Paul Heisterkamp (Eds.): Affective Dialogue Systems, Kloster Irsee, Germany, *Proceedings.*

[2] Cassell, J. (2000). Embodied Converstional Interface Agents. *Communications of the ACM*, 43, 70-78.

[3] Bickmore, T. (2003). Relational Agents: Effecting Change through Human-Computer Relationships, MIT Ph.D. Thesis.

[4] Bickmore, T. and Picard, R. (2005) "Establishing and Maintaining Long-Term Human-Computer Relationships" *ACM Transactions on Computer Human Interaction (ToCHI)*, 12(2): 293 – 327.

[5] Walker J. et al. (1994). Using a Human Face in an Interface. *Proceedings of CHI '94*, ACM Press, 85-91.

[6] Sproul et al. (1996). When the interface is a face. *Human-Computer Interaction*, 11, 97-124.

[7] Reeves, B., & Nass, C. (1996). The Media Equation: How people treat computers, television, and new media like real people and places, *Cambridge: Cambridge University Press.*

[8] De Meuse, K. P. (1987). A Review of the Effects of Nonverbal Cues on the Performance Appraisal Process. *Journal of Occupational Psychology*, 60, 207- 226.

[9] Cowell, A. J. & Stanney, K. M. (2005). Manipulation of non-verbal interaction style and demographic embodiment to increase anthropomorphic computer character credibility *International Journal of Human-Computer Studies, Volume 62, Issue 2, Pages 281-306*

[10] Baylor, A. L., Shen, E., & Huang, X. (2003). Which Pedagogical Agent do Learners Choose?The Effects of Gender and Ethnicity. Paper presented at the E-Learn (*World Conference on E-Learning in Corporate, Government, Healthcare, & Higher Education*), Phoenix, Arizona.

[11] Baylor, A. L. & Kim, Y. (2004). Pedagogical Agent Design: The Impact of Agent Realism, Gender, Ethnicity, and Instructional Role. *Presented at International Conference on Intelligent Tutoring Systems*, Maceio, Brazil, p 592-603.

[12] Hone, K., (2006). Empathic agents to reduce user frustration: The effects of varying agent characteristics. *Interact. Comput*. 18, 227–245.

[13] Baylor, A. L. (2004). Encouraging more positive engineering stereotypes with animated interface agents. Unpublished manuscript.

[14] Morency, L. P., and Darrell, T. (2004). From Conversational Tooltips to Grounded Discourse: Head Pose Tracking in Interactive Dialog Systems, *International Conference on Multimodal Interfaces*, pp. 32-37, College State, PA, October

[15] Cassell, J., & Miller, P. Is it Self-Administration if the Computer gives you Encouraging Looks? (in press) In F.G. Conrad & M.F. Schober (Eds.), *Envisioning the Survey Interview of the Future.* New York: John Wiley & Sons.

[16] Cross, J. F., & Cross, J. Age, sex, race, and the perception of facial beauty. *Developmental Psychology*, 1971, 3, 433-439.

[17] Moreno, R., & Flowerday, T. (2006). Students' choice of animated pedagogical agents in science learning: A test of the similarity-attraction hypothesis on gender and ethnicity. *Contemporary Educational Psychology*, Volume 31, Issue 2, Pages 186-207

Appendix G: Publication 2

Khan, R., & De Angeli, A. (2009). The attractiveness stereotype in the evaluation of embodied conversational agents. *Proceedings of INTERACT 2009 International Conference on Human-Computer Interaction* (pp. 405-417). Uppsula, Sweden: Springer.

The Attractiveness Stereotype in the Evaluation of Embodied Conversational Agents

Rabia Khan and Antonella De Angeli
Manchester Business School, University of Manchester, Booth Street West, Manchester, M15 6PB, UK
Rabia.Khan@postgrad.manchester.ac.uk, Antonella.De-Angeli@mbs.ac.uk

Abstract. Physical attractiveness is an important cue for social interaction. Psychology studies have long shown that physical attractiveness can elicit positive personality attributions as well as positive behaviour towards other people. This effect is explained by the attractiveness stereotype. In this paper, we investigate whether this stereotype apply to the interaction with virtual agents. We report the results of two experiments where the attractiveness stereotype was tested with and without interaction with the agent. Results indicate a strong effect of the attractiveness stereotype, showing that users tend to form and maintain a better evaluation of attractive agents than of unattractive ones independent of actual interaction with the agent or the agents' ethnicity. Implications for design are discussed.

Keywords: Embodied conversational agents, user evaluation, virtual bodies

1 Introduction

Since the media equation paradigm posited a link between computers and social actors [1], a large corpus of research has investigated the role of social cognition in HCI [2, 3]. Embodied Conversational Agents (ECA's) are a favourite target for this type of research as their anthropomorphic aspect tends to elicit social inference. There is evidence that virtual bodies carry with them stereotypical attributions and that users respond differently to ECA's based on their gender [4, 5], age [5, 6], and ethnicity [2, 5]. Stereotypes are widely shared generalisations about people as members of a social group, whereby group members are attributed similar characteristics on the basis of the categories to which they belong regardless of actual variation [8]. Stereotypes strongly influence social behaviour by providing default setting information for perception and action. Stereotype-based expectations are also believed to shape personality development due to social pressure [9].

In this paper, we investigate the effect of the attractiveness stereotype on the perception of, and behaviour with, ECA's. According to the attractiveness stereotype, nice looking people are perceived as more socially competent, more intelligent, friendlier, and more successful in life than less attractive people. Results of two experiments provided strong evidence that users apply the attractiveness stereotype in the evaluation of ECA's, independently of actual interaction with the agent or its ethnicity.

2 Related Work

Despite popular belief that "beauty is in the eye of the beholder", social scientists have demonstrated that attractiveness is defined by social consensus both within and across cultures and independently of gender [10]. Three main factors make a face attractive: symmetry, averageness and sexual dimorphism [11]. The more symmetrical a face is, then the more attractive it is perceived to be. Averageness refers to typicality of traits constituting a face, whereas sexual dimorphism signals the reproductive potential and sexual maturity of an individual. Research on sexual dimorphism suggests that attractive feminine traits corresponds to large eyes, high eyebrows, full lips, small nose, small chin, prominent cheekbones and narrow cheeks. Whilst, masculine traits such as square chins, thin lips, small eyes, and thick brows tend to signal dominance and status which enhance their mating value.

One of the earliest evidence of the attractiveness stereotype was reported by Dion and colleagues [12]. The authors asked participants to rate three photographs of fellow undergraduate students in terms of personality traits and behavioural characteristics. The stimuli differed on physical attractiveness: one picture represented an unattractive face, the second an average looking face, and the third an attractive face.

Participants consistently attributed to attractive individuals more socially desirable traits than to unattractive individuals. Attractive individuals were also deemed to lead better lives in terms of occupational success and relationship satisfaction than their unattractive counterparts. Since this work, a large corpus of psychological research has investigated the reliability of the attractiveness stereotype. This research can be differentiated into three main streams according to objectives and methodology.

The first stream focused on the definition of the content of the stereotype adopting the stranger-attribution paradigm [7, 9]. Participants were invited to rate personality traits, as well as behavioural, social and emotional characteristics of hypothetical individuals depicted in photographs, sometimes enriched by minimal written information. Two independent meta-analyses of the stranger-attribution literature confirmed the strength of the attractiveness stereotype independently of gender and age of both evaluators and targets [7, 9]. Both meta-analyses concentrated on North American participants and highlighted very similar trait components of the stereotype [13]. Large to medium sized effects were found on all dimensions related to social behaviour, confirming that attractive people are perceived as possessing better social skills, and being more popular and more extrovert than unattractive individuals. Large effects also appeared in the perception of sexual warmth, suggesting that attractive people, and in particular attractive females, are perceived as being more sexually responsive. Medium sized effects were evinced for dimensions related to cognitive skills and dominance: attractive people are perceived as more intelligent, rational and bright, as well as being more dominant and assertive than unattractive people. No effect of attractiveness was found on character perception (e.g., trustworthiness, sincerity and honesty) or on concern for other. Negative effects emerged on modesty, implying that attractive people are perceived as vainer than unattractive people.

The second stream of research [9] investigated the objectivity of the attractiveness stereotype via correlational research looking at the relationship between self-rated attractiveness and measures of personality, social skills and mental ability. A meta-analysis of this research provided evidence in favour of the attractiveness stereotype only with regard to personality traits related to social behavior (e.g., loneliness, self-consciousness and social anxiety), social behavior measures (e.g., number of friends and popularity with the opposite sex), and self-reported measures of sexual permissiveness.

The third research stream extended the stranger-attribution literature to more ecologically valid situations [10] with studies of social interaction, whereby the attractiveness stereotype was measured after actual interaction with a target. A set of meta-analyses revealed the persistence of the attractiveness stereotype even when the perceiver could make an informed judgment [10]. Attractive individuals (child or adult) were evaluated and treated more favourably than unattractive individuals by other people, even by those who knew them. These meta-analyses also revealed that attractive adults and children tended to display significantly more positive behavior than unattractive individuals.

Several theoretical frameworks have been invoked to explain the attractiveness stereotype. Fitness-related evolutionary theories posit that attractiveness is linked to health and reproduction fitness [10]. On the contrary, social expectancy theories stress the influence of socialization mechanisms, claiming that expectations about an attractive person influence people's interaction with that person who eventually change their self-perception and behavior in line with the social expectations [10]. Although no individual theory seems to explain the complexity of the effect, there is no doubt that attractiveness is a powerful and cultural independent cue driving interaction [10, 13].

A growing number of studies have investigated social affordances of 'virtual bodies', showing that their demographics subtlety affect user behaviour. For example, people tended to be more influenced by a virtual agent of the opposite sex [4] and preferred interacting with an agent of their same ethnicity [2, 5]. Attractive agents were regarded by users as being more persuasive and effective sales agents in purchasing goods [14]. A relevant corpus of research has addressed the perception of avatars, as mediators of human-human interaction in virtual environments [15, 16]. Results showed that people tended to perceive feminine avatars as being more attractive than masculine avatars, and anthropomorphic avatars as being more credible and attractive [15]. In online dating environments, users tended to create more attractive avatars, and participants represented by attractive avatars were more willing to approach members of the opposite gender [17, 18]. Investigations into user behaviour in Second Life [19] have found that users report on making their avatars not only similar, but also somewhat more attractive than themselves. Such users with avatars that are more attractive than their real selves reported being more outgoing, extrovert, loud, and risk-takers in Second Life than in real life [19].

3 Experiment 1

This experiment was designed to test the reliability of the attractiveness stereotype in the evaluation of ECA's within the stranger-attribution paradigm. It applied as closely as possible the procedure proposed by [12] for stimuli selection and testing but it used pictures of agent faces instead of real faces. Similarly to the original study, which selected photographs from a University yearbook, the agent faces were selected from a large data-base of agent embodiments used in ECA research [6]. Contrary to [12], however, only female faces were tested as the variance in attractiveness and realism of male faces was more limited and did not allow proper differentiation between stimuli. Based on the face-to-face literature, we formulated the following hypothesis: (H1) *Attractiveness will affect the initial impressions on embodied conversational agents: the more attractive the virtual agent, the more positive the user evaluation.*

3 .1 Method

Participants and Design. A total of 30 students at the University of Manchester (15 Male, 15 Female) took part in the experiment. Approximately 36% of participants were 18-25 years old, and the rest were between 26 and 35 years old. Attractiveness (3) was manipulated within–subjects. All participants evaluated three agents (attractive, average and unattractive).

Stimulus Materials. Six pictures of young female agents were used as stimuli in the study. These pictures were previously rated for attractiveness and realism by 545 independent evaluators. The 6 pictures were selected following the procedure applied in [12]. The 6 agents were assigned to one of two sets, each containing one attractive, one average and one unattractive face (Table 1).

Table 1. Agents used in the study.

Agent Set 1		
		
Attractiveness: 1.68	**Attractiveness:** 3.09	**Attractiveness:** 4.63
Realism: 1.98	**Realism:** 3.71	**Realism:** 4.39
Agent Set 2		
		
Attractiveness: 2.02	**Attractiveness:** 3.23	**Attractiveness:** 4.15
Realism: 2.39	**Realism:** 3.04	**Realism:** 3.37

The following selection criteria were applied: (*a)* human-looking faces from a white ethnic background; (*b*) high inter-rater agreement as to their physical attractiveness; (*c*) faces representing the very attractive and the very unattractive target were not at the extreme end of the attractiveness distribution; (*d*) faces had a neutral expression and; (*e*) neck and shoulders were displayed in the picture.

Procedure. The experiment took place in a usability laboratory. Participants were introduced to it as a study looking into the user's opinion of ECA's. Before the experiment, each participant was shown a short video

giving examples of four online agents available on the Internet and invited to provide comments on them. They were then shown one of the three pictures from either agent set 1 or set 2 on a computer screen and invited to evaluate them filling an on-line questionnaire. Presentation order was randomized and each picture was visible in a prominent position of the screen until the participants submitted the questionnaire. On completion, participants were presented with all three images of the agents they had evaluated and asked further questions about their physical appearance.

Dependent Variables. Participants were invited to record their impressions of each face along 7 dimensions. A measure of *physical attractiveness* was collected to validate the reliability of the experimental manipulation. It was measured by the relevant sub-scale of the Interpersonal Interaction Scale [20]. *Social competence* (unsociable – sociable, unfriendly – friendly, introvert – extrovert), *intellectual competence* (unintelligent – intelligent, emotional - rational, unambitious – ambitious), *social adjustment* (unstable – stable, immature – mature, poorly adjusted – well adjusted), *potency* (weak – strong, unassertive – assertive, submissive – dominant) and *integrity* (dishonest – honest, untrustworthy – trustworthy, insincere – sincere) were measured for hypotheses testing. These dimensions are well-known components of the attractiveness stereotype [13] and were used in this study as they may also apply to the evaluation of ECA's. The items within this investigation were taken from [13]. A*nthropomorphism* was measured by two likert-items (The Agent is human Like, The agent is machine like) from [21].

3.2 Results

Reliability analyses returned satisfactory results for each dimension tested in the study and each level of attractiveness (Cronbach alpha > 0.80). Seven indexes were computed averaging scores on individual items for each attractiveness level. Mean scores were entered as dependent variables into seven 3*2 mixed-design ANOVAs, with attractiveness (3) as within-subjects factor and agent-set (2) as between-subjects factor. Linear contrasts were run to test the difference between consecutive values of attractiveness based on a linear model [22]. Partial eta-squared (η^2) was computed as estimate of effect size. Partial η^2 = .01 indicate small effects, partial η^2 = .06 medium effects, and partial η^2 = 14 large effects [23].

Manipulation Check. The ANOVA on physical attractiveness returned a very strong effect for agent attractiveness ($F_{(2,56)}$ = 135.88, p < 0.001, partial η^2 = .83), and a significant interaction attractiveness * agent-set ($F_{(2,56)}$ = 12.29, p < 0.001, partial η^2 = .31). The interaction was due to the unequal distribution of attractiveness levels between the two agent-sets (Fig. 1). Although a significant linear trend was evident in each agent-set, the relative difference between attractiveness levels differed.

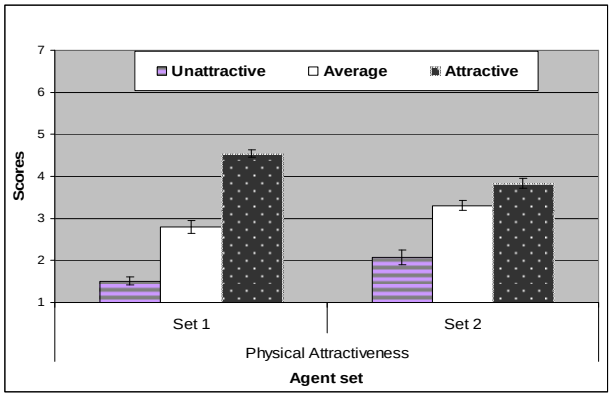
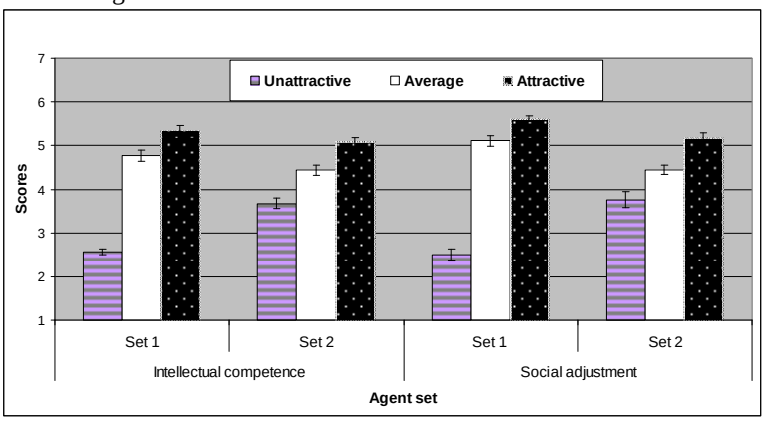


Fig. 1. Physical attractiveness scores as a function of experimental conditions.

Test of Hypotheses. Perception of *social competence* was strongly influenced by attractiveness, $F_{(2,56)}$ = 97.68, p < 0.001, partial η^2 = .77, with no interaction effect. Participants' evaluations increased linearly showing an improvement of about 1.5 points between consecutive levels of attractiveness.

The evaluation of *intellectual competence* and *social adjustment* revealed very similar patterns (Fig. 2). Both ANOVAs indicated a large main effect of attractiveness (in the order, $F_{(2,56)} = 51.99$, $p < 0.001$, partial $\eta^2 = .65$; $F_{(2,56)} = 45.52$, $p < 0.001$, partial $\eta^2 = .62$) and a significant 2-way interaction attractiveness * agent-set ($F_{(2,56)} = 7.22$, $p < 0.01$, partial $\eta^2 = .21$; $F_{(2,56)} = 9.19$, $p < 0.001$, partial $\eta^2 = .25$).
Although each agent set was affected by a significant linear trend, the increase between consecutive values of attractiveness was different. Set 2 followed a straight line, while the relative increase between the unattractive and the average looking agent in set1 was much larger than any other comparison, probably due to the lowest attractiveness scores of this agent.



. **Fig. 2.** Intellectual competence and social adjustment as a function of experimental conditions.

Participants evaluated more attractive agents systematically better on the *integrity* dimension, $F_{(2,56)} = 54.5$, $p < 0.001$, partial $\eta^2 = .66$. The amount of improvement differed between the two agent-sets as reflected by the significant interaction attractiveness * agent set, $F_{(2,56)} = 3.3$, $p < 0.05$, partial $\eta^2 = .12$ (Fig. 3). The evaluation of agents in set 1 was more strongly affected by attractiveness than the evaluation of agents in set 2. Both sets however returned significant results to the linear trend test.
The ANOVA on *potency* as dependent variable displayed a different trend of results, due to the large interaction effect attractiveness * agent-set, $F_{(2,56)} = 7.52$, $p < 0.05$, partial $\eta^2 = .21$. Set 1 followed the linear trend evinced in all other analyses. In contrast, the most unattractive agent of set 2 was assigned the highest level on potency.

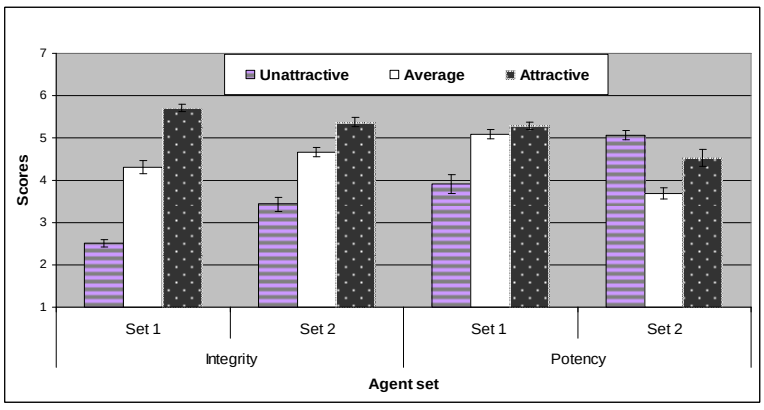


Fig. 3. Integrity and Potency scores as a function of experimental conditions.

Anthropomorphism returned only a large main effect for attractiveness ($F_{(2,56)} = 41.59$, $p < 0.001$, partial $\eta^2 = .59$). Both agent set 1 and 2 indicated a significant linear trend, but the increase varied between consecutive values of attractiveness.

3.3 Conclusion

The study provided strong support to the hypothesis that the attractiveness stereotype applies to ECA's, following the stranger attribution paradigm [7, 9, 13]. The components of the stereotype closely mirrored those of real humans. The large effect sizes for social competence, social adjustment and intellectual competence equalled the effect sizes reported in social psychology experiments. On the contrary, the effect size for integrity was much larger than that reported in studies with real human-beings. This suggests that the association between trustworthiness and beauty may be exasperated when the target is artificial and thus susceptible to limited attributions of intentionality.

The only exception to the attractiveness stereotype was the dimension of potency. This variable was found to be subject to variation also in studies with real human beings and it was hypothesised to reflect the North American stereotype which may not apply to collectivistic cultures [13]. The sample tested in our study was evenly split between Europeans and people from Eastern countries who may pay more attention to collectivistic values. However, the lack of effect on potency may also be due to the specific target analysed in this study. Indeed, it is reasonable to believe that potency assumes differential valence when applied to artificial agents or to real human beings. Users want ECA's to serve them [24]. Hence, in this context, dominance, assertiveness and strength may be associated to negative traits rather than to positive ones. This hypothesis was supported by several comments reported by participants during the final interview which highlighted difficulties in understanding and rating the dimension.

4 Experiment 2

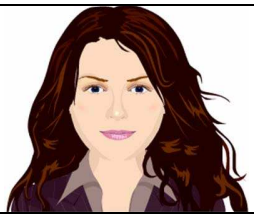
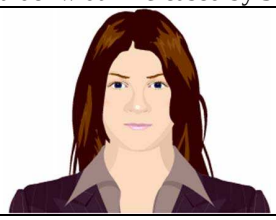
To address the reliability of the attractiveness stereotype towards ECA's in a more ecological setting, a second experiment was designed whereby evaluations were collected before and after actual interaction with the agent [10] Participants were invited to engage in a spontaneous conversation with an embodied chatterbot for 10 minutes. Given the very strong impact of attractiveness on first impression of ECA's evinced in experiment one, and following the interaction studies literature [10] we hypothesized that (H1) an *advantage of attractive agents over unattractive agents would be evident also after actual interaction with the agent.* However, we also expected that (H2) *the effect should be weaker after interaction as participants acquired more contextualized information to inform their evaluation.* Therefore, we predicted to find smaller differences between the evaluation of attractive and unattractive agents after usage. This decrease was expected to be associated to (H3) *a more negative view of the attractive agents after interaction*, as they may pay the price of the high expectations raised at first impressions.
To account for the problems evinced in study 1 using existing embodiments, the stimuli for experiment 2 were created by manipulating the appearance existing agents.

4.1 Method

Participants. Forty-eight students (21 female, and 27 male) at the University of Manchester participated in this experiment. Over 60% of participants were 26-35 years of age, and around 30% were 18-25. Participants were randomly assigned to experimental conditions in equal size groups.

Stimuli. Six Oddcast female agents of different races were systematically manipulated to decrease their physical attractiveness, based on the literature on facial attraction. Modification criteria are summarised in Table 2 below. A total of 15 agents were designed and pilot tested for attractiveness and realism with a sample of 58 participants. Four stimuli were selected from two models (one White and one Black female) which achieved the highest difference between the most and the least attractive pairs (Table 2).

Table 2. Attractive and Unattractive Agents.

Attractive Agent	**Unattractive Agent**
Nose - Proportional to face.	Nose - Widened by 50%.
Lips - Full.	Lips - Thinned by 25%.
Symmetry - Head and shoulders proportional to each other.	Asymmetry - Head width reduced by 16.5%. Head height reduced by 25%. Shoulder width increased by 30%.
	
Attractiveness: 4.98	**Attractiveness: 2.86**
Realism: 4.78	**Realism: 3.31**
	
Attractiveness: 4.76	**Attractiveness: 2.76**
Realism: 4.28	**Realism: 3.28**

Design. The experiment was based on a 2*2*2 design. Agents' attractiveness (attractive vs. unattractive) and ethnicity (white vs. black) were manipulated between-subjects. Evaluations were collected prior and after interaction with the agents.

Procedure. Participants were introduced to the experiment as a user evaluation of ECA's. Prior interaction, participants were required to evaluate a static image of one of the four targets using the same array of instruments employed in experiment one. Then, they were invited to chat with the agent on any topic they pleased for 10 minutes and left alone in the laboratory. The user wrote their input into a conversation window, whereas the agent spoke its answer back. Finally, participants evaluated the agent image using all the evaluation instruments.

4.2 Results

Mean scores were computed for all 7 dimensions measured in the pre- and post-test (Cronbach alpha > 0.80). Mean scores were entered into 7 mixed-design ANOVAs with Attractiveness (2) and Ethnicity (2) as between-subjects factors and Time (2) as within-subjects factors.

Manipulation Check. The ANOVA on physical attractiveness returned only a strong main effect for attractiveness, $F(1,44) = 46.23$, $p < 0.001$, partial $\eta 2 = .51$, supporting the reliability of the manipulation (mean difference = 1.15).

Test of Hypotheses. The analysis of *social competence* indicated a large main effect of attractiveness ($F_{(1,44)}$ = 48.38, $p < 0.001$, partial $\eta^2 = .52$) and evaluation time ($F_{(1,44)} = 23.79$, $p < 0.001$, partial $\eta^2 = .35$). The interaction attractiveness * time was also significant, $F_{(1,44)} = 6.63$, $p < 0.001$, partial $\eta^2 = .13$. Fig. 4 reports mean and standard errors. (score values on social competence as a function of attractiveness and time). It is evident that participants gave better evaluation to the most attractive agents; and their evaluation improved after the interaction. However, this effect was mostly due to people who interacted with the unattractive agent, as they improved their evaluation significantly more (mean = difference .95) than participants who interacted with the attractive avatar (mean difference = .25).

The analysis on *integrity* returned similar results, although all effect sizes were smaller. The main effect of attractiveness ($F_{(1,44)} = 9.76$, $p < 0.01$, partial $\eta^2 = .18$) and evaluation time ($F_{(1,44)} = 16.25$, $p < 0.001$, partial $\eta^2 = .27$) were significant. Attractive agents were evaluated better than unattractive ones (Fig. 4). The evaluation improved with time especially for unattractive agents, although the interaction does not reach statistical significance ($p = .11$).

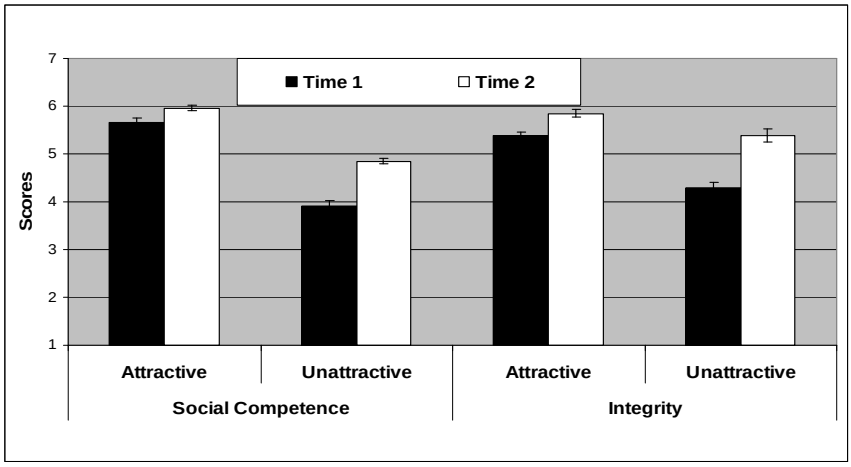


Fig. 4. Social competence and Integrity scores as a function of experimental conditions.

The ANOVAs on *intellectual competence* (IC), *social adjustment* (SA) and *anthropomorphism* (A) returned significant effects for attractiveness (IC; $F_{(1,44)} = 10.14$, $p < 0.01$, partial $\eta^2 = .19$; SA: $F_{(1,44)} = 18.45$, $p < 0.001$, partial $\eta^2 = .29$; A: $F_{(1,44)} = 16.79$, $p < 0.001$, partial $\eta^2 = .28$) and time (IC: $F_{(1,44)} = 12.49$, $p < 0.01$, partial $\eta^2 = .16$; SA: $F_{(1,44)} = 8.64$, $p < 0.05$, partial $\eta^2 = .16$; A: $F_{(1,44)} = 152.69$, $p < 0.001$, partial $\eta^2 = .77$). Attractive agents were evaluated systematically better than unattractive agents in all dimensions (Fig. 5). All evaluations significantly decreased after interaction. This drop was particular drastic in the case of anthropomorphism, as evinced by the higher effect size (A mean difference = .65).

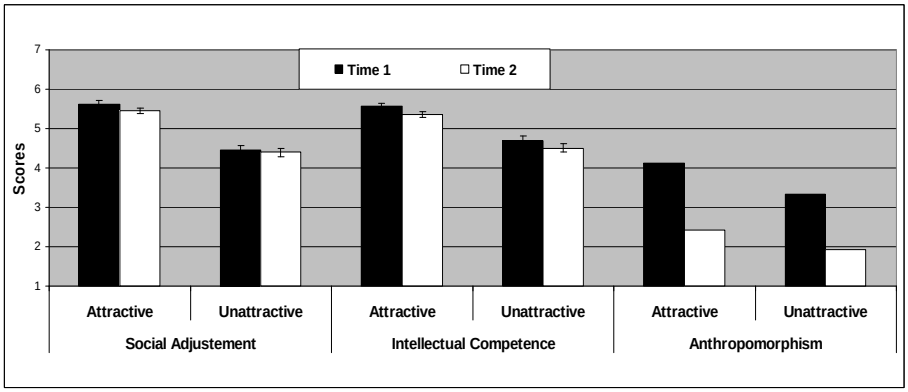


Fig. 5. Social Adjustment, Intellectual Competence and Anthropomorphism scores as a function of experimental conditions.

The evaluation of potency followed a completely different pattern, showing a medium sized effect for ethnicity ($F_{(1,44)}$ = 7.54, p < 0.01, partial η^2 = .14) and a marginally significant interaction ethnicity * attractiveness ($F_{(1,44)}$ = 3.74, p = 0.06, partial η^2 = .08). The interaction effect was due to the unattractive white agent being scored lowest in potency and to the unattractive black agent being scored highest.

4.3 Conclusion

Results of study 2 are summarized in Table 3, which displays effect sizes of significant effects. The effect of attractiveness (A) is robust showing that attractive agents are regarded better in all experimental dimensions, confirming H1. Participants changed their evaluation over time (T), but the direction of this change differed according to the dimension. In particular, social competence and integrity increased, whereas intellectual competence, social adjustment and anthropomorphism decreased. Thus, we reject H2 and H3.

Table 3. Summary of Experiment 2 results.

	A	**T**	**E**	**A*T**	**A*E**
Physical Attractiveness	.51				
Social Competence	.52	T1<T2 .35		.13	
Social Adjustment	.29	T1>T2 .16			
Intellectual Competence	.19	T1>T2 .22			
Potency			.15		
Integrity	.18	T1<T2 .27			
Anthropomorphism	.28	T1>T2 .77			

The only exception to the attractiveness stereotype was the dimension of potency, which also returned the only effect of ethnicity (E) found in the experiment. We argue that this dimension is rather subjective as participants may interpret it as being either a positive or negative trait in relation to an agent. It is worth noting that the agent's ethnicity did not have any major effect on the other dimensions.

5 Discussion

This paper contributes to the emerging literature on social implications of ECA's by showing that virtual bodies afford the attractiveness stereotype. This effect is very strong in first impressions (experiment 1 and experiment 2), but it pertains also to actual interaction (experiment 2). As in real life interaction, the stereotype particularly influences people's opinion of social competence, social adjustment and intellectual competence. The effect on integrity is generally stronger that in real life study, showing that appearance may influence character perception more strongly in ECA's than in real human beings, probably because ECA's are perceived as not to have hidden motives and agenda. Finally, the attributes of potency was found to be unaffected by attractiveness, probably because this is a culturally relevant attribute, or because of the specific target of judgment.
More research is needed to unveil the subtleties of user evaluation of embodied agents, this research provide some preliminary results and a methodology to foster this field.

References

1. Reeves, B., Nass, C.: The media equation: how people treat computers, television, and new media like real people and places. Cambridge University Press (1996)

2. Nass, C., Isbister, K., Lee, E.J., Truth is beauty: researching embodied conversational agents, Embodied conversational agents. ed. J. Cassell, Cambridge, MA: MIT Press (2000)
3. Moon, Y., Nass, C.: How 'Real' Are Computer Personalities? Psychological Responses to Personality Types in Human-Computer Interaction. Communication research. 23(6), 651--674 (1996)
4. Zanbaka, C., Goolkasian, P., Hodges, L.: Can a virtual cat persuade you? The role of gender and realism in speaker persuasiveness, in Proceedings of the SIGCHI conference on Human Factors in computing systems. ACM: Montreal, Quebec, Canada (2006)
5. Baylor, A.L.: The impact of pedagogical agent image on affective outcomes. In Proceedings of Workshop on Affective Interactions, Intelligent User Interface International Conference. San Diego, CA (2005)
6. Khan, R., De Angeli, A.: Mapping the Demographics of Virtual Humans. Volume 2 Proceedings of the 21st British HCI Group Conference HCI 2007. British Computer Society: Lancaster (2007)
7. Eagly, A.H., Ashmore, R. D., Makhijani, M. G., Longo, L.C.: What Is Beautiful Is Good: A Meta-Analytic Review of Research on the Physical Attractiveness Stereotype. Psychological Bulletin. 110(1), 109--128 (1991)
8. Fiske, S.T.: Social cognition and social perception. In M. R. Rosenzweig & L. W. Porter (Eds.), Annual review of psychology. vol. 44, pp. 155--194 (1993)
9. Feingold, A.: Good-looking people are not what we think. Psychological bulletin. 111(2), 304--341 (1992)
10. Langlois, J. H., Kalakanis, L., Rubenstein, A.J., Larson, A., Hallam, M., Smoot, M.: Maxims or Myths of Beauty? A Meta-Analytic and Theoretical Review. Psychological bulletin, 126(3), 390--423 (2000)
11. Rhodes, G.: The evolutionary psychology of facial beauty. Annual Review of Psychology. 57, 199--226 (2006)
12. Dion, K., Berscheid, E., Walster, E.: What is beautiful is good. Journal of Personality and Social Psychology. 24(3), 285--290 (1972)
13. Wheeler, L., Youngmee, K.: What is Beautiful is Culturally Good: The Physical Attractiveness Stereotype has Different Content in Collectivistic Cultures. 795--800 (1997)
14. Holzwarth, M., Janiszewski, C., Neumann, M, M.: The Influence of Avatars on Online Consumer Shopping Behavior. Journal of Marketing. 70(4), 19--36 (2006)
15. Nowak, K.L., Rauh, C.: The Influence of the Avatar on Online Perceptions of Anthropomorphism, Androgyny, Credibility, Homophily, and Attraction. Journal of Computer-Mediated Communication. 11(1), 153--178 (2006)
16. Nowak, K.L., Rauh, C.: Choose your 'buddy icon' carefully: The influence of avatar androgyny, anthropomorphism and credibility in online interactions. Computers in Human Behavior. 24(4), 1473--1493 (2008)
17. Vasalou, A., Joinson, A. N., Banziger, T., Goldie, P., Pitt, J.: Avatars in social media: Balancing accuracy, playfulness and embodied messages. International Journal of Human-Computer Studies. 66, 801--811 (2008)
18. Vasalou, A., Joinson, A. N.: Me, myself and I: The role of interactional context on self-presentation through avatars. Computers in Human Behavior. 25, 510--520 (2009)
19. Messinger, P.R., Ge, X., Stroulia, E., Lyons, K., Smirnov, K., Bone, M.: On the Relationship between My Avatar and Myself. Journal of Virtual Worlds Research. 1(2), 1--17 (2008)
20. McCroskey, J. C., McCain, T. A.: The measurement of interpersonal attraction, in Paper presented at the Annual Convention of the Western Speech Communication Assn. Honolulu (1972)
21. Baylor, A.L., Ryu, J.: The psychometric structure of pedagogical agent persona. Tech., Inst., Cognition and Learning. 2, 291--319 (2005)
22. Brace, N., Kemp, R., Snelgar, R.: SPSS for Psychologists: A guide to data analysis using SPSS using for windows (versions 12 and 13). 3rd ed. Houndmills: Palgrave MacMillan (2006)
23. Pallant, J.: SPSS Survival Manual: A step by step guide to data analysis using SPSS. (2007)
24. De Angeli, A., Brahnam, S.: I hate you! Disinhibition with virtual partners. Interacting with computers. 20(3), 302--310 (2008)