Evaluating the Perception of Virtual Characters by Manipulating Components of Their Physical Appearance

by

Katja Žibrek

A dissertation submitted to the University of Dublin, Trinity College
in fulfillment of the requirements for the degree of
Doctor of Philosophy

Supervisor:
Dr. Rachel McDonnell

University of Dublin, Trinity College

October, 2016

Declaration

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work.

I agree to deposit this thesis in the University's open access institutional repository or allow the library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

Some of the work presented in this thesis includes both unpublished and published work of others.

Katja Žibrek

October 25, 2016

Summary

This work explores the possibilities of studying the perception of 3D generated (virtual) characters using a theoretical background of social psychology. Our main intention is to present studies which explore responses towards virtual characters by manipulating specific components of their appearance. Since social perception relies on visual characteristics of people and can change the way they respond to the observed person, virtual representations of people can be affected as well. The literature on virtual characters presents some examples on how appearance could affect the perceiver's response. The most well-known phenomenon is the so called "Uncanny Valley" which describes the loss in appeal and increase in eeriness when the virtual character reaches a certain level of detail, closely resembling a human being. The research has since shown that the concept of uncanny valley is not as clear as proposed by Mori (1970) and some methodological and conceptual discrepancies have been revealed. In addition, disinterest and lack of appeal can be found in the responses to non-realistic characters as well. While virtual characters are becoming increasingly more prevalent in movies, gaming industry, as well as being used in education, therapy and are expanding to other areas of life, a better understanding of how they are perceived is of crucial importance.

Our work presents experiments which were designed using virtual characters in styles created by manipulating material properties, model geometry and rendering effects. For animated characters, natural human motion recorded with a motion capture system was used. In the first study, we focused on the perception of gender from motion cues, by adding the expression of emotion and analysing the possible effect of stereotypes on how gender is perceived. For the second study, we assessed combinations of different realism levels of shape and materials in an attempt to find the least and most appealing results and find how this manipulation affects perception of emotion. The last two studies were built in interactive and immersive environments where we used proximity in combination with subjective reports to explore the effects of agency and personality on peoples' attitudes towards virtual characters of different visual styles. Our last study is particularly important since it was conducted on an extremely large sample size (1106 participants) and used a between-group design to study the effect of appearance on people's perception.

Our findings show that overall, appearance is an important factor which does not only affect appeal but can change the way perceivers interpret the character's behaviour, empathise with it, and behave around it. We find indications that realism is a positive component of appearance which may even be preferred to stylisation, while certain manipulations of realism

in graphics elements will result in unappealing combinations. We also show social cognition is involved in the perception of virtual characters and the methodology to study interactions between real people can be used for virtual character as well, although some considerations when drawing conclusions should be followed.

We are confident that our work is an important contribution to the theoretical knowledge on perception in computer graphics and has valuable indications for psychology and social sciences, as well as an applicable value in the form of guidelines for character design.

Acknowledgements

First and foremost, I would like to thank my supervisor Dr. Rachel McDonnell for giving me the freedom and motivation to realise my research ideas, but also being someone I could learn from and grow in the process.

I would also like to thank Dr. Carol O'Sullivan for initially accepting my application to work with the GV2 team even though my background was rather different from computer science. Dr. Ludovic Hoyet should be given most credits for having the patience to introduce me to the technical part of my work and for being always ready to help.

Special thanks goes to all my colleagues from the lab, who were invaluable in the research process and with whom I became close friends in the past years. Kerstin Ruhland, Claudia Arellano Vidal and Ziggy for amazing friendship and movie nights; Rowan Hughes, Kenneth Ryall, Fintan McGee, Timothy Costigan and Anton Gerdelan for all the crazy jokes and nights out and about; Andrew Corcoran for taking me mountain biking; Emma Carrigan for being awesome; Rozenn Dahyot for matching leather bracelets, Elena Kokkinara for her cooking, music& soul sharing, and Mukta Prasad for making even the most mundane things in life endlessly entertaining.

An extended thanks goes to other friends I met along the way and who have been there to enrich my stay in Ireland, especially Romina Charifou for being a Jedi, and for making the best nut cake.

Staff at the Science Gallery Dublin, especially Shaun O'Boyle, who made sure our experiment was running properly at the exhibition. Also, Science Gallery Cafe for making the best coffee in Dublin, my fuel during the writing process.

Lastly, I would like to thank my family and friends at home for letting me go to pursue my future abroad.

Contents

Relevant Publications

Journal

1. Zibrek, K., Hoyet, L., Ruhland, K., and McDonnell, R. (2015). Exploring the Effect of Motion Type and Emotions on the Perception of Gender in Virtual Humans. In *Transactions on Applied Perception*, ACM.

Conference Papers

1. Zibrek, K., Hoyet, L., Ruhland, K., and McDonnell, R. (2013). Evaluating the effect of emotion on gender recognition in virtual humans. In *Proceedings of the ACM Symposium on Applied Perception*, ACM, 45–49.

2. Zibrek, K., and McDonnell, R. (2014). Does render style affect perception of personality in virtual humans? In *Proceedings of the ACM Symposium on Applied Perception*, ACM, 111–115.

Colaborations

1. Hoyet, L., Ryall, K., Zibrek, K., Park, H., Lee, J, Hodgins, J., O'Sullivan, C. (2012). Evaluating the distinctiveness and attractiveness of human motions on realistic virtual bodies. In *ACM Transactions on Graphics (TOG) – Proceedings of ACM SIGGRAPH Asia 2013*, ACM, 204:1–204:11.

2. Ruhland, K., Zibrek, K., and McDonnell, R. (2015). Perception of personality through eye gaze of realistic and cartoon models. In *Proceedings of the ACM Symposium on Applied Perception*, ACM.

3. Zell, E. Aliaga, C., Jarabo, A., Zibrek, K., McDonnell, R., Gutierrez, D., and Botsch, M. (2015). To Stylize or not to Stylize? The Effect of Shape and Material Stylization on the Perception of Computer-Generated Faces. In *ACM Transactions on Graphics (TOG) – Proceedings of ACM SIGGRAPH Asia 2013*, ACM, 184:1–184:12.

Posters

1. Zibrek, K., and McDonnell, R. (2015). Evaluating the Uncanny Valley with the Implicit Association Test (IAT). In *Proceedings of the ACM Symposium on Applied Perception*, ACM, 135–135.

Chapter 1

Introduction

Figure 1.1: Examples of visual styles on a virtual character model.

The creation of engaging virtual characters is an important task in the ever growing field of computer graphics. A huge amount of effort has been put into making the characters' appearance, motion and behaviour believable by relying on artistic expertise and, especially in recent years, using advanced scanners and physically-based shaders to replicate an exact appearance of a realistic human. Strikingly little is known however why certain characters still induce a negative response from the viewers and what makes a particular visual style of a character a better design choice than another, an area of research which falls into the domain of perception. Unfortunately, perception of virtual characters struggles with the lack of a concise methodological approach which is further mediated by an unstructured theoretical background. Our research draws on a simple premise – since our perceptual system is very sensitive for the processing of other humans it would be expected that the knowledge on how we perceive people could aid in understanding how we relate to virtual characters as well.

The current thesis explores this possibility by drawing parallels between social psychology and the perception of computer generated characters.

1.1 Motivation

Computer generated virtual characters are an important part of many 3D graphical applications. Commonly, they are used for entertainment purposes, such as substituting or reshaping the appearance of actors in movies, starring in animations or being used for games and simulations. In virtual reality environments, virtual characters can increase the users' immersion in social contexts with which they can even aid in physical rehabilitation or treatments of psychological disorders e.g phobias, depression, eating disorders, to name a few. In some games, users can build their own characters with unique appearance and thus create a virtual representation of themselves. In medical training, students encounter a virtual patient, with whom they need to interact in order to formulate a diagnosis and a treatment plan. The future holds endless possibilities for the use of virtual humans and with an ever expanding territory of computer technology, a demand to better understand how to create a suitable character for a particular domain is imperative.

When addressing the subject of engaging character traits, we can speak about the character's behaviour, voice, movement, appearance and in some cases interaction with the user. Each of these traits also varies greatly, e.g. appearance can vary from photo-realistic level and natural motion to a more cartoon-looking appearance and animation. The important part of designing a character is to know which characteristics will fit best to a particular purpose of use. For example, in which cases would a realistic looking character be a better design choice to a more cartoon-looking one? Some choices are still related to the technical limitations of a particular platform: rendering a sufficient amount of frames for stereo vision in head mounted displays limits the level of visual realism that can be presented in the virtual reality environment, for example. The design choice in these cases would be a trade-off between efficient visual realism on one side and technical performance on the other. However, as the technology is in constant development, we can anticipate the reduction of such compromises in the future. The need to therefore understand the perceptual consequences of a particular character design are of high practical value in the field of computer graphics as well as giving us insights into the nature of human perception in general. The studies presented in this thesis are a step towards understanding the perceptual processes surrounding specific traits of virtual characters. In particular, this work focuses on mismatching appearance elements and changing the render style of the character in order to explore their affect on appeal and perceived realism, but also a more complex effect on social cognition, such as emotions, personality, stereotypes and gender perception.

1.2 Scope and Limitations

The work presented here provides some new understandings of how appearance of virtual characters could influence our response towards them by approaching the research question from the social psychology background. It also utilises additional measures in order to explore different perceptual outcomes. There are many other aspects of human experience we could focus on, however, so our work has some limitations.

- While having a strong focus on perception of humans, we do not explicitly study the effect of the perceiver on the perception on virtual characters. Therefore we do not study how the viewer's social status, personality, emotional state, cultural differences, etc., affects his or her perception. We understand that these factors have an important effect on the perception, but they are out of the scope of our study.
- We only focus on some elements of characters' appearance. For example, we do not particularly study the effect of motion realism but only use natural motion to build the characters for experimental conditions. We are also limited in the variety of character appearance - we rely on the accessible library of models created by artists or produced by scanners. Our findings therefore cannot be faithfully generalised to all virtual characters.
- We also do not focus on other aspects which make the characters engaging - their voice, interaction with the observer, etc. Even though we include some of these aspects, we do not systematically study them.

More detailed limitations of particular experiments are further discussed by chapters.

1.3 Contributions

An important novelty of this thesis is that it approaches the research question from a theoretical background of the field of psychology and social psychology in order to understand the complexity of the perception of virtual characters. We argue that the perception and response to virtual characters is influenced by previous interactions with other humans, therefore the existing social schemas will be applied to virtual characters as well.

In addition, some measures of social psychology which are applied to our work, have not been considered for exploring the perception of virtual characters before. Most of the standard measures require some form of conscious evaluation from the observer. But the inclusion of certain indirect measures for example, would give evidence on how the virtual character is processed on an unconscious level, while behavioural measures will show if the effect is strong (or important) enough to influence the behaviour of the observer. We justify the use of measures commonly associated with studying the perception of real humans for the examination of virtual people by including a theoretical debate on the reasoning behind such a methodological approach.

Our work extends the previous research on appearance of virtual characters by analysing complex interactions between specific visual elements and socially relevant behaviour, such as emotions and personality. It also finds support and relevance for the use of characters of realistic appearance, especially when trying to amplify some characteristics of its personality and increase emotional involvement of the observer. This finding is an important addition since it expands the conclusions of previous research which show that realism is detrimental for the character's appeal.

We also present virtual characters in different platforms and levels of interaction – from video clips to interactive virtual reality environments. While the video stimuli help us to understand the perception of characters typically used in animations and other similar representations, virtual reality allows us to study new and more immersive ways of presenting the characters and collect behavioural responses from the participants.

1.4 Thesis Overview

Chapter 1: Introduction makes an overview of this work and the motivation behind it. Some limitations and contributions are outlined.

Chapter 2: Background provides a brief history of virtual characters and introduces the pipeline of character creation. Studies on perception of virtual characters based on their appearance are then presented. The response is further analysed in the context of neurological and social studies, which build the theoretical framework and methodological approach this work is based on. Lastly, studies of natural human motion applied to virtual characters are presented.

Chapter 3: The Effect of Mismatch on the Perception of Virtual Characters presents two studies where we intentionally mismatched specific elements of the character's appearance to determine how it would affect viewers' perception of the character.

Chapter 4: The Effect of Render Style on Perception of Virtual Characters describes a series of experiments which use one character rendered in different styles to explore viewers' response to it. In addition, all the experiments in this chapter use longer, more immersive, animated sequences of the character in order to increase the validity of our results.

Chapter 5: General Discussion and Future Work presents a discussion of our main results and the direction for the future.

Chapter 2

Background

When we think of computer generated (virtual) characters, we might have a vast variety of images in mind: from cartoon animations we encountered in famous ©Disney animations to elaborate human representations in computer games and movies. This wide variety of examples in character creation and their use is a significant result of developments in the field of computer animation, a digital successor to stop motion techniques typically used in traditional animation with 3D models and frame-by-frame animation of 2D illustrations. Throughout the thesis, the term "virtual characters" will be used to describe characters which fall in the domain of 3D computer animation.

2.1 History of Virtual Characters

The creation of a first 3D computer generated character is credited to William Fetter in the 1959 (Magnenat-Thalmann & Thalmann, 2005; Beane, 2012), who used computers to create 3D models of objects and even of a human body that came to be known as the Boeing Man (Fetter, 1982). This early seven jointed figure was used for ergonomic analysis, where pilot actions could be displayed by articulating the figure's pelvis, neck, shoulders, and elbows. In the 1960s, the field of computer graphics and animation started to expand and in 1970s most of the building blocks of 3D computer animation were laid, such as surface rendering by Henri Gouraud (1971) and texture mapping by Ed Catmull (1974). Fully animated characters appeared in advertisements and music videos in the early 1980s (Figure 2.1). But it was only ten years later that the technology had developed sufficiently for computer generated characters to be integrated into films, such as *Terminator 2: Judgment Day* (1991, Figure 2.1), *The Lawnmower Man* (1992), or a digital double in *Jurassic Park* (1993). This was the start of 3D animation receiving a wide spread commercial success and it wasn't long before full 3D animated movies were released: in 1995, Pixar released the first fully 3D animated feature film *Toy Story* (1995) featuring cartoon-shaped characters to great success critically and

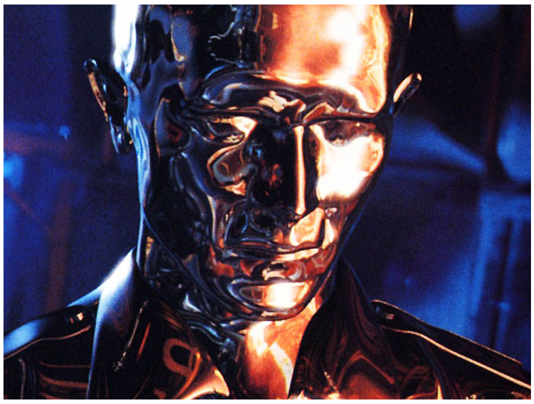
 

Figure 2.1: Left: one of the first examples of a computer generated human in popular media - music video for Mick Jagger's *Hard Woman* (1985). Right: complex liquid-metal material on a virtual character in *Terminator 2: Judgment Day* (1991).

commercially. In the 2000s more technology was being created to support the growing 3D animation industry: *Pixar's Monsters Inc.* (2001) showed that 3D fur could be successfully build onto a character. *The Lord of the Rings: The Fellowship of the Ring* (2001) pushed new techniques with crowd simulations. *Final Fantasy: The Spirits Within* (2001) attempted to create photo-realistic humans for the full 3D animated film. This film, however, did not do well critically but did influence the computer graphics by directing the research on the problems of creating believable human characters. Several attempts were made after to enhance the realism of these characters, but most of the time audience did not fully embrace the results (Figure 2.2).

In the 1990s, there was also a shift towards more interactive media, where real-time animation was employed. This posed new challenges to the field of character creation: the characters should not just look, talk, and behave like us in the virtual world, but also be able to see our avatars and to react to what we say and to the emotions we convey. In 1994 the Sony PlayStation home console system was released and was one of the first home console systems to be able to handle 3D graphics with hardware accelerations. NVIDIA later on released the first consumer-level graphics processing unit, the GeForce 256. These accelerators were needed because 3D gaming engines needed power to play the games at full quality. While these technological advancements made real-time interaction possible, this was usually at the cost of making the game characters less visually complex, e.g. using low poly meshes to represent a human figure. Throughout the years, the graphics and game components have developed to such an extent that they are able to support more realistic texture and geometrical structure of the characters. However, many games would involve cut scenes of cinematic sequences which could achieve higher photo-realism while disabling the interactive element of the game.

Figure 2.2: Examples of realistic virtual characters. Upper left - *Final Fantasy: The Spirits Within* (2001), upper right - *Animatrix* (2003), lower left - *Polar Express* (2004) and lower right - *Beowolf* (2007). While reaching high levels of realism, some characters may be getting “too close” to portraying a real human being but not quite reaching it. The end result can be unsettling.

Characters are used in areas other than films and games, e.g. education and training, simulations or therapy. The applications of virtual characters today are numerous and new areas of implementations are in development. Here are some examples of such applications (taken from Magnenat-Thalmann et al. (2005)):

- simulation-based learning and training (transportation, civil engineering, etc.), skill development, team coordination, and decision-making.
- ergonomic analysis in work environments and vehicles.
- virtual patients for surgery and plastic surgery.
- virtual teachers for distance learning, interactive assistance, and personalized instruction.
- virtual people for the treatment of social phobia and virtual psychotherapies.
- virtual inhabitants for virtual cities and architectural simulation with buildings, landscapes and lights, etc.

- virtual representations of participants in virtual conferences in order to reduce the transmission bandwidth requirements.

For example, an interesting implementation of virtual characters is in the simulations for treatment of social phobia (Herbelin et al., 2002). Social Anxiety Disorder (SAD) is one of the most common psychiatric disorders and is characterised as an intense fear of social performance in public situation, where potentially an interaction with other people will occur. A study using virtual reality (Pertaub et al., 2001) found that not only was social anxiety induced by the virtual audience, but the degree of anxiety experienced was directly related to the type of virtual audience feedback the speaker received (see Figure 2.3). For the treatment (Lee et al., 2002), virtual characters in virtual reality were systematically manipulated by the therapist to express different levels of yawning, chatting or clapping, indicating the amount of agreement or disagreement with the participant. A gradual increase of negative social evaluation would motivate the patient to adjust the anxiety level to achieve a more functional response.

Figure 2.3: Examples of the virtual characters in the study on fear of public speaking (Pertaub et al., 2001). Characters are responding to the participants speech by reacting interested, bored, etc. This behaviour was found to have a significant effect on how anxious the participant felt during the experience.

By now, people have become accustomed to seeing high-quality 3D animation and visual effects. Advances in technology and artificial intelligence are making game characters act increasingly lifelike. Other areas besides the entertainment industry, like education and training, are also increasing their interest in virtual characters to optimise the learning procedure. The future of virtual character development is exciting and unpredictable but not without considerable challenges. In the next section, the techniques to create virtual characters are described. Later, some of the limitations in character creation will be presented, which will serve as an introduction to the body of research in this area.

2.2 Character Production Pipeline

Production of virtual characters follows a particular pipeline. Here, two main stages which are important in character development, are briefly presented - modeling and animation. Other elements, e.g. post processing effects that contribute to the overall appearance of the character are briefly described as well.

2.2.1 Modeling and Texture

A model is a geometric surface representation of an object that can be rotated and viewed in a 3D animation software package. There are many ways to create a virtual model. It can be built with 3D animation software such as ©Autodesk Maya or 3ds Max[1] or even sculpted manually with packages like ©Autodesk Mudbox and ©Pixologic's ZBrush[2]. Artists usually use drawings or photographs of real people to retrieve the topology of the human shape (Figure 2.4).

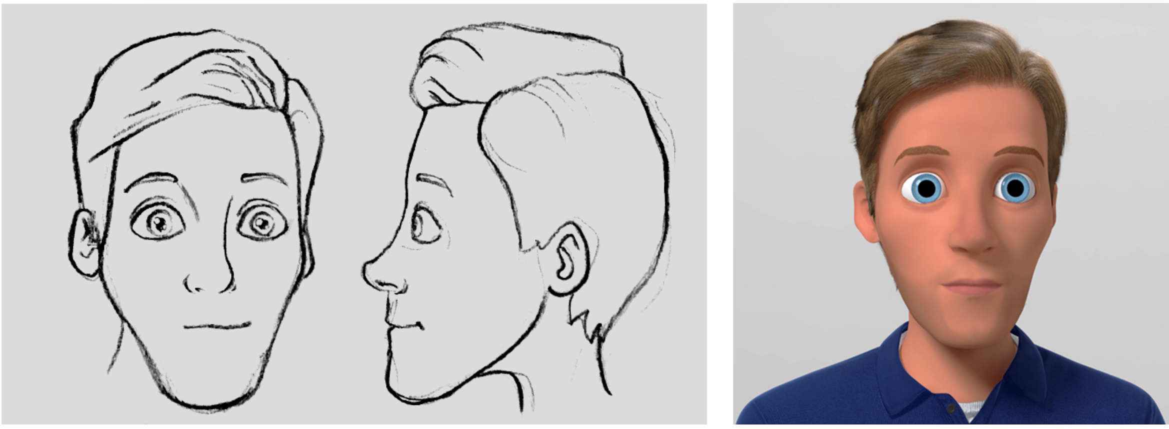


Figure 2.4: Artist created character models using multiple drawings that capture human topology of the face.

The models can also be generated procedurally with algorithms which create a 3D mesh, or retrieved through photographs and laser scanner technology. The scanning domain grew from original work on stereo capture of a face augmented with skin markings (Parke, 1974). Recent work includes a system that combines a recovered depth map and surface normals to generate a model (Nehab et al., 2005). A hybrid system of active light and augmented skin markings is the basis for the current state-of-the-art example of creating a photo-real human

[1]http://www.autodesk.com/. Retrieved 7 October 2016

[2]https://pixologic.com/. Retrieved 7 October 2016

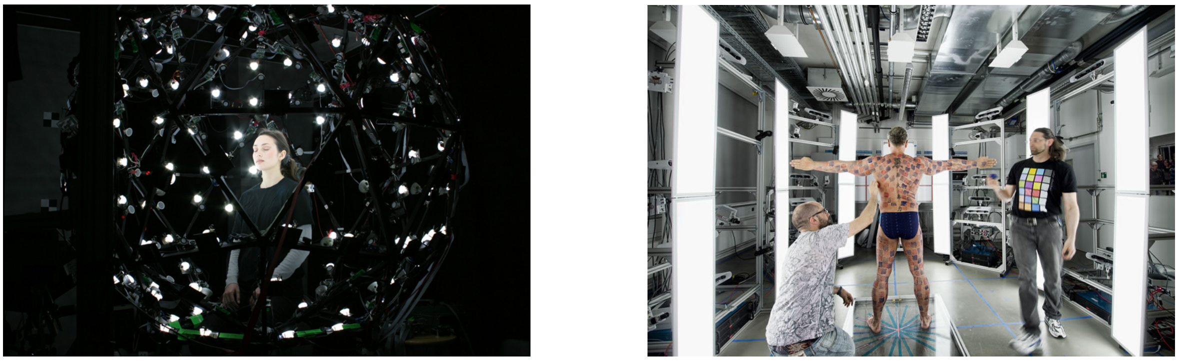


Figure 2.5: *Left:* Actor in a light stage. This technology was used to create accurate facial mesh for the virtual character in the *Digital Emily Project* (Alexander et al., 2009). *Right:* 4D scanner at Max Planck institute which can reproduce a 3D mesh with approximately 150,000 points.

face (Alexander et al., 2009). Simpler variations of the system also exist, e.g. single-shot capture of facial geometry (Beeler et al., 2010). The described systems are used to generate facial geometry but scanners that capture the whole body exist as well. The "4D scanner" used in Max Planck Institute for Intelligent Systems[3] uses custom speckle projectors for an accurate stereo reconstruction of a 3D shape (Fleming et al., 2016) (See Figure 2.5, *right*).

An important part of creating a virtual character is also the generation and correct application of textures on the character's mesh, a process known as texture mapping. Texture mapping is defined as mapping of an image onto a 3D surface. The idea is to map the colour of the image onto the corresponding colour of the object at each pixel (Catmull, 1974). In the texturing phase, texture artists can apply colour and surface properties to the geometric models. This task can vary from hand-painting the textures to using photographs to piece together the texture. Simple virtual characters usually have single texture maps for each surface, which add detail such as wrinkles. Texture artists can also paint directly on the 3D object in some in software packages.

For a more complex representation of the character's appearance, surface textures, such as bump maps, can be used as an easy and cheap way to generate shadows and raised surfaces, such as pores and wrinkles. A bump map is a grayscale image where darker areas represent depressions and light areas raised portions of the mesh. The perturbed normal is then used instead of the original normal. A more complex way to achieve higher detail is by normal mapping, where instead of grayscale image an RGB vector is used, giving the model an even smoother surface. For certain areas of the character, such as sweat on the skin or the eye liquid, specular maps create the illusion of shiny surfaces. Simple game and cartoon characters would usually have a combination of these maps, while realistic characters use a more complex structure, with elaborate shaders such as subsurface scattering for skin (Jensen

[3]https://ps.is.tuebingen.mpg.de/pages/4d-capture. Retrieved 7 October 2016

et al., 2001; Jimenez, 2012) (see Figure 2.6) and iris self-shadowing for the eye (Sousa et al., 2012). A lot of research has also been devoted to creating realistic hair due to a demanding task of rendering individual strands of hair (Marschner et al., 2003; Ward et al., 2007; Martin et al., 2014).

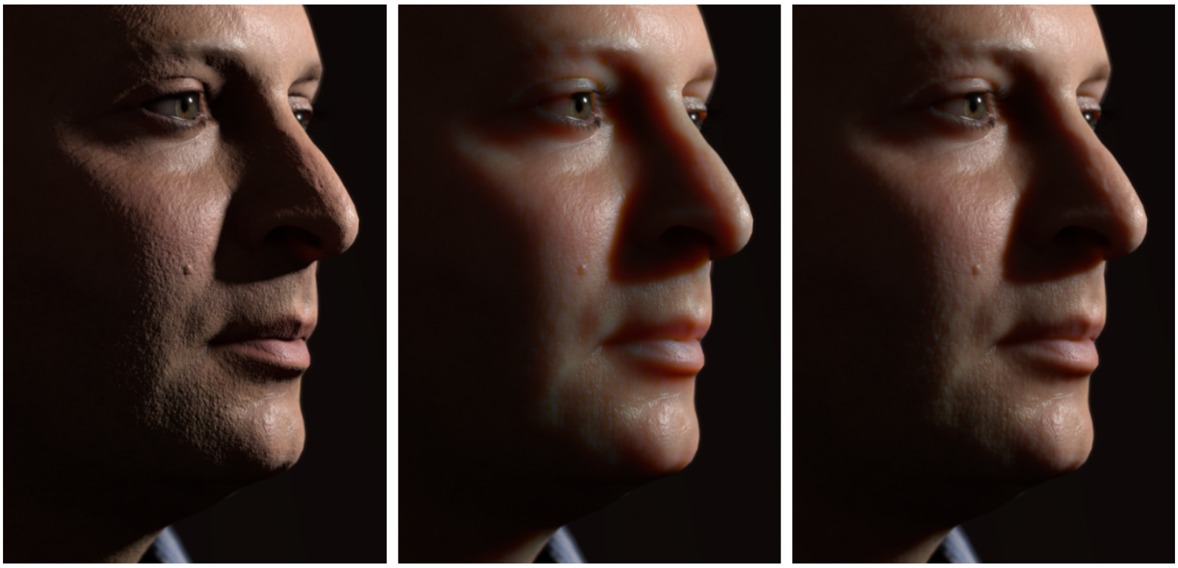


Figure 2.6: Examples of different subsurface scattering levels (*from left to right, respectively:* no, high, and medium level subsurface scattering.). This effect is used to give the appearance of light penetrating the layers of the skin epidermis, giving the character a more natural appearance (Jimenez, 2012).

2.2.2 Animation

In order to move the geometry of the character's surface, a system of control is applied, and the process is called rigging. A rig includes a skeleton with bones organized in a particular hierarchy whose motion deforms a certain part of the surface geometry (skinning). Each bone is associated with a number of vertices on the mesh so when the bone moves, the corresponding polygons of the mesh will move with it. Parts of the mesh can be associated with two or more bones and the motion of the polygons is then calculated on the basis of the weight each bone has on the vertex. Rigging is a highly demanding technical as well as artistic process, while some automatisation tools exist (see for example (Baran & Popović, 2007; Li et al., 2010)). In realistic models, special attention is given to the skin deformations for realistic appearance, e.g. wrinkling of the skin (Jimenez et al., 2011). A bone based model can be applied to move the muscles of the face as well, a more popular approach is using blend shapes. Blend shape animation is the method of choice for keyframe facial animation: a set of blend shapes (key facial expressions) are used to define a linear space of facial expressions (Joshi et al., 2005).

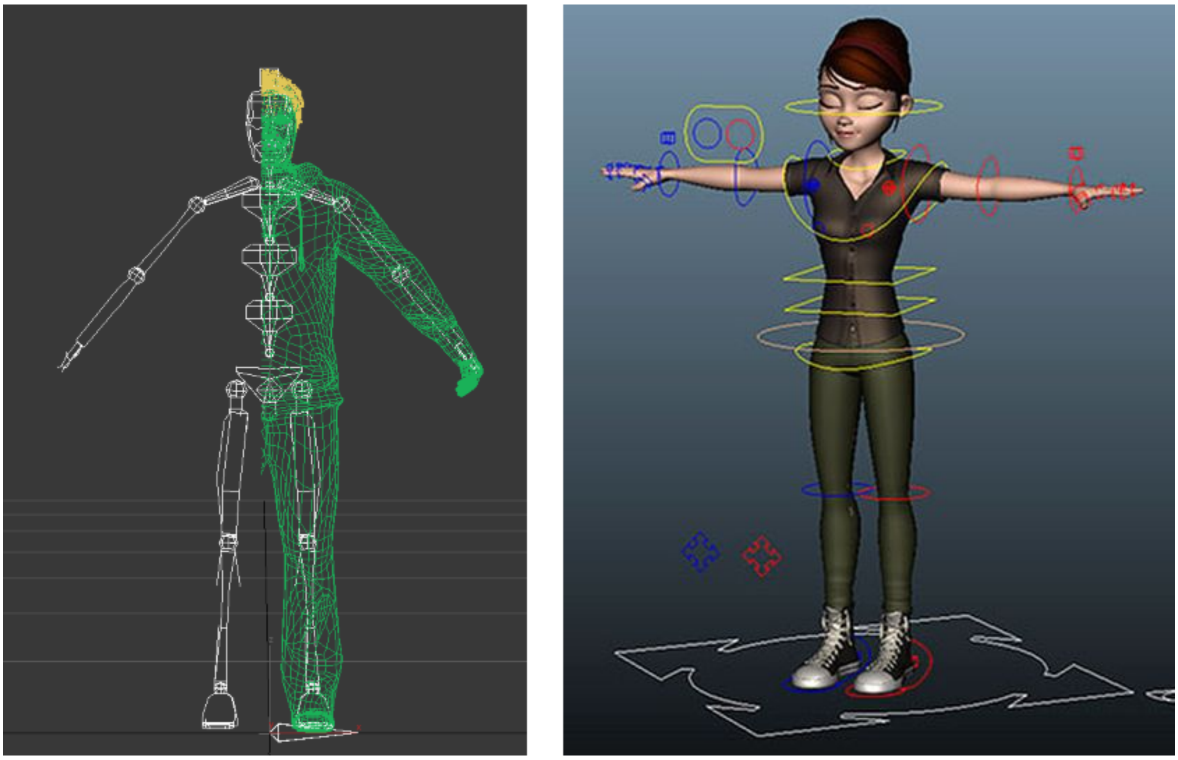


Figure 2.7: *Left:* 24-bone skeleton with an example mesh in wireframe mode. This skeleton was used for most characters in our research. *Right:* cartoon character's rig presented in T-pose. The colour lines around the character's mesh present the controllers for the range of motion of the rig.

The last step of the animation process is applying motion to the character. Here as before, the artist can fully animate the character himself using key poses and connecting the remaining frames in the animation with linear interpolation (such as B–spline) to create the sense of fluidity. This control over the character's motion gives the freedom to exaggerate and deform gestures and expressions of the character. Animators can also use some automatisation to speed up the process of animation, e.g. using kinematics. Forward kinematics calculates the end of the linked structure (end–effector) of the skeleton based on angles of all the joints. Inverse kinematics is similar but more complex: the position of the the end–effector calculates the joint position in reverse fashion, resulting in multiple possible positions which need to be controlled to obtain one stable position.

Fully automatic motion can be created using physics-based algorithms, which guide joint rotations. Physics elements such as mass, velocity and collisions are taken into account when calculating the end positions of the skeleton (see for example (Hodgins, 1996; Yang et al., 2004)).

For natural human motion, the most typical way is to record real actors with a motion capture system. Optical motion capture is a technique which uses a system of infrared cameras, surrounding the actor dressed in a tight suit with reflective markers placed on the major joints of his body. As the light coming from the cameras is reflected from the markers, the cameras can use this input to find the markers in real space – the data which is translated and reconstructed in the virtual space (See Figure 2.8). This marker data is then used to control bone positions and rotations of the character's skeleton in virtual space. If we want to map that motion data onto a virtual character, additional manipulation of data needs to be performed. Since the virtual character used to display motion will only rarely have a matching morphology to the one of the actor, retargeting of the motion needs to be performed. This can be done by importing the motion data into another software, for example ©Autodesk 3ds Max, and retargeting it to the bipad (skeleton) hierarchy provided.

Figure 2.8: Motion capture setup at Graphics, Vison and Visualisation group at Trinity College Dublin. The actor is walking in the motion capture area, wearing a tight-fitting suit with reflective markers. The motion is recorded with Vicon optical mocap system of cameras, arranged circularly around the capture area.

New technology is developing in the direction of video facial tracking, where an actor is wearing a helmet with a camera attached and positioned in front of the face. The video of facial motion can than be analyzed and retargeted to a virtual character by specific computer vision softwares, e.g. in Faceware Technologies[4]. Recently, ©Cubic motion in combination with ©Unreal Engine and ©Ninja theory released a demo showing cutting edge real-time

[4]http://facewaretech.com/. Retrieved 7 October 2016

facial animation[5], where actor's motion was accurately transferred onto a modified virtual model of the actor (Figure 2.9).

Figure 2.9: Actor's performance being captured and retargeted in real-time onto a virtual character for the *Hellblade: Senua's Sacrifice* demo at Game Developers Conference 2016.

Animated characters can then be put in a virtual environment, where additional post-processing effects can define the overall appearance (render) of the character. A rendered character can look very different to its original model. Lighting of the scene and the character, for example, is of great importance in the end result, since it can put emphasis on certain areas of the mesh and even change the perceived intesity of emotion (Wisessing et al., 2016). Additional shaders can change a realistic representation of a character into a hand-drawing style character, but keeping the original realistic geometry of the mesh (Figure 2.10).

2.2.3 Conclusion

The creation of virtual characters has reached the point where we can create engaging virtual characters for a wide variety of platforms and uses. However, while it is important that an artist can create a very appealing and expressive cartoon character, equal importance is given to the accuracy in portraying believable appearance and motion of realistic characters. The technology for creating realistic characters is constantly advancing but it faces considerable challenges. For example, it is arguable whether any fully computer generated realistic virtual character created to this day is convincing enough to pass as real (Tinwell et al., 2011). Also, very little is known what makes unrealistic (cartoon) characters appealing and big animation

[5] http://www.cubicmotion.com/cubic-goes-real-time-at-gdc/. Retrieved 30 September 2016

Figure 2.10: Similar meshes of a character in different visual styles. The character on the far right is rendered differently from the cartoon version shown next to him by using UE4's inbuilt cel-shader effect.

companies such as ©Disney and ©Dreamworks rely on artistic experience and instinct when creating human characters. These questions, if answered, could provide useful guidelines for the creation of characters for a specific purpose (e.g. training, therapy) and could potentially avoid making bad design choices which would disengage the audience. In the next section, the process of perception is described and its implications on the responses to virtual characters are outlined.

2.3 Perception of Virtual Characters

Before we explore the perception of virtual characters, we first focus on the general use of the term *perception*. Perception comes from the Latin word *perceptio* which literally means "to sieze" or "to understand" (Schacter et al., 2011). Perception can describe at least two processes - processing of information from the senses (visual, tactile, audio, etc.) by the brain and interpreting this information to represent and understand the environment (Bernstein, 2010). The first process transforms low–level information such as shape of an object, into high–level organisation such as recognizing the object as a box. The second process is based on already acquired information about the world and therefore works by activating existing mental schemas about the world and provides a context for the information from the senses to be interpreted. The first process is also known as bottom–up processing and the second as top–down processing. A simple model of this process is described in Figure 2.11. In the context of virtual characters, we will present an extended model of perception which includes studies of neurology (bottom–up approach) and social cognition (top–down approach).

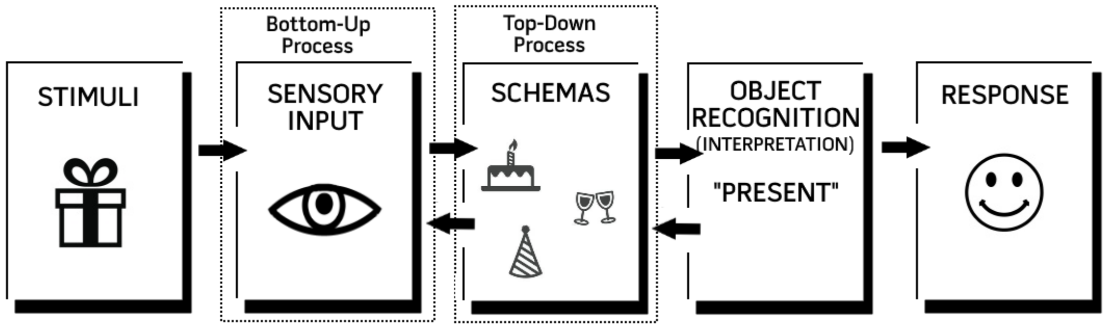


Figure 2.11: Simple model of perception. The object is first seen by the sensory system (visual input) and then interpreted in the context of the created schema (being at a birthday party). The object is recognized (present) and ends in the response of happiness. The arrows represent the direction of effect, where there is a feedback loop between sensory information and created schemas. Schemas also include cognitive states of the person (his attention, motivation, personality) which can affect the sensory organ to perceive selectively. For example, it the person does not trust his friends that gave him a present, he might react concerned or even frightened at the sight of the box. The first process is also described as bottom-up, because it starts as a sensory signal, while top-down is an established set of associations in the brain. Another example is in this very picture. The boxes that describe separate steps in the model are registered by the sensory system as broken lines (bottom-up process) yet our brain connects them in a rectangular object, seemingly casting a shadow (top-down process).

2.3.1 Real and Non–real Characters

Evidence presented by Neuroscience showed that there is a fundamental difference in how our brain processes information from real people as opposed to non-realistic (cartoon) characters. By investigating the active areas in the brain while the participant was viewing real life videos in comparison to rotoscoped versions of the same footage, which turned the realistic appearance into a cartoon version (Mar et al., 2007), some differences have been found. For live-action footage, medial parietal areas were activated (precuneus, posterior cingulate) which are associated with social cognition and mentalising, taking another one's perspective especially with respect to their thoughts rather than body sensations (hunger). It is also involved with the perception of self and the processing of emotions. The right middle frontal gyrus was activated for live footage also, which is responsible for perceptions of intentionality, making judgments about other people. For the cartoon footage, the brain puts more effort into visual processing (cartoons had more colours, brighter) and interestingly, observing cartoons was more associated with the reward system in the brain (Kringelbach, 2005), making it seem that cartoon stimuli are seen as more (visually) rewarding than real actors. Similarly, other research (Han et al., 2005) provided evidence that behaviour of cartoon characters does not have a social impact on adult viewers as real-life actors do (engagement of medial prefrontal cortex and the cerebellum loop). These areas are possibly engaged in mentalising about other humans or forming a theory of mind of other humans.

Studies on autism also produced more indications which point to impaired social perception due to inadequate processing of real human faces and, in some cases, an intact processing for cartoon characters. Individuals with autism have significant deficits in face perception. The kinds of face recognition errors made by persons with autism suggest that they do not employ a normal configural or holistic face processing strategy (Hobson et al., 1988). However, activation in a particular child with autism showed, that he developed normal activation in the fusiform face area and the amygdala for cartoon characters in Digimon (Grelotti et al., 2005) but not real faces, which is the case for people without autism. A recent study by Carter et al. (Carter et al., 2016) showed that exaggerated motion (commonly used for cartoon characters) increased social behaviour from children with Autism Spectrum Disorder, showing the potential importance of motion stylisation. The activation of medial brain areas was found to be important for development of the social brain from infancy onward (Johnson & Tassinary, 2005). The subcortical route not only detects the presence of faces, and orients the newborn towards them, but might also activate relevant cortical regions such as the lateral occipital, fusiform and orbitofrontal cortices. People with autistic spectrum disorder show abnormalities in the amygdala thus indicating deficits in aspects of social cognition.

The mentioned research showed that there is a difference in how we perceive real as opposed to cartoon characters. While real people induce social responses, cartoon characters seem to be more visually rewarding. To explore the difference further,in this thesis we first focus on stylisation as a factor that influences the perception of non-realistic virtual characters. We use the term *stylisation* as opposed to *realism* to emphasise visual qualities (such as shape, colour, texture, etc.) which were intentionally modified and exaggerated away from realism. When the goal is to achieve visual qualities that reflect human appearance as much as possible, we label these characters as *realistic*.

2.3.2 Stylised Characters

Virtual characters can be created as completely abstract representations, with exaggerated facial and body features, or created in a style we usually find in cartoons and computer games. Studies on aestheticism (overview in (Cupchik, 1995)) explored the effect of style on the perceivers. The stylistic component evokes an emotional and generally more diverse response to the object than a conventionally exhibited object e.g. seeing an object in real life as opposed to seeing it illustrated. Style is thus used to induce a more complex perception of the object, which evokes feelings and allows the viewer to make his own personal interpretations of the visual content and construct his own meaning around it. However, this study also found that perceivers themselves affect the overall visual experience. For example, a naive user will prefer art which is closer to real life whereas an experienced viewer will look for novelty in the artwork.

In graphics, stylisation is applied by the technique of nonphoto-realistic rendering (NPR). NPR has attracted much interest in the computer graphics community over the last decade and has established itself firmly alongside the quest for increasing realism. Techniques that allow automatic creation of images that convey complex meaning and support high degrees of abstraction have applications ranging from illustration to information visualization to artistic expression.

Stylisation techniques have beneficial effects on perception as well. In augmented reality, different types of stylisation on the perception of real and created objects were explored (Fischer et al., 2005; Haller et al., 2005) and the results showed that stylisation significantly decreased the discernibility between real and virtual objects. This finding is particularly useful for increasing immersion in augmented reality when the observed object does not possess the same level of realism.

There have been studies done on how stylization could aid the perception of computer generated faces. It has been shown that stylization can considerably improve memorisation of novel faces, without compromising identification of the person if shown on a photograph (Gooch et al., 2004). Winnemoller et al. (2006) conducted two user studies that aimed at evaluating the proposed stylization technique using naming and memory game tasks. In the first task, which followed the same protocol as Gooch et al. (2004), participants were significantly faster at naming stylised images than nonstylised images. In the second task, participants played a memory game using either stylised or nonstylised faces. Again, participants were quicker to finish the memory game with stylised faces. Taken together, these studies suggest that stylization provides crucial benefits for static stimuli in tasks with a heavy memory load. Wallraven et al. (2007) concentrated on the impact of stylization on dynamic stimuli and found stylisation improved recognition of the 3D generated avatar. The same study, however, also found a decrease in intensity in recognition of expression when the face was highly stylised (brush strokes), possibly indicating the adverse effect of dynamic noise produced by brush strokes.

Interestingly, stylisation can also be used to achieve an empathic response from the observer. In the discussion proposed by Misselhorn (2009), the author explains how it is possible that we can empathise with objects. Empathy is defined as "taking ones perspective" when the person is expressing an emotional response. It is also possible to emphathise with objects that we know are not real and alive through the conceptual reality that the object possesses and the viewers identification with that concept. The object can exhibit this conceptual meaning through modifying its appearance away from visual accuracy. This is done through visual simplicity - the reduction of realism increases its conceptual performance. On the other hand, the concept of realism is harder to achieve as it requires more complex visual information.

Another explanation for this emotional response is anthropomorphism. This term describes the tendency to imbue the real or imagined behaviour of nonhuman agents with humanlike characteristics, motivations, intentions, or emotions (Epley et al., 2007). According to this theory, people have a natural ability to anthropomorphise - since very early in the development children attribute human qualities to nearly any object (clouds, stuffed animals, dolls) because the knowledge of their own mental world (of self) is more readily accessible than the actual knowledge of the objects. By growing up, we build more realistic representations of the objects and are more likely to anthropomorphise when the agent exhibits human-like traits (motion that resembles human motion, appearance, voice, etc.) and when we are motivated to do so: when faced with unpredictability and when there is a strong desire to socialise. In the movie *Cast Away* (Cast Away., 2000), the protagonist was stranded alone in an island with no human or animal company and therefore developed a social and emotional connection with a volleyball he affectionately called Wilson. While this is an extreme case, it is a good example of humans' need to anthropomorphise when they lack socialisation.

In human-computer interaction, people are more willing to anthropomorphise if the agents exhibit more human-like qualities: robots and virtual characters with human qualities can therefore elicit more emotional response from the users, especially if they are motivated to do so. However, if the characters or robots become too close to resembling a real human, they might induce a negative response from the perceivers. This effect is also known as the "Uncanny Valley" which is explained in the next section.

2.3.3 Realistic Characters and Uncanny Valley

The first attempts to describe people's reactions towards more realistic synthetic human representations were made while examining humanoid robots (e.g. (Kobayashi et al., 2003; Minato et al., 2004)). Mori (1970) warned that robots should not be made too similar to real humans because this would evoke unpleasant feelings in the viewer. To characterize this relationship, Mori used a hypothetical curve and coined the sudden dip in this curve at almost human-like levels as the "Uncanny Valley" (Figure 2.12). The same function was later proposed for the physical appearances of computer generated characters (see for example: (Seyama & Nagayama, 2007; MacDorman et al., 2009; Tinwell et al., 2013; Yamada et al., 2013)). The uncanny valley has increasingly been experienced by the public when characters in movies or video games appeared to be "not quite right" (See Figure 2.2).

In graphics, the attempts to measure the likability (some studies used the term appeal, affinity or positive response) of virtual characters approaching realism were performed in various ways, according to the understanding of the uncanny valley phenomenon. Some studies focused on the literal understanding of uncanny valley, where almost humanlike characters will always receive more negative responses. Schneider et al. (2007) studied the subjective responses of people to characters from various forms of media, which were ranging from abstract to almost human-like characters. They found that the combination of traits which is most accepted by people is clearly non human appearance of the character with the ability to emote like a human. Flach et al. (2012) expanded this to characters which were familiar (appeared in popular animations or games) as opposed to unfamiliar characters and found that the uncanny valley is more moderate when characters are previously known by the subject.

Other authors focused on changing different characteristics of visual appearance of the character which may enhance or reduce the uncanny valley. MacDorman et al. (2009) showed participants images of human-like virtual faces with different textures applied (ranging from photo-realistic to line texture models) and different polygon count of the mesh. Results suggested that decreasing texture photo-realism can actually make the face look less eerie and more attractive, however, contrary to the uncanny valley graph, the eeriest character was actually a stylised virtual face. Similarly, Hyde et al. (2013) showed that realistic characters were found to be more pleasant than the cartoon characters, however their realistic model did not represent a high level of photo-realism. These studies therefore highlighted an important issue with the methodological approach in studying uncanny valley - the choice of models for both realistic and unrealistic spectrum of characters will have an affect on the subjective response. Another study by McDonnell and colleagues (2012) tackled this problem by using a larger range of styles from abstract to highly realistic renderings applied to one 3D virtual mesh, scanned from an actor (see Figure 2.13). For the less realistic range, they specifically used styles which were found to be appealing by the audience before (e.g. Disney appearance for the Toon CG character) and which were created by an artist. The realistic range included different levels of detail of the scanned texture of the real actor. They found that rendering style affects the appeal and trustworthiness of the characters but realism does not necessarily reduce appeal, since unappealing evaluations could be found for both realistic and abstract characters.

Other research was conducted to investigate the effect of render style as well. A recent study (Volante et al., 2016) using rigorous methodological approach found a relationship between the level of realism of the virtual patient and medical students' empathy and engagement with them. A virtual patient rendered in a realistic style was more able to induce feelings of empathy and anxiety as medical students were observing the decline in his health.

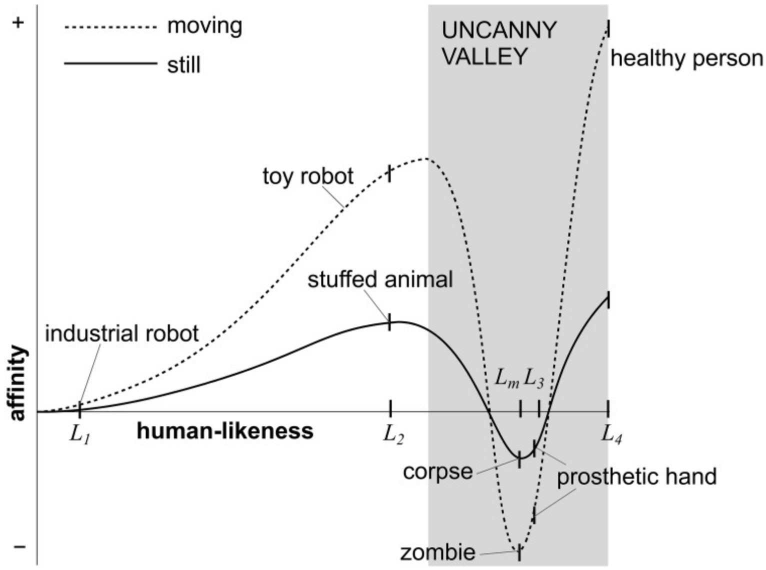


Figure 2.12: Hypothetical graph of the uncanny valley by Mori (1970) for still and moving representations of robots. When appearance of the robots is approaching realism, we see a drop in the perceived familiarity. This effect is even more apparent when the robots are moving (dashed line). Figure is adapted by Katsyri et. all (2015).

Since engagement with the virtual patient is important for memory retention it was deemed helpful for students to train on patients that were rendered in a realistic style. Similarly, Ring et al. (2014) found that realistic characters are more appropriate for medical tasks. Another recent study (Patel, 2015) however, found little effect of realism on deciding about a character's fate in an ethical dilemma; their results put more importance on the context (narrative) around the character. In conclusion, these studies have not found proof of realism being the decisive factor of appeal or the lack of appeal.

A different approach focused on morphing images of characters in order to produce linear transitions from one end of character realism to another (Seyama & Nagayama, 2007; Yamada et al., 2013; Cheetham & Jancke, 2013). Seyama et al. (2007) showed that the uncanny valley was confirmed only when morphed faces had abnormal features such as bizarre eyes. Therefore there are features of the human face which are more sensitive to manipulation of realism. They explain this evidence by recognizing that our perceptual system is very sensitive in processing facial information. This idea was further developed in the categorisation ambiguity hypothesis (Yamada et al., 2013; Cheetham & Jancke, 2013), where it was not only shown that the lowest likeability is at the morph object which is most difficult to categorize, but that this response is more prominent when the morph is between a real human and an inanimate object or representation of a human. Studies focusing on neurocognitive mechanisms attribute negative evaluation for a specific render style to the difficulty of categorizing images in a particular category, resulting in a competing visual-category representations dur-

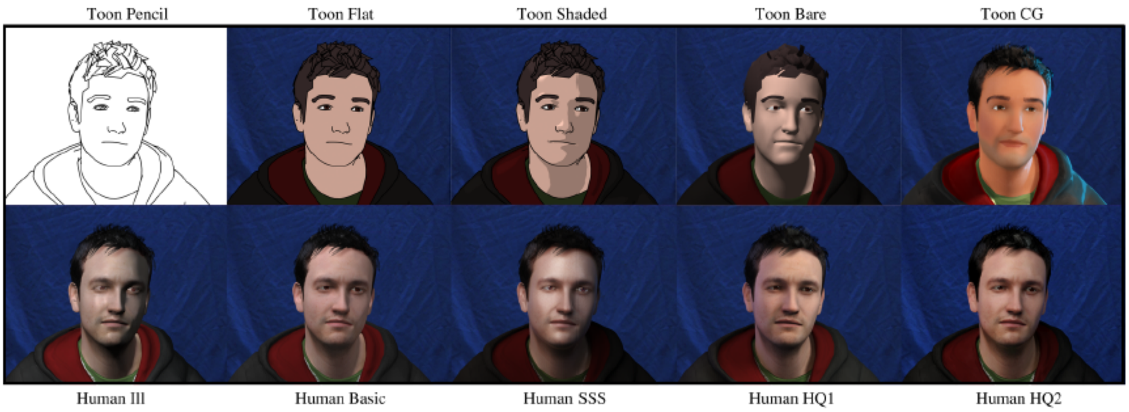


Figure 2.13: The ten render styles, ranging from abstract to realistic, used in the experiment of MacDonnell et al. (2012).

ing recognition (Ferrey et al., 2015). Negative effect for such images occurs to the extent that selecting one interpretation over the other requires inhibition of the visual-category information associated with the non-selected interpretation. This brain process, where conflicting information is more difficult to process as opposed to familiar or typical information, is also known as predictive coding (Saygin et al., 2012; Chaminade et al., 2007). Cheetham et al. (2014) found however, that perceptual discrimination is not necessarily associated with negative affect and can sometimes also serve as an amplifier of underlying positive or negative affect.

Some authors focused more on the ability of motion to enhance negative or positive response to the characters. Studies of Thompson et al. (2011) and Piwek et al. (2014) found that more realistic movement elicits higher affinity and gave no evidence of uncanny valley. Motion anomalies however (Hodgins et al., 2010; Tinwell et al., 2013) have an important effect on the affinity towards virtual characters, especially if they are in the upper part of the face. There are also important interactions between the realism of the appearance and motion of virtual characters. Chaminade et al. (2007) conducted an experiment, where participants had to decide whether they see biological or synthetic motion on characters that ranged from low to high levels of antropomorphism. The results showed that people were more prone to see motion as biological, the less the character's appearance was antropomorhic, possibly because more detailed representations of the character trigger processes in the brain which are less automatic and require higher cognitive involvement, thus reducing perceptual bias.

Some studies believe that the negative effect could be avoided by increasing attractiveness of the virtual model. Many studies on the perceived attractiveness of human faces used the technique of merging different photographs to achieve an average appearance. There was speculation that this technique impacts ratings of attractiveness not just because it averages the shape, but also because it removes blemishes and other skin irregularities (Alley & Cunningham, 1991). Several studies confirmed that texture changes do result in a significantly more attractive face (Benson & Perrett, 1992; Little & Hancock, 2002). Publications in the cosmetics domain also help explain the observed effects on appeal: Fink and colleagues (2006) created textures from photographs of women of different age, and evaluated these textures on a single female virtual character. Renderings with pure skin have been rated as younger and more attractive than renderings with strong variations in skin pigmentation. This observation was confirmed in a follow-up study (Fink & Matts, 2008), which showed that blurring the skin texture can increase attractiveness.

The uncanny valley has been an important research question in many studies that explored likability of virtual characters. Regardless of the importance and strong research focus, a consensus on whether the uncanny valley exists as described by Mori (1970) has not been reached (Kätsyri et al., 2015). The reasons may lie in the lack of standardized measures that would be applicable to the wide range of characters that researchers are using, different understandings of the concept of uncanny valley, inconsistencies in methodology, etc. It could simply be that the human perceptual system is too sensitive and can spot small details of imperfection in the characters' motion and appearance. This presents a very significant problem and many animation and game studios have reverted back to lower levels of realism in fear of not meeting the requirements for the character to be believable.

There is also the possibility that the uncanny valley is not just a side effect of technological imperfections but might be exposing an underlying psychological organisation. There is an indication that the uncanny valley is impossible to traverse due to a growing habituation of the viewer to technical trickery (Tinwell et al., 2011). Realistic and seemingly autonomous characters could raise the question of whether we might not all just be soulless machines (MacDorman et al., 2009) and small but visible mistakes in appearance and motion could trigger an innate fear of dying and associated psychological defenses for coping with the inevitability of death (Greenberg et al., 1994), an idea first proposed by Sara Kiesler (in MacDorman (2005)). There is therefore a possibility that the perceived realism could be touching on our own personal and human identity (Ramey, 2005).

However, this very possibility that we are reacting from a personal fear and make harsher judgments of photo-realistic characters might be pointing to another important view of perception. Is there a possibility that realistic virtual characters could potentially activate processes of social cognition similar to the ones that are triggered when viewing real people? Which ones would that be and how would that influence peoples' attitudes towards them? Also, as described in the neurological evidence, there is a fundamental difference in perception between stylised as opposed to real people, where social cognition does not seem to be automatically processed when viewing stylised characters. In the next section, we explore the aspects of social cognition when perceiving virtual characters.

2.3.4 Virtual Characters and Social Cognition

There is a general consensus in neuroscience that the human brain is flexible and constantly adapting to new stimuli, not just by making functional but also structural changes (Draganski et al., 2004; Sagi et al., 2012). The observation and engagement with the virtual worlds and characters is a recent achievement of our culture and the human perceptual system is still adjusting to this novelty. Evidence from neuroscience shows, that when the brain is faced with novelty, it will search through the already acquired (familiar) information and compare it to the new observation: it does so by predicting response and generating errors if the prediction has not been met (Rao & Ballard, 1999; Friston, 2010; Saygin et al., 2012). If the system works by predicting an outcome, it minimizes the amount of energy used to analyze information. This process is called the "free-energy principle" which states that any self-organizing system that is at equilibrium with its environment must minimize its free energy (Friston, 2010). Perception is therefore an active process that involves generating predictions about the environment, as well as the brain's own states (Bar, 2009; Barsalou, 2009).

Similarly, in cognitive neuroscience, one of the explanations of flexible processing is believed to be attained through the operation of two complementary mental modules: the neocortical and the hippocampal learning/memory systems (McClelland et al., 1995). The first is a slow-learning mechanism, forms stable perception of the world and is resistant to change (see schema in perception, Figure 2.11). The second is a fast learning system and forms a temporary representation of novel or surprising stimuli. These stimuli can then go to the neocortical processing pathway through the process of consolidation and modify the stable perception of world. The separation of these two processes is a fundamental requirement of the perception process (Macrae et al., 1999; Sherman et al., 1998; Smith & DeCoster, 2000) and the most obvious effect on the perception of other people is that it activates *categorical thinking* (Allport, 1979). This means that rather than thinking about people in the terms of their unique constellations of attributes, perceivers prefer instead to construct them on social categories (e.g. race, gender, age) about which they already have a general knowledge. To derive evaluations about a target subject, often a stereotype-based judgment is formed (Brewer, 1988). The profit of having these judgments are obvious - they allow for fast and automatic processing of information (care should be taken with the word automatic - it is used as an umbrella term for all features that are unconscious, efficient, uncontrolled (Klauer et al., 2012)). This speed of processing was explored in research using

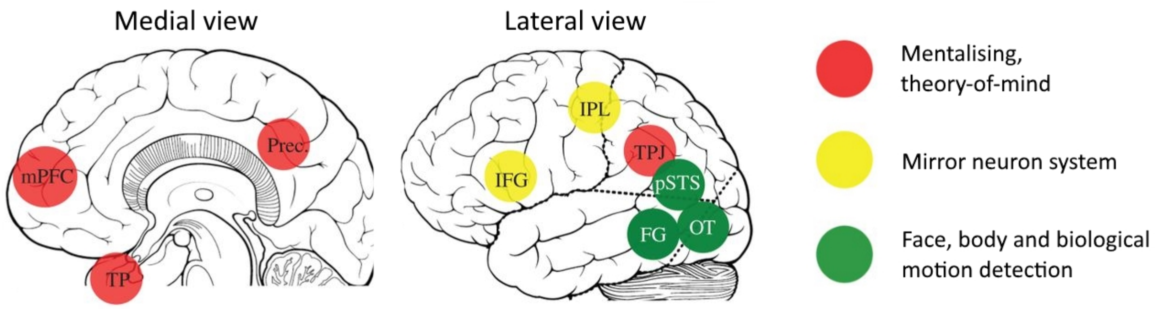


Figure 2.14: Brain areas associated with the network which is consistently activated during social tasks. The area mPFC is shown on the of the image (*left*), whereas only a portion (posterior part) of STS is labeled (*right*), however it extends over the whole depression (sulcus). Image taken from Cross et al. (2015).

various methods of social cognition, such as the Implicit Association Test (Greenwald et al., 1998; Gawronski et al., 2003) where people had a faster response to information which were rich in stereotypes (Klauer et al., 2012).

In addition to the dual process theory, when people think about others in a social way, unique neural patterns are triggered, particularly the mPFC (medial prefrontal cortex area in the brain) which activates with social cognition, and STS (superior temporal sulcus), associated with judgments of intent from interpreting motion cues (see Figure 2.14). This is true even if people are thinking about objects or animals (Mitchell et al., 2005), and as previously mentioned people have a good ability of anthropomorphizing (Epley et al., 2007). Fascinatingly, people can more easily think of animals as people than they would do about people from the "edge of society", such as drug addicts or homeless people (Harris & Fiske, 2006). Therefore as opposed to making objects human-like through anthropomorphism, we can also do the opposite – dehumanize humans (Haslam, 2006), especially because we perceive them as being from a different group (Demoulin et al., 2009) or because they behave bizarrely, such as people with mental disorders (Kramer & Buck, 1997). In short, those whom people seek to exterminate and therefore have no motivation or desire to understand and thus have no chance of future interaction with are the most likely to be dehumanized (Osofsky et al., 2005). Neurologically, people from the outgroup trigger neurological response of disgust (e.g. insula) but not patterns of social cognition (Mitchell et al., 2005).

Since we can think socially about objects, we should under the right conditions do so for stylised virtual characters as well. Intuitively, this goes against the formerly mentioned difference when perceiving real as opposed to stylised characters, where we seemingly use social cognition for real people but not stylised characters. However, the answer could be in the two different processing speeds - fast and slow. With real people, the process of forming social thought is fast, based on former knowledge and experience. Stylised characters can trigger emotional response and in turn we are able to associate social states to them due to our incentive to anthropomorphise, but the path is not automatic – evidence shows that people need to be prompt to *think* about cartoon characters in a social context to trigger the mentalising process (Han et al., 2005). Something similar could be happening to characters created to appear realistic. There is evidence that human like character will load the cognitive process (Chaminade et al., 2007; Saygin et al., 2012), which means that the processing of realistic characters is not automatic but it engages slower processing pathway. One study used an indirect method which was based on the time it takes for the brain to process information (Haeske, 2016). The reasoning behind such a design was that longer exposure to the stimuli would provide enough time for the decision to be made about eeriness since an uncanny character would need more time to be cognitively processed. In their experiment, participants rated the eeriness of computer-generated faces that varied in human likeness. The presentation times of the stimuli varied – a couple of seconds, 100ms and unlimited. However, the results showed the time of the exposure to the stimuli did not change the eeriness ratings of the virtual characters, signifying that the processing of these characters is fast and not slow as thought before. However, even though the exposure timing and the way it links to processing information was well justified, we argue that the way the response was collected from the participants could be problematic. In their design, participants were able to think about the answer after, which could engage the mentalising process, even if the image was shown briefly. We conclude, that the exposure time of the stimuli is not as important as the response from the participant, which should not be consciously processed.

Additionally, cognitive load itself is not a predictor of negative response since new evidence shows that higher cognitive difficulty (slow process) can be an amplifier of negative or positive affect (Albrecht & Carbon, 2014), that the stimuli which should fall in the uncanny valley due to difficulty in categorising them as real or not real faces do not induce negative response in observers (Cheetham & Jancke, 2013; Cheetham et al., 2014). Also, due to the brain adaptability hypothesis, where there is a necessary interplay between both fast and slow systems (Macrae & Bodenhausen, 2000), slower processing should not automatically predict negative response.

Based on these arguments, we formulate an extended model of perception specifically for virtual characters (see Figure 2.15). Following this model, an additional methodological strategy is considered for the study of virtual characters, which was introduced in Chapter 4. We argue that based on the evidence, people are able to form social cognitive constructs about virtual characters but this process may not be automatic and it can yield unexpected outcomes. In addition, subjective responses might not be enough to understand people's attitudes towards virtual characters since social cognitive process are not always accessible

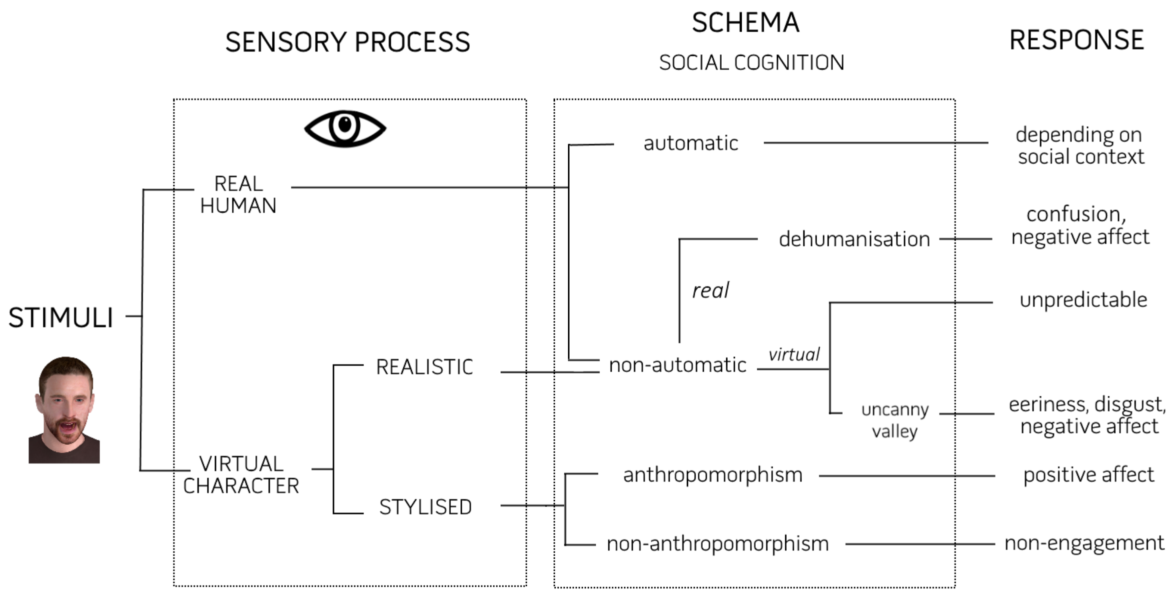


Figure 2.15: Our extended model of perception based on social cognitive schema for virtual characters and real people. First, we separate the perception of real humans as opposed to virtual characters. A virtual character could hypothetically be perceived as real which would indicate that the automatic (fast) mental schema for humans was triggered. In the case of non-automatic (slow) processing the scheme for virtual characters is shared with the human, but for human it will lead to dehumanisation, whereas for virtual characters it may lead even further to uncanny valley or it might have an unexpected, even positive outcome. We further make a distinction between realistic and stylised characters, since the latter do not share the same social schema with real humans and the processing of social information is always non-automatic.

to conscious processing (Klauer et al., 2012). We propose the use of indirect methods in addition to subjective responses to extract more information about the perception of virtual characters. In the next section we describe methods which were used in our research.

2.4 Methodology in Perception of Virtual Characters

Peoples' attitudes towards virtual characters can be measured in various different ways. For low level bottom–up processes, brain studies using fMRI and EEG can be used. Experimenters can also monitor participant's heart rate, respiratory rate and skin conductance to track changes in anxiety levels of people who are observing the character. Peoples' eye gaze can give a lot of information on their attention to particular areas of the character, indicating areas of interest or disturbance. These measures are referred to as physiological measures and have many advantages: they are quantifiable and do not require participants conscious evaluation. Reasons against using these measures could be poor accessibility and cost of the machines, additional expertise for analysing the results and non–direct association between physiological and mental response.

Peoples' responses can also be collected through observation and qualitative measures. These methods are helpful when we do not have much knowledge about a particular problem we are investigating and do not know how to approach it. Qualitative approach will provide a wide range of data but it will be difficult to analyse in a concise way and subjected to noise.

The most commonly used measures, however, are subjective responses, where people are asked to answer questionnaires, make a comparison or decision. Subjective responses are usually obtained by questionnaires. Likert scales and semantic differential scales are used in the attempt to quantify data. The first asks about the intensity of a statement (e.g. "On a scale from 1 to 7, how eerie is the character?") while the second gives the opposite descriptors on each end of the scale (e.g. response scale ranges from *good* to *bad*). These scales have certain disadvantages. Likert and differential scales show diversity in how people map the responses on the scale - they avoid extremes of the scale, make changes along the way on how they assess stimuli, etc. (Bartz et al., 2008). Including repetitions of similar but not exactly the same scales to provide repetitions of the same answer improves reliability of these measures. To avoid subjective mapping of answers, a forced choice task can be used, where a limited range of options is given, and the participant must choose one of them that mostly represents his opinion. In the Two–Alternative Forced Choice (2AFC) experiment design, speed and accuracy of choices between two alternatives given a timed interval are tested (Blackwell, 1952). This method is also an indirect measure of response, since it does not include conscious processing. However, limiting the responses a participants can use, results in the loss of useful information (Bartz et al., 2008).

There is currently only one standardised measure which measures attitudes towards artificial humans and this is the Godspeed Questionnaire, introduced by Bartneck et al. (2009) and revised by Ho et al. (2010). This instrument uses 4 indices with high internal reliability – warmth, humanness, eeriness and attractiveness. This terminology is commonly used to measure the mentioned uncanny valley. A lot of research studies however do not use this same terminology and the lack of universally agreed upon operational definition for attitude towards virtual characters is a critical issue of the field (Kätsyri et al., 2015).

In our research we present a couple of methodological approaches, commonly used in social psychology, which were used alongside subjective measures. Particularly, we present the attribution bias (Gilbert et al., 1988) and a behavioural measure of social psychology – proximity (Bailenson et al., 2003).

2.4.1 Attribution bias

To study the response to virtual characters based on the ease of cognitive processing and at the same time avoiding the time–based responses presented by Haeske et al. (2016), we decided to use the approach of social cognitive methods. One promising example comes from studies of social cognition. When people try to understand the behaviour of other people (process of attribution), the general consensus in research is that their mental process will go through at least two stages – spontaneous identification and deliberate controlled inference (Trope, 1986; Quattrone, 1982; Gilbert et al., 1988). For example, when observing aggressive behaviour, people will automatically identify the person as aggressive and then consciously assess if the situation made the person behave aggressively. In this way they will "correct" the formerly established automatic perception of the person as aggressive by deliberately processing the situational context. Experiments conducted by Gilbert et al. (1988) provided an explanation for this process. His hypothesis was that conscious attribution to situational factors requires more cognitive power, therefore, when people are busy with another mental task they will fail to examine the context and would attribute behaviour to the person and not the situation. In one of their experiments, the participants listened to a person, who was assigned a pro– or anti–abortion speech. The nonbusy group of participants only observed the person while the second group needed to formulate their own speech at the end of the clip, therefore they were cognitively busy planning their response as they were observing the person in the clip. The participants who were not busy used situational information to explain the behaviour of the person –namely, he was assigned the particular essay portion by others. However, the busy participants were unable to use the situational information, therefore their attribution relied heavily on the person in the clip himself and not the situation that made him behave so. This shows that people under high cognitive load will be more likely to attribute behavioural responses to the person and people without the same cognitive load will be free to examine the situational factors and therefore correct their response.

How could this effect the attribution of behaviour to virtual characters? Based on the fast and slow mechanisms of processing, the realistic characters trigger a prediction error which would need to be corrected. While the outcome of that is unpredictable (could be positive or negative), the process itself could interfere with the attribution of the behaviour to situational factors and associated them more with the character. In the case of stylised characters, cognitive difficulty would not be expected since these characters do not trigger the social cognition automatically (Han et al., 2005), hence we should not expect prediction based errors which would slow down the processing of visual information. Therefore, we expect to see a different pattern of attribution bias for characters which attempt to reach realism as opposed to the stylised characters, which would be more likely to be attributed situational rather than personality origin of their behaviour.

2.4.2 Proximity

Personal space is a the distance between two or more people. Experimentally, it can be studied in many ways but the common idea is to measure the distance to the target person which the participant chooses, which is still comfortable for him or her. One approach is to study the distance participants choose when they are instructed to sit near the target person. Another way is analysing stop distance, where participants are instructed to indicate when the target person, which is approaching them, should stop moving. In projective studies, participants manipulate other objects, such as dolls, to indicate their personal space, and natural observation studies, where experimenters observe peoples' proximity patterns when they are engaged in real–life social situations (Hayduk, 1983). There are several factors which influence personal space, including culture (Hall, 1966), race (Rosegrant & McCroskey, 1975), physiology (McBride et al., 1965), age (Willis Jr, 1966), interpersonal relationships (Evans & Howard, 1973) and social attraction (McCroskey & McCain, 1974). Invasion of personal space would result in anxiety, stress or discomfort of the person who is being approached. In situations where there is a need for the proximity rules to be broken (such as crowded trains) usually eye contact is avoided to prevent further discomfort.

Some researchers started to use virtual reality to study personal space (Krikorian et al., 2000; Bailenson et al., 2001; Sommer, 2002). The works of Bailenson et al. (Bailenson et al., 2001, 2003, 2004) gave the area of proxemics another dimension by moving the observation into the virtual environment. This allowed the behaviour of people to be measured extremely accurately in the highly controlled virtual environment and at the same time keep the ecological validity (Bailenson et al., 2003). The usual setup includes a task that prompts participants to move close to the virtual character, presented in virtual space. This was achieved by asking participants to memorise certain features of the characters' clothes. While participants were moving, their trajectory was accurately recorded and minimal front and back distance was used to asses the level of comfort people take when approaching a virtual human. It was found that participants rarely come closer than 40 cm to the character, they approach them closer when they move towards their back (Bailenson et al., 2001), people keep similar distance from agents (behaviour guided by computers) and avatars (behaviour manipulated by a real person) when the level of engagement is low (Blascovich, 2002), otherwise they tend to keep further away from avatars. A very important part of low-level engagement which makes a great difference in how close people move is apparently eye-gaze (Blascovich, 2002; Bailenson et al., 2003).

In our studies, we employ the proximity measure as a behavioural response to virtual characters. We do this since self-report questionnaires are effective when measuring how people perceive an embodied agent, but not necessarily how they will respond to him, as shown by the study of Bailenson et al. (Bailenson et al., 2005). In the same study, interaction effects were found for behaviour realism and visual realism of the character, indicating that appearance does affect how people perceive the characters. In particular, the results showed that realistic characters elicited embarrassment reactions from participants. According to our extended model of perception (see Figure 2.15), this could signify the triggering of social schema, similar to when we engage with real people. This is crucial for our work, where we use more advanced realistic characters and additional measures of social cognition. We will also compare this behavioural measure with a self-reported level of co-presence, which according to Bailenson et al. (2005), is measuring the same aspect as proximity but provides information on the level of cognitive processing.

2.5 Information From Natural Human Motion

The main focus of this thesis is to study how differences in appearance of the virtual characters affect our perception of them. However, as already mentioned in Section 2.3.3, there are important interactions with appearance and motion - motion can amplify a negative or positive response to the character. Since motion capture is used throughout the presented work for the creation of stimuli, more information on perception of biological motion is provided in the current section.

When building virtual characters, natural human motion can be used as an animation technique. Movement of the body or its parts makes a substantial contribution to nonverbal communication. The early work exploring biological motion (Johansson, 1973) was made possible by putting lights on parts of the human body in a darkened room. To the observer, when the lights were static, they resembled a meaningless cluster of dots, but when moving, these dots were grouped to create the perception of a human form engaged in a readily identified activity (walking, running, dancing, ect.). These point-light displays (PLD) proved that humans are extremely capable of recognizing and categorizing human motion even with very little information. The use of PLD technique has been applied to animation in graphics, where human motion is captured using a motion capture system (see Figure 2.8). Studies have shown that the biological motion of humans is rich with information on gender of the mover (Kozlowski & Cutting, 1977; Johnson & Tassinary, 2005; Hill et al., 2003), his emotional state (Ekman, 1992; Atkinson et al., 2004) and even personality (Argyle, 1988; North, 1975; Hyde et al., 2013).

2.5.1 Gender

Several studies have shown that gender can be identified by observing body motion alone. For example, men and women have a different way of walking, and these differences are apparent to the observers, even when observing only the ankle motion from point-light walkers (Kozlowski & Cutting, 1977). Studies identified that hip sway often indicates female motion, while shoulder movement indicates male motion (Johnson & Tassinary, 2005). Similar studies found that recognition of sex is possible from different motions as well, e.g. throwing (Johnson et al., 2011) or conversation (McDonnell & O'Sullivan, 2010).

Other studies also focused on the recognition of gender from facial motion, using either androgynous faces (Hill & Johnston, 2001; Morrison et al., 2007) or point-light displays (Hill et al., 2003). These studies found that gender can be recognized from facial motion only, but that specific cues, such as excessive nodding, blinking and overall amount of movement, play an important role in recognizing the female sex (Morrison et al., 2007). Similarly, Hill et al. (2001) showed that head motion alone is less useful for discriminating sex than facial motion.

Some studies that focused on motion analysis from virtual characters found that appearance of the character interacts with the overall perception of the motion as well. This does not come as a surprise, given that humans heavily rely on appearance in every day life. Even when we are not familiar with a person, we make judgements based on the perceived sex, the way he or she is dressed, skin appearance, hairstyle, etc. While many studies would use PLD's or genderless models to avoid interaction between motion and appearance, some studies explored this relationship. For example, body shape has an effect on the perceived sex of walkers (Johnson & Tassinary, 2005; McDonnell et al., 2009). An exaggerated male or female body shape of the virtual character thus influences the way sex is perceived from the walking motion (McDonnell et al., 2009).

2.5.2 Emotions

The motion component is particularly important when we want to build emotions onto a virtual character. People will engage with the character and be motivated to observe it for longer if it accurately expresses emotions. This is because the perception and interpretation of other people's emotion is essential for effective social interaction. And since we put so much importance on emotions, the ability to recognise and distinguish between different emotional states has a prominent role in perceptual processes.

In the field of psychology and neuroscience, many definitions of emotions exist and the exploration of their origin and development is an ongoing research focus (Lewis et al., 2008).

The most general definition describes emotions as subjective experiences, where the core feeling is that of pleasure or pain (Frijda, 1988). Usually, the role of emotion is the regulation of behaviour (moving towards or away from a particular object) through the activation of different bodily responses. Several approaches to emotion classification exist in the literature as well, from defining emotion as discrete categories (Ekman, 1992) or as dimensions (Mehrabian, 1980; Plutchik, 2001). Ekman's approach to emotions (1992) was aimed at identifying emotions which are universally recognized and where similarity in their physical expression can be observed. He classified them as basic emotions: anger, happiness, sadness, fear, surprise and disgust. This classification also provided a simple as well as systematical approach for the study of emotion recognition from motion, which provided a comprehensive way to map emotions onto virtual characters.

The development of emotion studies on virtual characters originate from studies using PLD's for recognition of movement. Emotions can be identified from either body motion (for full body motions see for example (Atkinson et al., 2004; Crane & Gross, 2007), upper body motions (Volkova et al., 2014) and arm motions (Pollick et al., 2001)) or facial motion (Bassili, 1978). Similar results have been found on motion applied to fully modeled virtual characters (Ennis et al., 2013). Additionally, recognition rates are improved when body and facial motions are combined (Clavel et al., 2009), showing that both facial and body cues have a specific role in conveying emotion. Research also shows that there are particular areas of the face (Cunningham et al., 2004; Schyns et al., 2009) and certain body motions (De Meijer, 1989) that convey most of the information about the portrayed expressions, e.g. corners of the mouth are important in recognizing happiness, which suggests that the recognition of emotion can be improved when this behavioural data is present.

The perception of emotion is highly dependent on the perceiver as well. For example, females are much better at recognition of emotion and better at expressing them as well (Hall, 1979). Using video references of female and male actors portraying emotions, Battocchi et al. (2005) found that emotions are better recognized on female actors overall. Also, since emotions have a role of regulating behaviour, perceiving someone's face which is expressing anger or fear alerts the observer that the situation might be threatening to him as well. Brain studies show that some areas in the brain, specifically the amygdala, tends to activate while viewing fearful and angry faces, as opposed to happy, surprised, and sad faces (Calder, 1996). Since the amygdala region is activated in response to danger, it is believed that negative emotional expressions, such as anger and fear, trigger a defense response in the perceiver. This reaction is also related to empathy, which differs from the mentioned concept of mentalising (Singer, 2006) by relating to someone else emotionally rather than mentally. Empathy can be described as matching of affective experience between a participant and a target individual (Rameson & Lieberman, 2009) and can be tested with brain imaging techniques (Rameson & Lieberman, 2009), behavioural response (Bouchard et al., 2013) or subjective reports (Davis, 1983).

One of the explanation of reduced appeal of the virtual characters approaching realism is also that the appearance and motion imperfections interfere with efficient emotion perception from faces (Tinwell et al., 2011) or dampen the expression intensity (Hyde et al., 2013). The effect of stylisation on emotion detection however seems to rely more heavily on the type (or abstraction level) of stylisation (Wallraven et al., 2007). Our research further investigates the interaction between appearance and motion and their effect on emotion perception.

2.5.3 Gender Stereotypes and Emotion

The perception of gender and emotion from motion is closely related. Gender stereotypes impact both the production and perception of emotions (Brewer, 1988). Many studies done with subjects living in Western societies show that emotions are perceived to be gender specific (Hess et al., 2004; Fischer et al., 2004; Plant et al., 2000). Overall, in these societies females are believed to express more emotions than men but some emotions are more male specific, e.g. anger, contempt and pride (Plant et al., 2000). It is also believed that females express more emotions which signal powerlessness, e.g. sadness and fear (Fischer et al., 2004). Studies conducted on static images (Hess et al., 2004) also found differences in the way the same emotional expressions are perceived on female and male static faces. This is particularly the case for happiness and anger. While expressing anger, women were rated as more angry than men, and in the happy condition, men as happier than women. Hess et al. (2004) attributed this finding to the contrast effect, since a woman showing anger or a man expressing happiness might violate the viewers expectation of gender stereotype and therefore would intensify the perceived emotional expression. Using video references of female and male actors portraying emotions, Battocchi et al. (2005) found that emotions are better recognized on female actors overall, especially anger and sadness. While both studies explored the interaction between gender and emotion, neither separated appearance of the model from dynamic motion. They used either video references, where the sex of the actor was apparent, or static images and thus left out the component of dynamic motion. The study of Johnson et al. (2011) combined both aspects by exploring sex recognition bias on the perception of throwing a ball under different emotions using point-light displays. They found that an angry throw is perceived as more male and a sad throw as more female, which confirms previous findings. However, this study only focused on throwing motion, not taking into account that different types of motion carry more gender cues which could influence the bias. In our research (see Chapter 3.1) we tested gender bias on motion with obvious gender cues (walking) and less obvious cues (conversation).

2.5.4 Personality

In order to create a more complex and engaging behaviour of virtual characters, one can consider designing them to express personality traits. Here, we first provide a definition of personality as described by research in psychology, and then present studies from mixed fields, some of them using virtual characters, which contributed to this particular domain.

Personality refers to individual differences in characteristic patterns of thinking, feeling and behaving. The theory of personality has been the subject of intensive study in psychology research. Due to its continuous examination and re-evaluation, the "Big Five" theory (see for example (Goldberg, 1990; Costa & McCrae, 1992; John et al., 2008)) is perceived by many to be the standard description of human personality. The Big Five is a hierarchical model of personality traits with five broad factors (extraversion, agreeableness, conscientiousness, openness to experience, and emotional stability). Each bipolar factor (e.g., extraversion vs. introversion) is further described by specific facets and traits (e.g. extraverts are talkative, sociable, but people with low emotional stability are seen as anxious and easily upset). Some studies showed that there are correlations between facets of the Big Five, and condensed the original model to the "Big Three", "Big Two" and the "General Factor of Personality (GFP)" (Musek, 2007). GFP is summarized as the difference between the personality that can be described as having high scores on all the facets of the Big Five and the personality that scores low on these facets. Musek (2007) described the Big One as an optimum blend of all socially valued dimensions.

Personality does not just contribute to the variation of human behaviour, it can also to some extent predict it (Paunonen & Ashton, 2001). Therefore, if a person reacts to a certain situation in a distinct way, it can be predicted that he or she will react that way in a similar situation as well. This predictive behaviour can be deducted from observable behaviour, self-ratings or ratings of others. Studies from different areas of science show that people can assess personality of other people even at zero-acquaintance (Borkenau & Liebler, 1992; Back et al., 2011; Mehl et al., 2006), based on physical appearance from photos of real people (Naumann et al., 2009), type of motion (Neff et al., 2010; Argyle, 1988; North, 1975) and even use of language (Gill & Oberlander, 2003; Dewaele & Furnham, 2000).

Research conducted with synthetic motion and/or virtual characters found that certain traits of the personality can be expressed through body and facial motion (Argyle, 1988; North, 1975). Extraversion is known to be related to exaggerations in body motion and the general speed of body motion, while introverts use fewer outward directed gestures and touch themselves more (Argyle, 1988). However, determining how to represent such affective and individual qualities in a computational framework remains an active research problem (Neff et al., 2010). Increased facial motion has been found to make the character more extraverted (Hyde et al., 2013). Some inadequacies in a character's motion can also transfer to the observed personality of the character. For example, Tinwell (2013) found that virtual characters with inadequate upper facial animation exhibit personality traits associated with psychopathy. Gaze behaviour was found to be very closely linked to personality as well. Various studies using eye-tracking and video recordings analysed the connection between the known personality traits of participants and their eye gaze behaviour. These studies also confirmed a link between eye motion and Big Five traits (Rauthmann et al., 2012; Matsumoto et al., 2010; Franks, 1963).

While studies of personality and the appearance of virtual characters are yet to be explored, there are many indications that appearance influences the perceived personality, as observed in real people (Borkenau & Liebler, 1992; Naumann et al., 2009; Back et al., 2011). General neatness and looking healthy is related to conscientiousness, while the persons that exhibit the very opposite appearance are more prone to be perceived as open to experience (Naumann et al., 2009). Extraversion and agreeableness is more evident from dynamic cues, such as having a smile on the face.

While examining the perception of personality, we need to take into account that the observers of personality are known to be biased. Studies from psychology show that a person's perception of another person's personality is influenced by their own personality (Serfass & Sherman, 2013; Paunonen & Hong, 2013), especially when the information about the person observed are sparse (Ready et al., 2000) or when the observed person exhibits traits which the observer is sensitive about (e.g. neurotic observers will be more sensitive to perceptions of criticism (Serfass & Sherman, 2013)). There is also an effect of gender – women are known to be better decoders of both posed and spontaneous nonverbal behaviour and are better in encoding it as well (Hall, 1979; Ambady et al., 1995).

For our research in Chapter 4, introducing longer exposure times to video clips where a virtual character was expressing personality traits, was an important novelty and therefore a difficult task. We relied on the Big Five theory when creating personalities for virtual characters and we treated the high score on the Big Five factors as "positive" personality traits, and their polar opposites as their "negative" personality traits, since there is indication that the Big Five factors could be described with either socially attractive or unattractive traits (Musek, 2007). We relied on the outcomes of the above mentioned studies to direct a professional actor and map his performance to a virtual character and study the effect of appearance on the overall perception of this character.

2.6 Summary of Chapter 1

In this chapter, we presented research which is related to the studies conducted and presented in the following two chapters. The history and character production pipeline gave an example of how characters can be created – some of the techniques which were used in creation of our stimuli also. Perception of virtual charters, the main interest of our research, gave an overview of the appearance components of virtual characters. We also presented our extended theoretical model of perception on which we based some of our measurement choices (subjective ratings, attribution bias, proximity) and the motivation to studying complex behaviour of the character, such as emotion and personality, which can be faithfully expressed through observable visual cues, e.g. motion. The next chapters describe our research, where specific research questions are explained in more detail in the introduction of each chapter.

Chapter 3

The Effect of Mismatch on the Perception of Virtual Characters

The main focus of this chapter is to explore how mismatches of visual elements effect the perception of virtual characters. Previous research has found that mismatches are particularly crucial for the loss in appeal, be that the incompatibility of texture and mesh (MacDorman et al., 2009) or motion and appearance (Saygin et al., 2012). The effect can be attributed to predictive coding (Saygin et al., 2012), where brain processing creates prediction errors when presented with opposing visual information and induces a negative response from the viewer. However, the effect of mismatch on perception needs further testing, particularly when deciding which elements of appearance are more crucial for the loss of appeal and what other outcomes could a mismatched presentation have on the perception of virtual characters.

In this chapter, we present two studies where we intentionally created a form of mismatch when creating the virtual character. For the first study, we applied motion from male and female actors on the virtual characters of opposite sexes. We were not only interested how this would affect the perception of gender but also emotion categorisation, since it is known that emotion perception is affected by gender stereotypes (Johnson et al., 2011). This way we also investigate how social cognition is involved in the perception of virtual characters, as outlined in section 2.3.4. The second study is more closely concerned with the uncanny valley effect by exploring the mismatch in realism between shape and material and how this influences the perceived appeal. In addition, we also investigated how the mismatch in shape and material could affect the recognition and intensity of emotions.

3.1 Motion and Shape Mismatch

The creation of virtual characters far exceeds the natural world with respect to matching animation with appearance properties of the character. This helps us to create never before seen creatures, or provides practical solutions for production, e.g. using motion from the same actor on different characters. However, there are some limits to this possibility. It is known that people can recognize the underlying sex of the actor since motion is rich in gender cues, so the sex of the actor should be matched with the character sex. In this study, we explore how a mismatch of appearance and motion sex could bias our perception of realistic virtual characters, and how this is further mediated by the type of motion and emotion that the character is expressing.

We build upon studies which showed that sex and emotions can be recognized from motion and that gender stereotypes for emotional expressions influence the perception of sex. Johnson et al. (2011) studied this relationship on throwing motion, whereas we drew from the studies on recognition rates for different types of motion. For conversational motion, focusing on facial motion (Hill & Johnston, 2001; Hill et al., 2003), above chance recognition rates for both sexes were found (ranging from 58% to 61.9%). However, recognition rates from walking motions were generally found to be higher (around 70% (Kozlowski & Cutting, 1977)), and mainly due to the information coming from shoulder and hip motion. In this study, we therefore compare gender ratings for walking and conversational motions and expect the ratings for conversational motions to less accurately match the actual actor's sex. Furthermore, since we expect our conversational characters to exhibit less gender cues than walking characters, we hypothesise that emotions expressed through conversation would affect the heuristic of sex judgments to a greater extent than they would for walking motions.

Finally, studies so far investigated emotion and/or sex recognition on point-light displays or one virtual model to display the motions, whereas previous work has shown that the choice of virtual model affects the perception of both motion (Chaminade et al., 2007) and sex of the virtual character (McDonnell et al., 2009). For this reason we used two models, one male and one female, and mismatched the gender of the underlying motion and sex of the model to examine if this would dampen or exaggerate gender and emotion recognition.

To test our assumptions, we conducted two experiments. In Experiment 1 we studied gender perception for basic emotions expressed while walking, where we expected gender judgments would not be highly affected by the displayed emotion. Experiment 2 was designed to test the same assumptions but for conversational motions, where we expected gender stereotypes around specific emotions would influence the perception of gender to a greater extent than in the previous experiment. We conclude by comparing results and discussing the outcomes of both experiments.

Figure 3.1: Examples of motion capture as performed by a male and a female actor, expressing happy (*left*) and sad (*right*) walks. The actors were free to express their own interpretation of the emotion.

3.1.1 Stimuli Creation

To address the question of gender and emotion perception and interaction on virtual characters, we chose to use natural motions captured on real actors, in order to capture the subtleties and timings of real movements. The motion capture data used for both experiments in this study was acquired using a 21 camera Vicon optical system, where 52 markers were placed on the body. The motion capture data was taken from 4 female and 4 male actors (see examples of actors in Figure 3.1). Two virtual characters of different sex were used to display the motions of the actors (*Man* and *Woman* model). We used congruent (where character sex matched actor sex) and in-congruent (where character sex did not match actor sex) stimuli. Other details of stimuli creation are discussed separately for Experiment 1 and Experiment 2.

For Experiment 1, actors were instructed to portray emotions through their walk. We did not capture face or finger motion. The captured body markers were used to compute the joint angle animations and mapped onto the virtual characters in Autodesk 3ds Max 2014®. We created walking characters, displayed in central, full-body view and walking towards the camera. The faces of the virtual models were blurred out to direct participant's attention towards the body motion. Also, static facial expressions might influence the overall emotional display of the walk. Eighty 3-second video clips were generated in total: 2 Models (Woman, Man) $\times$ 8 Actors (4F, 4M) $\times$ 5 Emotions (Angry, Fear, Happy, Sad + Neutral). Movies were created at 30 frames per second and at 1240 × 900 resolution and displayed on a 24-inch LCD screen.

For Experiment 2, we considered facial motion to be extremely important for recognition of conversational emotions, therefore we added 36 markers on the face to capture facial motion alongside body motion. The actors were asked to remain stationary in the middle of the recorded capture space. The basic emotions were conveyed through acting out sentences that we selected from a validated list of affective sentences for spoken emotion identification (Ben-David et al., 2011) e.g. “Get out of my room!” for anger or “It’s a beautiful day outside!” for happiness (for full list of sentences, see Table 3.1). From this validated list, we chose three sentences for each emotion (including Neutral). Since we wanted to examine the effect of motion alone, the voices of the actors, although captured, were not used in this experiment. No eye or finger motion was captured. As before, the body markers were used to compute the joint angle animations and then applied to virtual characters. The facial motion was exported separately to the body as 3D marker motion. A bone-based approach that used linear blend skinning was used to drive the facial geometry. In order to retarget the motion from actor to model, the markers were aligned to the head and then automatically adjusted to the position of the bones of the face of the character. The characters’ facial bones were then constrained to their corresponding optical markers to produce the animation. We generated 3 clips for each emotional expression, since different sentences offer more variety in their interpretation. Two hundred and forty 3-second clips were generated in total: 2 Models (Woman, Man) × 8 Actors (4F, 4M) × 5 Emotions (Angry, Fear, Happy, Sad + Neutral) × 3 Sentences. The character was displayed in the center of the screen, facing forward at the beginning of each clip. We selected a medium close up view as we wanted to provide information from both the facial movements and upper body of the actors, while the legs were stationary. Movies were created at 30 frames per second and at 1240 × 900 resolution and displayed on a 24-inch LCD screen.

3.1.2 Experiment 1: Walking

In this experiment, we were interested in evaluating to what extent emotions would influence how observers perceive the gender of walking actors. Since evaluating the influence of emotions on gender perception judgments depends on the accurate portrayal of emotions by actors (and therefore the accurate recognition by viewers), we also included an emotion recognition task in this experiment to ensure the expected emotions were recognized above chance level through the walk.

Emotion	Happiness	Sadness	Anger	Fear	Neutral
Sentences (Ben-David et al. (2011))	It's a beautiful day outside! I am going on holidays! This is the best day of my life!	Nobody sat beside me at lunch today. My best friend is moving away. This weather is depressing.	Stop what you are doing and listen to me! Get out of my room! I'm sick of you being late	Look out, there is a car coming! This place is creeping me out. Someone is following me.	My spoon is on the table. There are magnets on the fridge. Containers have a blue lid.

Table 3.1: A subset of emotional sentences taken from Ben-David et al. (Ben-David et al., 2011) that were used in our experiment.

Experiment Design

Twenty-four participants (15M-9F; average age: 24) took part in the experiment. They were recruited mainly from the university, had normal or corrected to normal vision and were naïve to the purpose of the experiment. Also, University ethical approval was granted for the experiment, and participants received a €5 book voucher to compensate for their time. Participants were asked to read and sign the informed consent and read through the information sheet, where they were familiarized with the details of the experiment. Afterward, they were seated in front of a computer screen and asked to follow the instructions displayed before them. The experiment was divided into two blocks: a) Gender Rating and b) Emotion Recognition. In each block, participants saw one model with all motions applied first, followed by the same motions on the other model. We randomized whether participants viewed the Man or the Woman model first. Clips were randomized within each block to avoid ordering effects. Participants saw 80 animations in each block: 2 Models (Woman, Man) $\times$ 8 Actors (4F, 4M) $\times$ 5 Emotions (Angry, Fear, Happy, Sad + Neutral). An example of stimuli is shown on Figure 3.2.

a) Gender Rating

In the first block, participants were asked to rate the gender of the character, based on motion cues. They were informed that the displayed motion would not necessarily match the sex of the model. After each clip, they rated Gender on a 5-point Likert scale from 1 (Very Male) to 5 (Very Female) using the number keys on a keyboard.

We decided to use the term *gender* for the rating task and the term *sex* for the actual underlying motion (male, female) based on the distinction between *gender* as a sociocultural

Figure 3.2: Example of stimuli for Experiment 1. A Woman and a Man model with male actor's motion applied, portraying Fear (left), and female actor's motion portraying Sad (right).

construct associated with maleness and femaleness and *sex* as a demographic category (Unger, 1979). Because we instructed the participants to pass judgment on a 5-point scale of how female or male the motion was, and did not ask to categorize the motion as either male or female, our task is closer to the sociocultural use of the term. However, our study does not presuppose that the origin of the distinction in motion between males and females is of sociocultural or biological nature, as pointed out by Deaux (1993).

b) Emotion Recognition

We introduced the second block to explore and evaluate the accuracy of emotion portrayal by the actors. The effect of emotion on gender can only be justified if the emotions are correctly identified above chance level.

In this block, participants categorized the displayed emotion. They were asked *"Which of the 5 listed emotions is the character expressing?"* and selected an emotion by pressing the corresponding key on the keyboard, marked with A (Anger), F (Fear), H (Happy), N (Neutral) and S (Sad). We instructed them to use the Neutral condition only when they believed that the character did not express any emotion. They were also informed at the beginning of the experiment that one of the listed emotions would always be portrayed in each clip.

Results

Results were analyzed using repeated measures ANalysis Of VAriance (ANOVA) with within-subjects factors: *Actor Sex*, *Model*, and *Emotion*, and between-groups factor *Participant Sex* on all of our data. In all cases, we averaged over the data for the four actors for Actor Sex. For significant differences, estimates of effect size are reported using partial eta-squared. Where main interaction effects were found, we conducted Newman-Keuls tests for comparison of means to further explore the results. Degrees of freedom have been corrected using Greenhouse-Geisser estimates of sphericity (marked with an asterisk *) and Levene's test was used for testing homogeneity of variances. For clarity, we present the results grouped for each block (Gender Rating and Emotion Recognition) with the results relevant to our research in the form of questions and answers.

a) Gender Rating

We first investigated the results for gender ratings. The participants were able to distinguish between male (Avg: 2.27±0.08) and female (Avg: 3.76±0.10) motions overall ($F(1, 22) = 85.63, p \approx 0$). This ability to rate the gender of the motion was unaffected by the character model used to present the motion. Overall, it is possible to recognize gender from walking motions, as previous studies have shown (Johansson, 1973; Kozlowski & Cutting, 1977).

Does emotion affect gender rating? Ratings for gender of the motion were affected by Emotion ($F^*(2.66,\ 58.52) = 10.17,\ p \approx 0,\ \epsilon = 0.67$). Post-hoc analysis showed that Angry (Avg: 2.68±0.07) was considered significantly more male than Sad (Avg: 2.91±0.08), Happy (Avg: 3.08±0.05), Neutral (Avg: 3.13±0.06) and Fear (Avg: 3.26±0.10), with $p < 0.01$ in all cases. Sad was also found to be significantly more male than Fear ($p < 0.002$) and Neutral ($p < 0.05$). These results show that *different emotional conditions affect gender ratings.*

An interaction also occurred between Actor Sex and Emotion ($F(4, 88) = 17.43, p \approx 0$) (see Figure 3.4), where post-hoc analysis showed that all emotions portrayed by female actors were rated as significantly more female than their male counterparts ($p < 0.001$ in all cases). For females, Sad and Angry ($p < 0.03$) were rated as the least female emotions, followed by Fear ($p < 0.03$ in all cases). Neutral and Happy were rated as most female emotions ($p < 0.03$ in all cases) where Happy was rated as more female than Neutral but the difference is not significant. For males, Fear was rated as the least male motion ($p < 0.001$ in all cases), Sad and Neutral were rated equally ($p < 0.009$ in all cases) but less male than Happy ($p < 0.008$ in all cases). Angry was considered the most male of all motions ($p < 0.001$ in

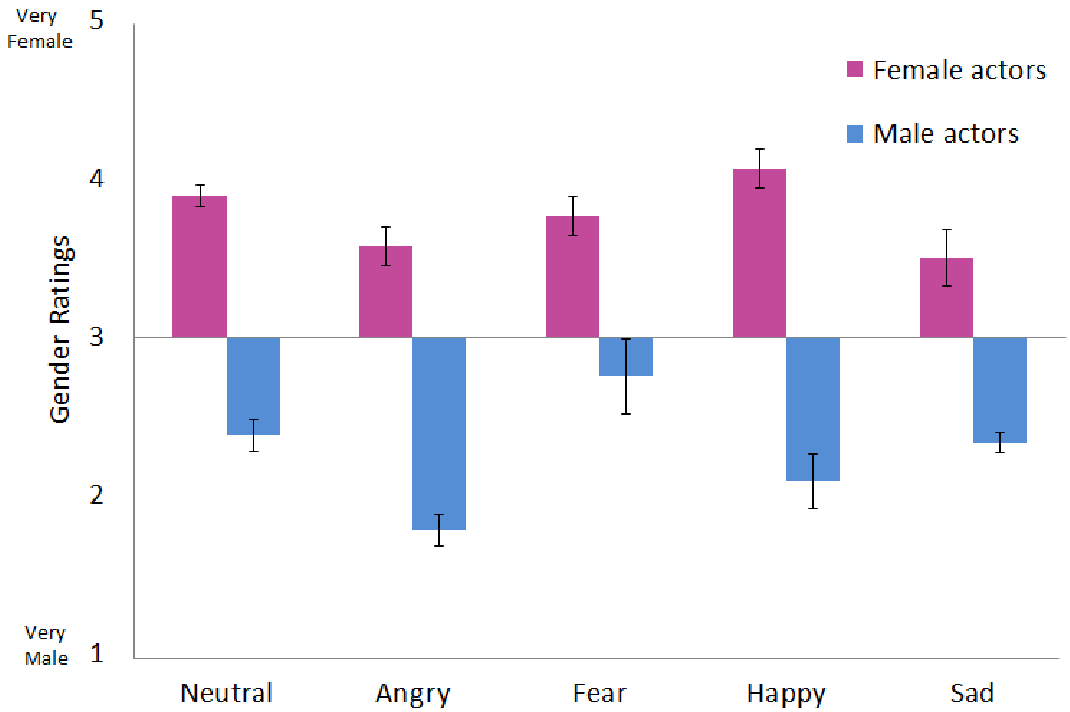


Figure 3.3: Experiment 1: An interaction between Actor Sex and Emotion according to gender ratings for walking motions. Error bars show standard error of the means.

all cases). Angry was therefore perceived as the most male on male and the least female on female motions, which confirms the studies on the effect of gender stereotypes on production and perception of emotion (Fischer, 1993) and previous findings on gender bias from body motion (Johnson et al., 2011). Therefore, gender ratings for some emotional conditions show a possible influence of stereotypes on gender perception. This perception bias is further discussed in Section 6.2.

To determine whether participants could identify gender of the motion from different emotional conditions, we decided to compare the mean gender ratings for male and female motion with the average rating, which indicates the absence of visible gender cues. To test this, we conducted one-sample t-tests for each Emotion $\times$ Actor Sex combination (i.e., gender ratings significantly different from the constant 3). The ratings for male and female Angry, Neutral, Sad and Happy were significantly different from the middle estimate of gender ($p \approx 0$ for all). Male Fear was seen as less male but still significantly different from being rated as ambiguous ($t(15) = -2.6, p < 0.02$). We therefore conclude that *emotion affects gender judgments of walking motions but these changes never differ greatly from the actor's actual sex.*

Does appearance of the model influence gender ratings? There was no main effect of Model on gender ratings, but an interaction between Model and Emotion was found

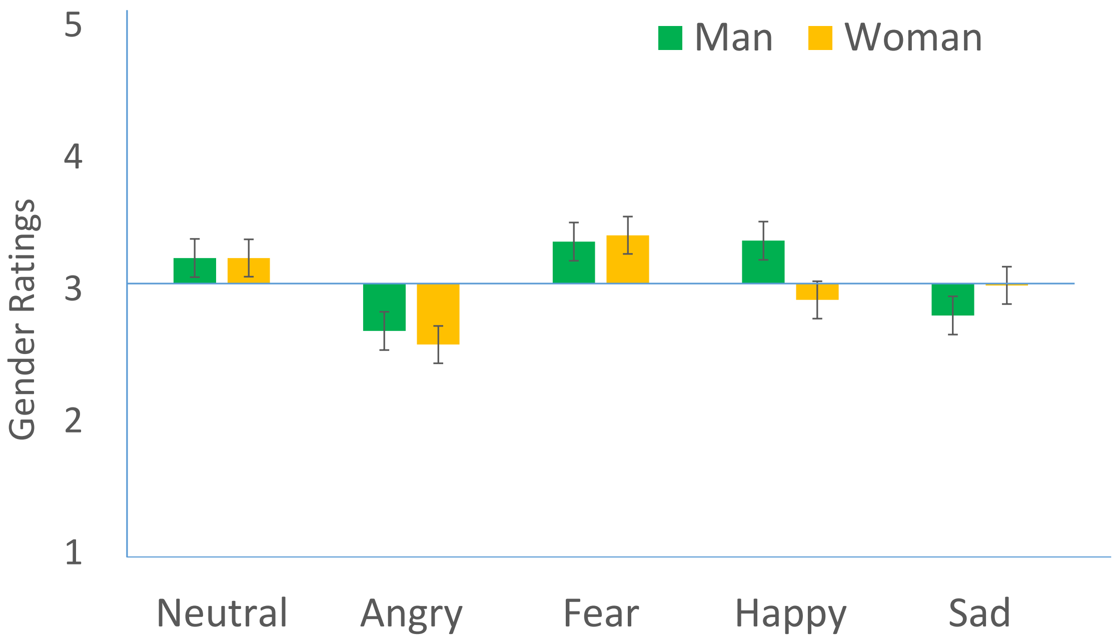


Figure 3.4: Experiment 1: An interaction between Emotion and Model according to gender ratings for walking motions. Error bars show standard error of the means.

($F(4, 88) = 3.17, p = 0.017$), where Happy and Angry were seen as more male on the Woman model than on the Man model ($p < 0.02$ for all). Previous research (Hess et al., 2004) showed that anger on a female face looks more intense than on a male face due to the contrast effect, since anger is perceived to be a male emotion. Our results might be showing this contrast effect as well, since Angry on the Woman model appeared more male. However, our results did not present the expected contrast effect for Happy, since it was perceived more female on the Woman model.

There was also an interaction between Participant Sex, Actor Sex and Model ($F(1, 22) = 11.13, p = 0.003$). Female participants could identify male walks on the Woman model more easy than on the Man model ($p < 0.008$). A similar result was found for male participants, who found female walkers easier to identify on the Man than on the Woman model ($p < 0.02$). These results might also be related to the above mentioned contrast effect (Hess et al., 2004) as it appears to have been easier for participants to identify the gender of the actor of opposite sex displayed on an incongruent model. However, it remains uncertain as to why the sex of the observer matters in this interaction. We can conclude that *the appearance of the model affects perception of gender*. These results could also be highly related to the contrast effect, however further experiments will be necessary to fully explore this relation.

b) Emotion Recognition

To explore the role of emotions and to test if the emotions were correctly identified above chance level by the participants, we analyzed results for emotion recognition.

Were basic emotions correctly portrayed by walking motions? Overall, participants were very accurate at identifying emotions (70% accuracy on average), with 20% being the chance level. However, we found a main effect of Emotion ($F^*(2.35, 51.58) = 20.76, p \approx 0, \epsilon = 0.59$) and post-hoc analysis showed that Neutral and Sad were significantly more recognized than any other emotion ($p < 0.005$ in all cases). The next most recognized was Fear ($p < 0.03$ in all cases), followed by Angry and Happy which were the least recognized emotions ($p < 0.03$ in all cases). It is possible that the Sad emotion was easiest to recognize, due to the significant head cue (lowering of the head) which could only be present for this emotion. Neutral walk was highly recognized but other emotions were more likely to be confused with Neutral as well (see Figure 3.5), especially in the case of Happy and Angry. This result is similar to the study of Crane and Gross (2007) where they found sadness being the most and anger the least recognized of the tested emotions. Since we have not used facial motion for our stimuli, this effect could be attributed to the lack of facial information, which is further discussed in section 3.1.4. Overall, our results show that *basic emotions were accurately portrayed through walking motion.*

Does sex of the actor affect the recognition of emotion? Since studies have found that females are better at portraying emotions (Brody & Hall, 2000), we analyzed our results for possible differences in emotion recognition in relation to actor's sex. We found a main effect of Actor Sex ($F(1, 22) = 41.14, p \approx 0$), which showed that participants were more accurate at emotion recognition when viewing female (77.29%±0.04) than male (63.54%±0.03) motion. An interaction between Actor Sex and Emotion ($F^*(2.84, 62.37) = 27.40, p \approx 0, \epsilon = 0.71$) showed that this was due to the fact that emotions Angry, Fear and Happy were more easily recognized on the female actor ($p < 0.003$ in all cases). Participants confused both male Angry and Happy walk with Neutral, and male Fear walk was often mistaken with Sad. The recognition rates and missclassifications for each emotion according to actor's sex are seen in the confusion matrix (Table 1). This shows that *emotions expressed by females through walking are easier to recognize than emotions expressed by males.*

Does the sex of the virtual model change the perception of emotion? Appearance can also influence emotion perception (Hess et al., 2004), so we were interested in

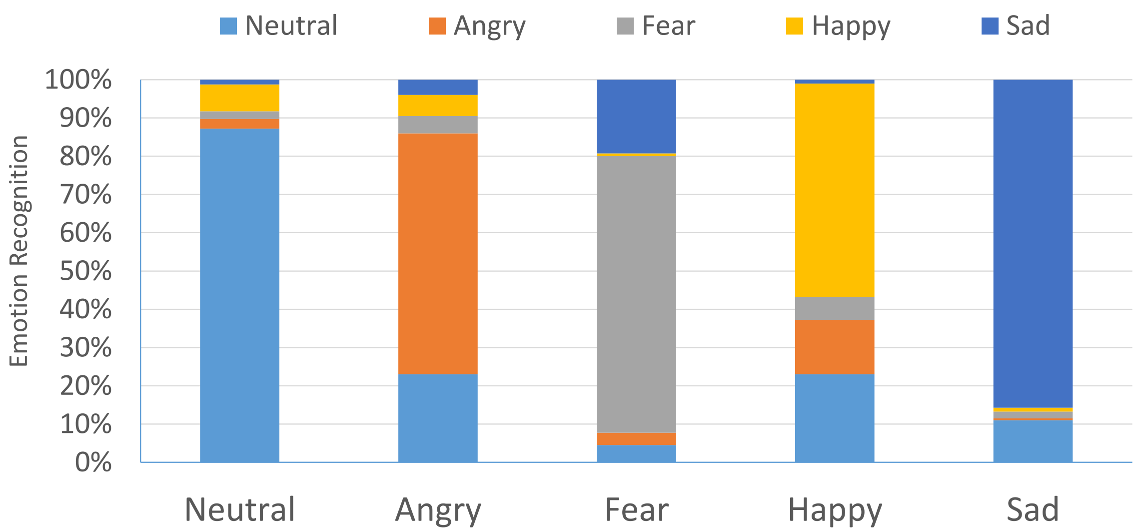


Figure 3.5: Experiment 1: Main effect of Emotion (*horizontal axis*) on emotion recognition (*coloured stacks inside the bars*) for walking motions. Happy, even though recognized above average (over 20%), was least accurately identified and was confused with Neutral and Angry. Angry emotion was confused with Neutral.

the differences in emotion recognition for the models we used. There was no main effect on the accuracy of emotion recognition, but we found a slight interaction with Emotion ($F(4, 88) = 2.48, p < 0.049$) where the Man model was slightly better for recognizing Fear (73% on the Man and 67% on the Woman model) but was worse for recognizing Angry than the Woman model (58% on the Man and 64% on the Woman model; $p < 0.05$ for all). Therefore, *the appearance of the model has a slight influence on how specific emotions are perceived.*

3.1.3 Experiment 2: Conversation

In Experiment 1, we found that the underlying sex of the actor is recognized even if actors are expressing emotions that could bias the perceived gender. We hypothesize that this finding is due to the existence of strong motion cues in walking that make the perception of gender less prone to emotional bias. Therefore, in this experiment we focus on another type of motion, conversation, which does not rely as heavily on the physiological differences between sexes as walking does. We studied the same interaction: how basic emotions affect gender perception of conversational motion and how the choice of model affects this interaction. As in Experiment 1, we also assessed the accuracy of the recognized emotions by asking the participants to report which emotions the characters were expressing.

Figure 3.6: Stimuli for Experiment 2. Left: Female motion applied to the Woman and the Man models (Happy emotion). Right: Male motion applied to both models (Angry emotion).

Experiment Design

Similar to Experiment 1, participants rated the gender of the character based on their perception of motion in the first block. In the second block, they were asked to categorize the type of emotion the virtual character was expressing. In both blocks, participants were presented with the same set of randomized 240 video clips depicting virtual characters conveying different emotional sentences (Figure 3.6). In each block, half of the participants viewed the Man model stimuli first and the other half viewed the Woman model first. Trials lasted 3 seconds each and participants were not presented with audio. In total, participants saw 240 animations in each block: 2 Models (Woman, Man) $\times$ 8 Actors (4F, 4M) $\times$ 5 Emotions (Angry, Fear, Happy, Sad + Neutral) $\times$ 3 Sentences. Sixteen participants (8M-8F; average age: 31.6) took part in each block of this experiment. Fifteen participants completed both blocks (7M-8F) and two additional participants completed just one block each. As before, they were recruited mainly from the university, had normal or corrected to normal vision and were all naïve to the purpose of the experiment. University ethical approval was granted for the experiment, and participants received a €5 book voucher to compensate for their time.

Results

As in Experiment 1, results for both the Gender Rating and Emotion Recognition tasks were analyzed using repeated measures ANalysis Of VAriance (ANOVA) with within-subjects factors: *Actor Sex*, *Model*, and *Emotion*, and between-groups factor *Participant Sex* on all of our data. In all cases, we averaged over the data for the four actors for Actor Sex, and over the three sentences. For significant differences, estimates of effect size are reported using partial eta-squared. Degrees of freedom have been corrected using Greenhouse-Geisser estimates of sphericity (marked with *) and Levene's test was used for testing homogeneity of variances. Where interaction effects were found, we conducted Newman-Keuls tests for comparison of means to further explore the results. No effect of Participant Sex was observed for any tests conducted so this is not discussed further.

a) Gender Rating

As in Experiment 1, we found that participants were able to distinguish between male (Avg: 2.53±0.03) and female (Avg: 3.49±0.05) motions overall ($F(1,14) = 244.26, p \approx 0$). This implies that it is possible to recognize gender from conversational motions as well, which confirms the findings of previous studies (Hill et al., 2003).

Does emotion affect gender rating? We were mainly interested in the effect of emotion on gender ratings. We found that the ability to rate the gender of the motion was affected by Emotion ($F^*(1.60, 22.43) = 19.43, p \approx 0, \epsilon = 0.40$). In general, Angry (Avg: 2.47±0.11) was considered significantly more male than Neutral (Avg: 3.03±0.08), Fear (Avg: 3.09±0.05), Happy (Avg: 3.24±0.05), or Sad (Avg: 3.20±0.07). There were no significant differences in gender ratings for other emotions. As in Experiment 1, we found that *emotions affect gender ratings.*

An interaction occurred between Actor Sex and Emotion ($F(4,56) = 53.20, p \approx 0$) (Figure 3.7) as well, where post-hoc analysis showed that all emotions expressed by female actors were rated as significantly more female than their male counterparts ($p < 0.001$ in all cases). For females, Angry was rated as significantly less female than all other emotions ($p < 0.001$ in all cases). Neutral was rated as more female ($p < 0.001$ in all cases), followed by Fear and Sad, which were rated equally ($p < 0.001$ in all cases). Happy was rated significantly more female than all other emotions ($p < 0.001$ in all cases). For males, Sad and Neutral were rated equally as the least male of all emotions ($p < 0.009$ in all cases). Fear was considered more male ($p < 0.009$ in all cases), followed by Happy ($p < 0.003$ in all cases), and finally Angry was rated as the most male of all male emotions ($p < 0.003$ in all cases). Therefore, it might be more difficult to identify gender when a female actor is expressing anger, and for male expressions of sadness and neutral sentences. Happiness, however, appears to facilitate the recognition of gender for both males and females. The possible influence of stereotypes on gender perception is further discussed in Section 6.2.

As in Experiment 1, we checked whether participants could identify gender by conducting one-sample t-tests for each Emotion $\times$ Actor Sex combination (i.e., gender ratings significantly different from the constant 3). We found that for Angry, participants could not determine the gender of the female motions, but could do so for male motions ($t(15) = -9.63, p \approx 0$). For Neutral and Sad, male motions did not convey gender information, but female motions did (Sad: $t(15) = 5.93, p \approx 0$; Neutral: $t(15) = 3.14, p < 0.007$).

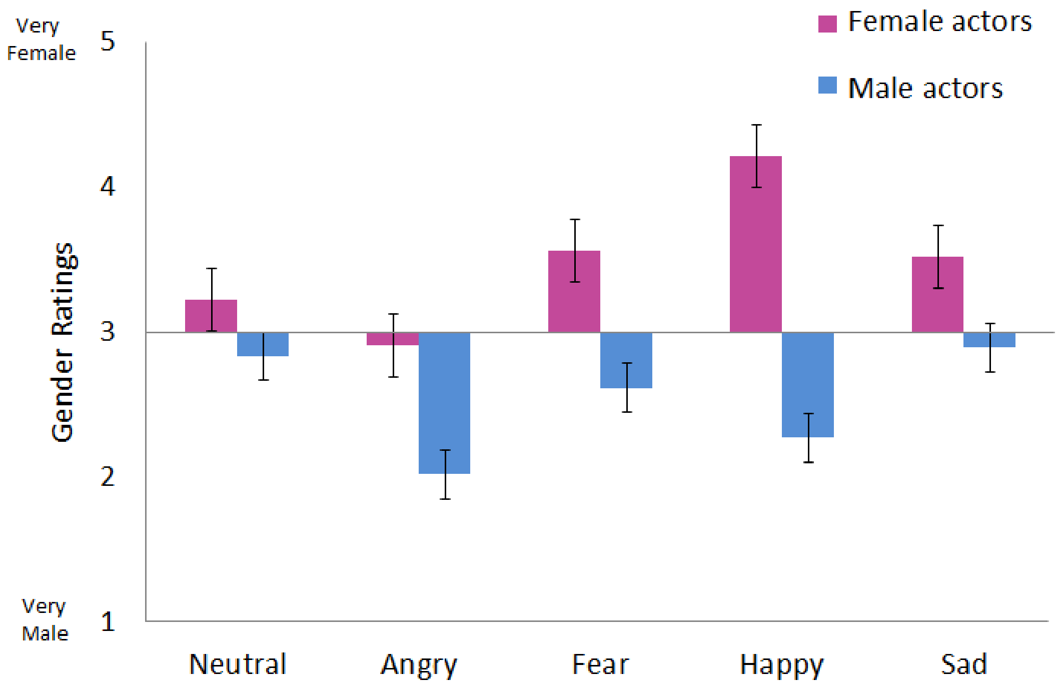


Figure 3.7: Experiment 2: An interaction between Actor Sex and Emotion on gender ratings for conversational motions. Error bars show standard error of the means.

For Happy and Fear, male motions were recognized as male (Happy: $t(15) = -10.72, p \approx 0$; Fear: $t(15) = -6.15, p \approx 0$) and female motions as female (Happy: $t(15) = 13.81, p \approx 0$; Fear: $t(15) = 7.93, p \approx 0$). Therefore, some gender specific emotions (female anger, male neutral and male sadness) prevented an accurate recognition of gender. However, female actors were generally not rated as male and male actors were not perceived to be female. We conclude that *emotions affect gender judgments of conversational motions but do not cause misperceptions of gender.*

Does appearance of the model influence gender ratings? Finally, the model used to present the emotions did not affect ratings (no main or interaction effects). These results imply that *appearance does not affect gender perception of conversing characters.*

b) Emotion Recognition

As in Experiment 1, we included the emotion recognition task and validated that our sets of conversational motions are correctly identified above chance level.

		EMOTION RECOGNITION (Experiment 1)					EMOTION RECOGNITION (Experiment 2)				
		Neutral	**Angry**	**Fear**	**Happy**	**Sad**	**Neutral**	**Angry**	**Fear**	**Happy**	**Sad**
FEMALE actor	**Neutral**	82.8%	1.6%	4.7%	8.3%	2.6%	59.1%	5.2%	3.4%	3.2%	29.2%
	Angry	10.9%	73.4%	5.2%	9.4%	1.0%	0.5%	95.6%	2.1%	0.0%	1.9%
	Fear	4.2%	1.0%	92.7%	1.0%	1.0%	1.6%	6.5%	79.2%	1.3%	11.5%
	Happy	21.9%	3.6%	9.9%	63.5%	1.0%	1.8%	5.5%	0.8%	89.1%	2.9%
	Sad	17.2%	2.1%	4.7%	2.1%	74.0%	11.7%	4.5%	8.4%	1.1%	74.5%
MALE actor	**Neutral**	86.5%	3.1%	1.0%	6.3%	3.1%	46.4%	9.1%	2.1%	9.4%	33.1%
	Angry	36.5%	48.4%	5.2%	2.6%	7.3%	0.3%	96.1%	2.9%	0.3%	0.5%
	Fear	6.8%	6.8%	47.9%	1.6%	37.0%	8.9%	24.2%	52.4%	3.4%	11.3%
	Happy	26.6%	24.0%	1.0%	44.8%	3.6%	3.2%	9.1%	0.3%	86.8%	0.8%
	Sad	7.8%	0.0%	1.0%	1.0%	90.1%	24.0%	8.9%	9.2%	4.2%	53.9%

Figure 3.8: Confusion matrix: Emotion Recognition for walking motions (Experiment 1) and conversational motions (Experiment 2), averaged over Model.

Were basic emotions correctly portrayed by conversing characters? We found that overall, participants were very accurate in identifying the acted emotions (68% accuracy on average), with 20% being the probability of recognition being due to chance. We also found a main effect of Emotion ($F^{*}(1.49, 20.82) = 35.39, p \approx 0, \epsilon = 0.37$) and post-hoc analysis showed that Angry and Happy were significantly more recognized than any other emotion ($p < 0.001$ in all cases). Sad and Fear were rated equally accurately, and Neutral was the least recognized emotion ($p < 0.009$ in all cases). As in Experiment 1, *basic emotions are recognized from conversational motions but with different accuracy.*

Does the sex of the actor affect the recognition of emotion? We found a main effect of Actor Sex ($F(1, 14) = 179.90, p \approx 0$), which showed that participants were more accurate at emotion recognition when viewing female (79.48%±1.36) than male (67.09%±1.48) motions. An interaction between Actor Sex and Emotion ($F(4, 65) = 28.96, p \approx 0$) showed that this was due to the fact that all emotions except Angry and Happy were more easily recognized on the female actor ($p < 0.04$ in all cases). The confusion matrix shows where misclassifications occurred (Table 1). From these results, we can conclude that *females are better at portraying conversational emotions than males.* Specifically, male sadness, fear and neutral portrayals are less recognizable from conversation, in our study. Anger is stereotypically a male emotion so it is not surprising that it was correctly portrayed by males. Even though happiness is more attributed to females, happiness might be easier to portray with facial motion by both males and females. This explanation is supported by the study of Ennis et al. (2013), where they demonstrated that facial motion is important in disambiguating happy and angry emotions from body motions.

Does appearance influence emotion recognition? Model was found to have an effect on accuracy ($F(1, 14) = 9.18, p < 0.009$), with motion recognition being more accurate on the

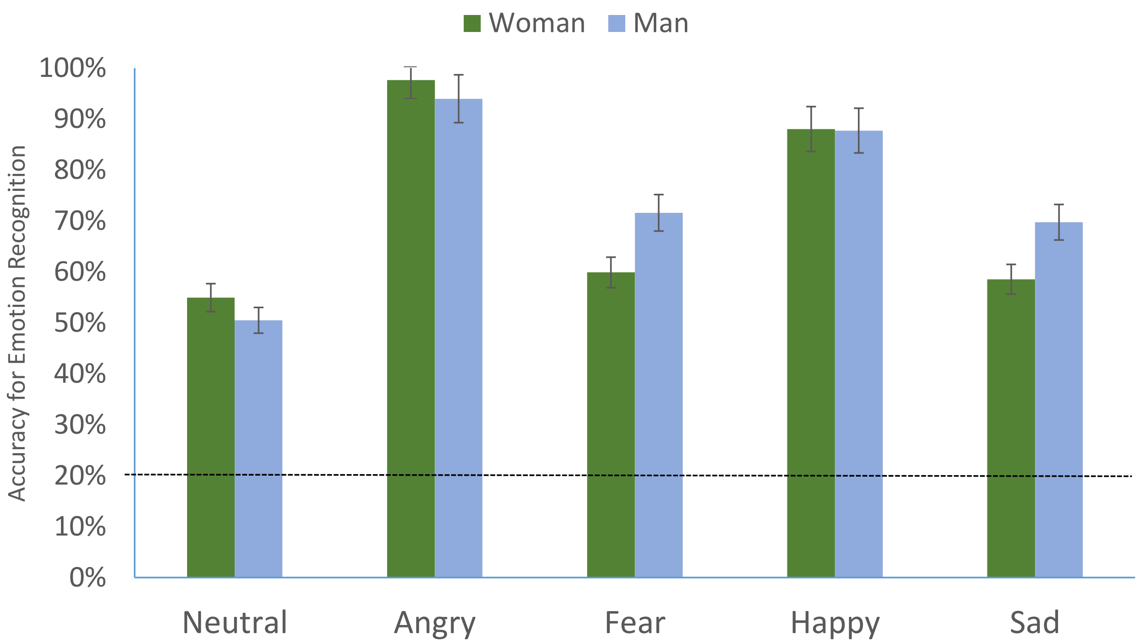


Figure 3.9: Experiment 2: Interaction between Emotion and Model on emotion recognition for conversational motions. Fear and Sad were better recognised on the Male model. We can also see that Angry and Happy were most recognised emotions overall. The dashed line presents chance level of recognition (20% accuracy), error bars show standard error of the means.

Man (74.74%±1.45) than on the Woman model (71.83%±1.41). We also found an interaction between Emotion and Model ($F^*(2.00, 28.03) = 10.88, p \approx 0, \epsilon = 0.50$), which was due to the fact that Fear and Sad were detected more easily on the Man than on the Woman ($p < 0.001$ for both), see Figure 3.9. Therefore, *virtual model has an influence on how specific emotions are perceived.*

3.1.4 Cross-experimental analysis

We conclude our analysis of results by comparing the results of Experiment 1 and Experiment 2. This was done to evaluate if the type of motion (walking and conversation) had any effect on how emotions affect gender perception. We hypothesized that conversational motion would be more prone to gender bias because it does not contain as many gender cues as walking motion, so the gender would be more difficult to determine and emotions would have a stronger effect on gender ratings. We also compared emotion recognition accuracy for both motion types.

Results

To understand how the two different types of motions, walking and conversing, affect gender and emotion interaction, we further analyzed the results of Experiment 1 and Experiment 2 in relation to each other. We first conducted repeated measures ANalysis Of VAriance (ANOVA) with within-subjects factors: *Actor Sex*, *Emotion* and *Model*, and between-groups factors *Motion Type* and *Participant Sex* on all of our data from both experiments. In Experiment 2, we used 3 different sentences to portray each emotion for variety, as our actors portrayed conversational emotions differently across sentences. In Experiment 1, we used 1 example of a walk per emotion, since our actors were consistent in their portrayal of emotion across walks. Therefore, participants in Experiment 2 saw 3 times more stimuli than in Experiment 1. However, since the order of stimuli was randomized and the task was not one that could be learned as the experiment progressed, we conducted cross-experimental analysis by using mixed model ANOVA. In order to conduct ANOVA, we averaged ratings over the 3 sentences from Experiment 1 to give us an average rating of emotion. As before, we tested for assumptions of ANOVA, using Greenhouse-Geisser correction for sphericity and Levene's test for homogeneity of variance, and used transformations on the raw data where assumptions where not met. If this was not successful, we used a non-parametric tests (Wilcoxon Signed Rank test for repeated measures and Mann-Whitney U test for comparisons of independent samples). Where assumptions were met, Newman-Keuls test was used for comparison of means to further explore interaction effects.

a) Gender Ratings

Main effects were investigated to explore how the effects of emotion, sex of the actor and model could influence the results regardless of the Motion Type for which no main effect was found. A main effect was found for Actor Sex ($F(1,36) = 120.80, p \approx 0$), which shows that gender was rated according to the underlying sex. This result was expected since both previous results from Experiment 1 and 2 showed this effect. The other main effect was found for Emotion ($F^*(3.06, 110.27) = 24.27, p \approx 0, \epsilon = 0.76$). The post-hoc conducted for the second effect indicates that overall, Angry was rated as significantly more male than any other emotion, with the average rating of 2.60 ± 0.06. Two interactions were found as well: between Actor Sex and Emotion ($F(4,144) = 41.10, p \approx 0$) where similar patterns of gender ratings emerged as in Experiment 1 and Experiment 2, and an interaction between Actor

Sex, Model and Participant Sex ($F(1, 36) = 5.31, p = 0.027$), as in Experiment 1.

Is the underlying sex of the actor easier to detect on walking or conversational motions? We assumed that actor's sex would be more apparent from walking than conversational motions, since identifiable sex cues exist in walking motions that make sex identification easier and less prone to gender bias. A significant interaction between Motion Type and Actor Sex occurred ($F(1, 36) = 4.53, p = 0.040$), where males were rated more male and females more female on walking motions than on conversational motions. This result shows that *actor's sex is indeed more apparent on walking than on conversational motions.*

Are conversational motions more prone to the gender bias that exists around emotions than walking motions? We found an interaction between Emotion and Motion Type ($F^*(3.06, 110.27) = 5.34, p = 0.002, \epsilon = 0.76$), where post-hoc analysis revealed that Angry was rated significantly more male for conversers than walkers and Actor Sex, Motion Type and Emotion interaction ($F(4, 144) = 13.33, p \approx 0$), where post-hoc analysis showed that gender was better recognized for walking motions in the case of Angry, Sad and Neutral male motions ($p < 0.02$, for all) and Neutral and Angry female motions ($p < 0.001$, for all). However, no differences were found between Happy and Fear. Gender biases could be observed in Sad (male) and Angry (female) conditions, where the gender of the conversers was perceived as more ambiguous (Figure 3.10) than in the walking condition. We also expected that participants would be better at recognizing gender from Neutral walking condition than in conversation, because of differences in gender cues between conversations and walking, and our results confirm this expectation. We can therefore conclude that *where differences in gender perception between the two types of motion were found they showed a greater influence of emotion on gender stereotypes.*

Does appearance of the model influence gender ratings for both types of motion? As in previous experiments, we investigated the effect of model on gender ratings. We found only an interaction between Participant Sex, Actor Sex, Model, and Motion Type ($F(4, 152) = 2.56, p = 0.041$), again, showing a slight contrast effect when female participants rated male motion on a female actor as more male and male participants rated female motion on a male actor more male. However, this effect was prominent only for walking motions ($p < 0.03$, for both). *Appearance of the model has a slight effect which depends on the type of motion and sex of the observer.*

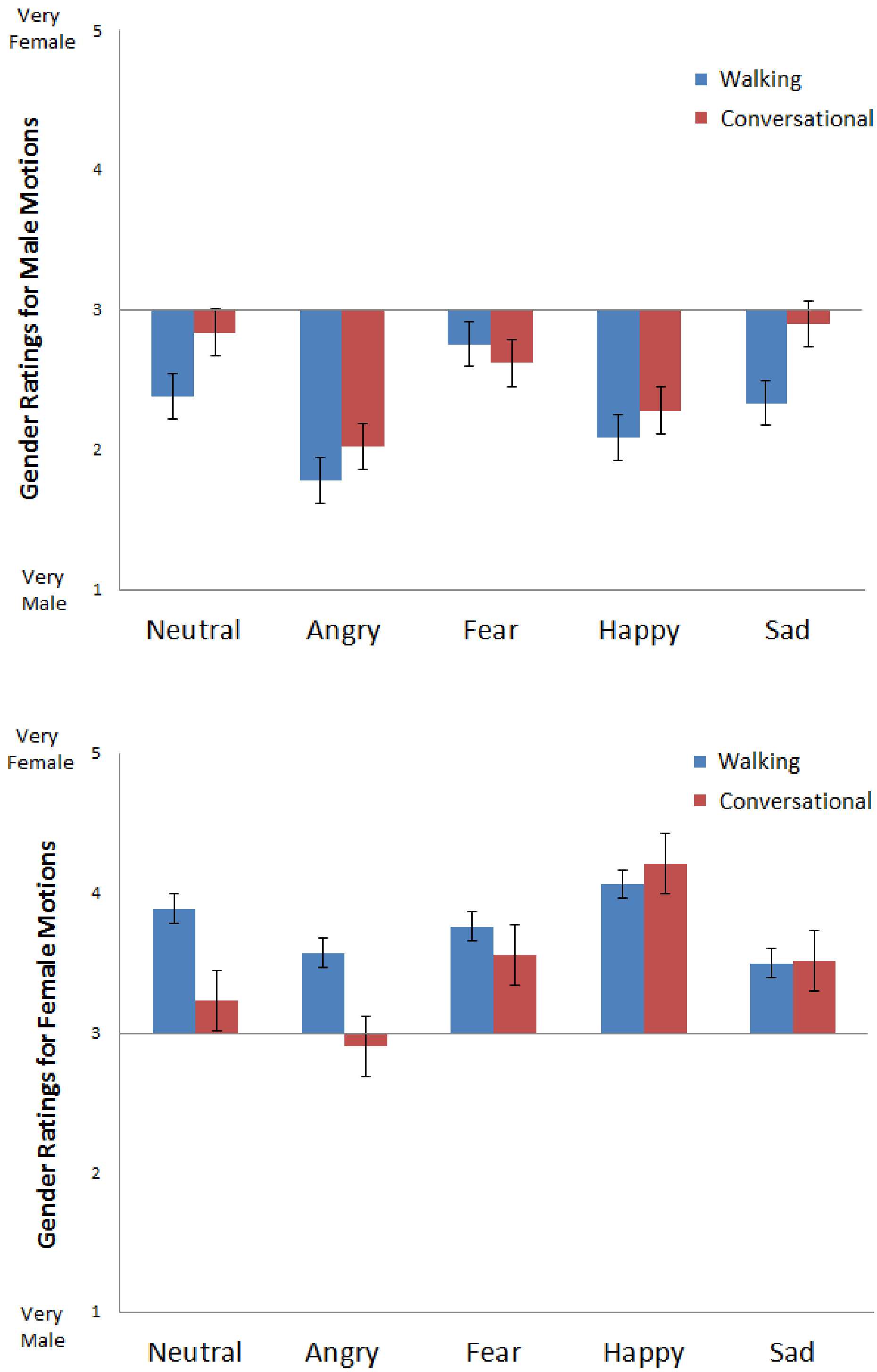


Figure 3.10: Comparison of interactions between gender ratings and emotion for different types of motion (walking and conversation) for male (*top histogram*) and female (*bottom histogram*) motions. Error bars show standard errors of mean gender ratings.

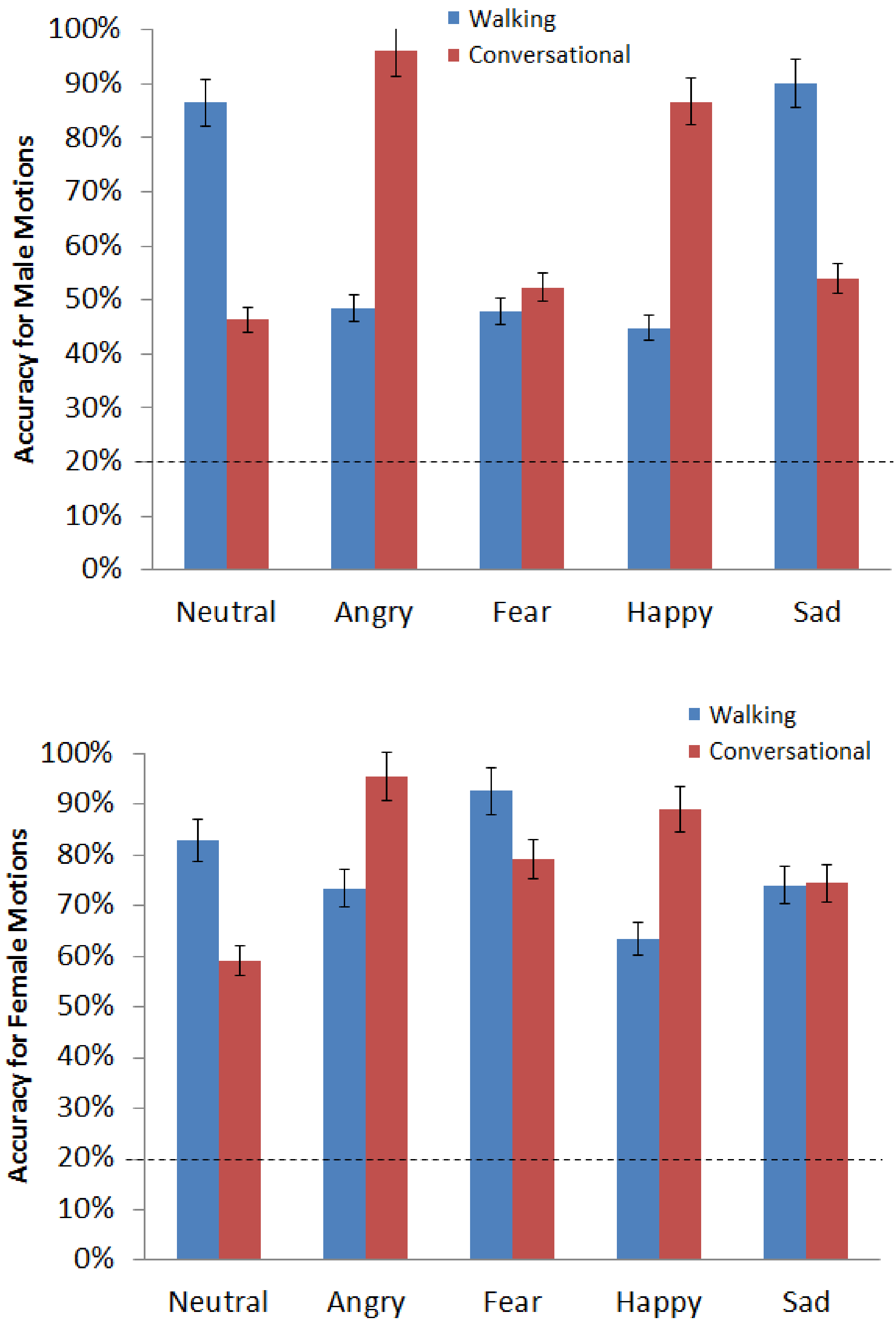


Figure 3.11: Comparison of means between different types of motion (walking and conversation) for male actors (*top histogram*) and female actors (*bottom histogram*) on emotion recognition. The dashed line presents chance level of recognition (20% accuracy), error bars show standard error of the means.

b) Emotion Recognition

For this set of data, the assumptions for ANOVA were not met and transformations on our data were not successful, therefore we used the non-parametric Mann-Whitney U test for independent samples to compare variables of both groups and report significant differences. First, we were interested to determine if there were significant differences on recognition of emotion according to Participant Sex but found no differences. We therefore continued to analyze recognition accuracy for male and female actors separately, averaged over models, to explore the differences between recognition accuracy of the two motion types. Finally, we used Wilcoxon signed-rank test for repeated measures to compare differences in recognition accuracy according to Model.

Are emotions, expressed through different types of motions, recognized with the same accuracy? There are significant differences for all emotions between the two types of motion, except for Sad expressed by female actors and Fear portrayed by male actors (see Figure 3.11). Walking was better for the recognition of Neutral (Male actors: $U = 34.5, p \approx 0$, Female actors: $U = 67.5, p = 0.001$), Fear expressed by females ($U = 50.5, p \approx 0$) and Sad expressed by male actors ($U = 16.0, p \approx 0$). Angry and Happy emotions were better recognized from conversation (Male Angry: $U = 0.0, p \approx 0$ and Happy: $U = 7.5, p \approx 0$, Female Angry: $U = 86.0, p = 0.003$ and Happy: $U = 44.0, p \approx 0$). Overall, *neutral motions are easier to recognize on walking, whereas anger and happiness are more recognized on conversational motions.*

Does appearance of the model influence emotion recognition for both types of motion? Significant differences between Model were found according to Motion Type, Emotion and Actor Sex. Man model improved the recognition for conversational motions in the cases of Sad (for males: $Z = 2.02, p = 0.042$, females: $Z = 2.30, p = 0.021$) and female Fear ($Z = 2.48, p = 0.013$). For walking motions, the Man model improved the recognition for male Neutral motion ($Z = 2.03, p = 0.042$). The Man model also improved accuracy for male Fear for both types of motions (conversation: $Z = 2.95, p = 0.003$ and walking: $Z = 2.31, p = 0.021$).The Woman model improved recognition only for female Angry conversational motion ($Z = 2.40, p = 0.016$). These results might show a slight preference of the Man model when it comes to accuracy, especially for conversational motions and emotions that are stereotypically not attributed to males (Fear, Sad). Similarly, Woman model improved the recognition of Angry conversation, emotion which is attributed to males. Further tests would have to be conducted to understand why the model would have this effect for these specific conditions. Our results therefore indicate that *some models are better for the recognition of emotions according to a particular motion type than others.*

Discussion

The results of the cross-analysis show that for both walking and conversational motions, gender is generally rated according to the underlying actor's sex. Also, both motions presented some variations in gender perception when emotional conditions were introduced. These variations, as we predicted, were mostly in favor of the hypothesis that certain emotions are attributed more to one gender than the other. This effect was most prominent for anger, an emotion commonly attributed to males (Brody & Hall, 2000) and was perceived as more male when expressed by male actors and perceived less female when expressed by female actors. Similarly, sadness and fear, which are commonly attributed to females, caused male motions to appear less male overall. The only emotion that did not follow our hypothesis was happiness, which is an emotion commonly attributed to females. Unexpectedly, happy motions carried gender information even for males. Therefore, it is possible that there are specific ways in which males and females express this emotion that could explain why stereotypes did not have a stronger effect. From our motion capture session it was possible to notice that female actors were expressing happiness more freely, while male actors were more reserved. Given the fact that happiness was properly recognized overall, we believe that this difference in the way happiness was portrayed by the actors might have made gender easier to identify.

In our study, we were particularly interested in the differences between walking and conversational motions. We argued that motions which contain less gender cues would be more affected by gender stereotypes that exist around certain emotions. Firstly, we noticed a difference in gender judgments for neutral conditions between both types of motion which confirmed our belief – gender was less apparent in conversational than walking motions. We assumed that this was due to the presence of physiological gender cues in walking as opposed to less obvious or absent gender cues in conversation. Furthermore, in certain emotional conditions, gender bias appeared to be more prominent for conversing characters: male sadness was rated as significantly less male in conversation than in walking condition and female anger was perceived to be less female in conversation as opposed to angry walk. However, the gender was not perceived as coming from the opposite sex, which means there might still be some gender information present in the conversational motion that would prevent stereotypes from completely influencing the judgment. This result is different to the result of Johnson et al. (2011) where they found that angry throwing motions resulted in the perception of actor's sex being perceived as male, even when female actors were portraying it. Based on our assumptions, throwing motion may not contain as many gender cues as conversations, therefore gender judgments in their study could have relied mostly on the portrayed emotion. In the case of happiness and fear, we expected male actors to appear less male in the conversational motion as opposed to the walking motion. However, we found no difference in the effect of gender bias for these two motion types.

The actual appearance of the model also had a small effect on the perception of gender. Male participants perceived female walking as more female on the Man model and female participants perceived male walking more male on the Woman model. These results show a slight contrast effect, as proposed by Hess et al. (2004). Furthermore, the role of participant sex in assessing the gender of the opposite sex could be explained with mating strategies proposed by evolutionary psychology (for an overview see Hugill et al. (2010)). Since reproduction and selectivity of partners is an important task, humans have a high sensitivity for assessing non-verbal cues from the people of the sex they are attracted to. It could be that male participants in our study were more sensitive towards female motion which became even more apparent on male models and female participants spotted male motion on a female model more easily as well.

Confusion with other emotions gave us information on how well the emotions were perceived by participants. All emotions were recognized above chance, confirming that the variation in our gender judgment tasks were due to the target emotions. Even though all emotions were recognized above chance, and as known from the literature (Brody & Hall, 2000), females are better at portraying emotions which might make the emotion categorization task easier. Even though emotions expressed by male actors were more often confused with each other, they were considerably well recognized in our study.

We also found that emotions were more difficult to recognize on one but not the other type of motion. Male anger and happiness were more difficult to recognize on walking motions, where in conversation they were the most recognized emotions. This could mean that some types of motions are better for emotion portrayal or that there are certain areas of the body which are more important for the expression of certain emotions, e.g. happiness is more difficult to identify if there is no facial information present. This is not an unusual result since previous studies have found different recognition rates for emotion from body, facial and combined expressions (Clavel et al., 2009; Aviezer et al., 2012; Ennis et al., 2013). Particularly interesting is the study of Aviezer (2012) where highly intense emotional expressions that were taken from real-life situations were correctly identified from body motion but not from isolated facial expressions. It appears therefore that intensity of emotion plays a part in recognition from certain parts of body.

Male fear appeared to be the only emotion which was difficult to recognize in both types of motions. This could mean that male actors were not as successful at portraying fear as female actors. This could be an indication of a cultural effect on the production of the emotion. For example, males are not expected to portray fear therefore their portrayal of this emotion was harder to categorize by the participants. Furthermore, as mentioned above, gender ratings showed more difficulties in perceiving the underlying gender when male actors were portraying fear. Both results could point to the cultural effect on the production and perception of emotion.

Even though we cannot explain how appearance affected gender ratings, we found that the choice of model could improve emotion recognition. Some emotions (mainly fear and sadness) were easier to recognize on the Man model.

3.1.5 General Discussion

In this study, we introduced some new features to the understanding of how certain basic emotions can influence the perception of gender of virtual characters. We proposed that: a) stereotypes for emotion will bias gender perception, b) this effect will be stronger for motions that do not exhibit obvious physiological cues, and that c) there would be an effect of virtual model on gender perception. Our results generally confirmed all three hypothesis. The way we perceive gender from motion does depend on whether specific emotion is expressed, how this emotion is expressed and who expresses it. Some emotions are attributed more to one gender than the other and therefore can change our perception of gender, especially if the type of motion itself does not reveal the underlying gender sufficiently. In our case, anger and male sadness were the emotions that biased gender perception the most, especially when expressed through conversing, rather than walking characters. Furthermore, the appearance of the model can also affect the perception of gender, particularly by making the underlying sex of the actor more apparent when it is mismatched with the sex of the virtual model.

However, in our study emotions did not invert the perception of gender. It is possible therefore, that attribution of emotion to a specific gender does not influence gender perception to a great extent when there are sufficient gender cues present in the motion already. A good rule to follow when creating virtual humans using natural human motion would thus be to match actor's sex with that of the virtual character. But in cases or types of motions where gender is not apparent, some emotions can influence the perception of the actor's actual sex. For example, if we create a male virtual character, angry motion would make him look more male, while fearful and sad motion would make him appear less male.

Our result is also arguable from the point that we only found this effect for anger and male sadness but not other emotions. This could be due to the fact that social stereotypes do not influence perception of emotion but its production as well, as already suggested. For example, showing fear is seen as a sign of weakness, therefore males would be more reserved while expressing it. This reservation in expression and not the emotion itself might make gender easier to identify by the observers. This would not show the effect of biased gender perception, but the very portrayal of emotion by an actor would be under the cultural influence on the production of the emotion. We believe that some emotion recognition rates in our experiment might have been lower because they were portrayed with more reserve as well. This possibility also prevented us to do gender judgment analysis only on best sets of recognized emotions, since that would artificially skew the data. A similar systematical bias on emotion recognition is presented in the paper from Johnson et al. (2011), where neutral, happy and sad throwing motions were often confused with each other but angry was not. Overall, emotions in our experiment were recognized above chance so we believe that the gender analysis per emotion was conducted as appropriately as possible given the intrinsic complexity of studying simultaneously emotion and gender.

Since we found a slight effect of the virtual character's appearance on the perception of gender, using the appropriate sex of the actor to match the appearance of the character can avoid any confusion. However, this might not be true for all social subgroups. E.g. transsexuals, are shown to exhibit motion specifics of the sex they identify themselves with.

We are also aware of some limitations of our study. Due to technical difficulties, we could not capture finger or eye motion for the conversing characters, which may improve the believability of the motion, since the missing data can be detected from the observer (Jörg et al., 2010; Hodgins et al., 2010). Due to the limitations of our motion capture system, we were also unable to record facial information for the walking motions which might have improved emotional recognition and affect gender ratings. We are aware that in the cross-analysis we compared stimuli that are slightly different and that statistical differences could be due either to the lack of facial motion or differences in motion types. However, given the similar trends in the results of the cross-analysis, such analysis already provides us insights about how emotions differ between different types of motions. Another limitation is that we only used two models to display the motion. Furthermore, a gender rating study was used in our experiment design, where we measured judgments of gender, which could differ from actual perception of gender. Other measures (e.g. eye-tracking) could give additional information about how we perceive gender of virtual humans. However, we are confident that the conducted research furthered our understanding about the effect of emotion on gender perception.

With this study we were able to show how complex the interaction between different structural components of the virtual character actually is. The mismatch in appearance and motion can make the identification of underlying gender easier (contrast effect), but it is also important *what* the character is expressing and *how*. Results also show evidence of social cognition influencing the perception of virtual characters: people's stereotypes could effect the overall perception of the character and, as hypothesised, this is mediated by how difficult this task is (how many gender cues are available).

3.2 Shape and Material Mismatch

While the previous study investigated the effect of motion and appearance mismatch, this study explores another type of mismatch related to static components of appearance: shape and material realism. We use the term "shape" to describe the properties of the character's polygon mesh, whereas "materials" encompass texture and other optical elements that give the shape its overall appearance.

I collaborated as fourth author on this study which was accepted to SIGGRAPH Asia 2015. The other two research groups were graphics group at Bielefield University, Germany and University of Zaragoza, Spain. My main contribution was the formation of the experiment design, analysing the results and putting importance on perception of emotion of the virtual character. However, this study provided interesting findings which are important for our research focus, since two main aspects that primarily define the appearance of a 3D character were manipulated.

The main aim of this study is to analyze how different combinations and mismatches of shape and materials affect the perceived realism, appeal, eeriness, and familiarity of the characters. Our experiment design is inspired and justified by the current trends in feature animation, which have recently used different combinations of stylised shapes and materials to depict 3D characters. Examples include highly stylised shapes and textures in Pixar's *Toy Story* (1995) movies, or the somewhat less stylised shapes but photo-realistic materials in *The Adventures of Tintin* (2011). Furthermore, in a review of recent advances in facial appearance capture, Klehm and colleagues (2015) mention the need for deeper insights into human perception of facial appearance. They note the complexity and the importance of focusing on important features, which we address by carefully isolating the effects of the parameters being studied in each test. We use static pictures as stimuli, as it has been found that much of the information that people use to evaluate virtual characters is available in a still image (McDonnell et al., 2012).

Mismatch in elements of realism has been found to produce negative effects (Seyama & Nagayama, 2007; Burleigh et al., 2013) and Green et al. (2008) concluded that there is less tolerance to deviations from original proportions in the cases where faces are more attractive and human-like. A previous paper (MacDorman et al., 2009) analysed combinations of mesh and texture realism of the virtual character, however mesh realism was manipulated only by

Figure 3.12: Our face scanning setup (right) and comparison between photographs and virtual reconstructions of our actor (left).

polygon count (the realistic mesh had the highest number of polygons) while our study uses facial shapes which range from scanned 3D representation of a human face to very abstract cartoon characters, which were handcrafted by artists.

Several studies confirmed that texture changes do result in a significantly more attractive face (Benson & Perrett, 1992; Little & Hancock, 2002), especially the effect of blurring the skin texture (Fink & Matts, 2008). Therefore we apply texture blur to our stimuli and test the effect on the perception.

Moreover, we additionally investigate how the combinations affect the perceived intensity of different facial expressions (sadness, anger, happiness, and surprise). Stylisation has been found to enhance the recognition and intensity of expression of a virtual character, particularly in the case of cartoons (McCloud, 1993) or illustrations (Gooch et al., 2004), but also when presented as a dynamic stimuli which were not heavily stylised (Wallraven et al., 2007). We wanted to see how the emotion expressions will be affected by our combinations as well.

3.2.1 Stimuli Creation

Our experiments required the design of different levels of stylization of the same character. Additionally, for each stylization level we modeled four of the universal facial emotions: anger, happiness, sadness, and surprise (Ekman, 1992), plus a neutral expression. We discarded disgust and fear because their status as basic expression was questioned recently (Jack et al., 2014) and they are harder to identify by observers.

The creation of the stimuli for this experiments was a demanding task and was a result of collaboration between research groups from three universities (Trinity College Dublin, Bielefield University and University of Zaragoza) and two professional artists.

Our realistic characters are based on real people of about average attractiveness without ethnic bias to the group of participants. To generate the realistic models, the graphics group at Bielefield University replicated the multiview-stereo face scanner of Beeler et al. (2010), which reconstructs high-resolution textured point clouds from the photographs of six cameras arranged as pairs around a person (Figure 3.12). Since all photographs are taken simultaneously, the scanning process is instantaneous and therefore well suited for capturing different facial expressions. Each pose representing one emotion was captured several times, and the most convincing one was selected by a group of about twenty people of different cultural backgrounds, while referring back to Ekman's guidelines.

Since the scanner only captures the frontal part of the face and fails to faithfully reconstruct eyes and hair, the template head model was fitted to the measured point cloud using a non-rigid registration approach similar to Weise et al. (2011). Regions of missing data are therefore filled in by the template model, which additionally provides a 2D parameterisation of the model. This parameterisation is used for texture mapping, with texture images being generated automatically from the photographs. The hair style, the eyes, and the teeth were manually sculpted and adjusted to fit the scanned model. Figure 3.12 shows one example of our reconstructed models.

While a realistic character can be obtained from 3D scans of a real person, no automatic solution exists to generate increasingly stylised versions. Therefore, professional 3D artists produced the required stylised shapes and materials from our realistic characters, taking inspiration from commercial animation films (see examples in Figure 3.13). For our first set of experiments we used three stylization levels for shape and material. The extended stimuli for the later experiments used two more stylization levels.

We were interested in analyzing the effect and interaction of shape, material, and textures. Therefore, we transferred all material properties of the baseline characters to the other character shapes. The inter–surface mapping for the texture transfer was computed based on a dense correspondence map established using the non–rigid registration technique of Zell and Botsch (2013).

Rendering of all stimuli was performed using Mental Ray, with each character being placed in front of a light gray background. The lighting setup consists of a key light and a rim light, and photon mapping is used for global illumination. For the skin all three characters use the same multi–layer skin shader with subsurface scattering, with diffuse albedo specified by a

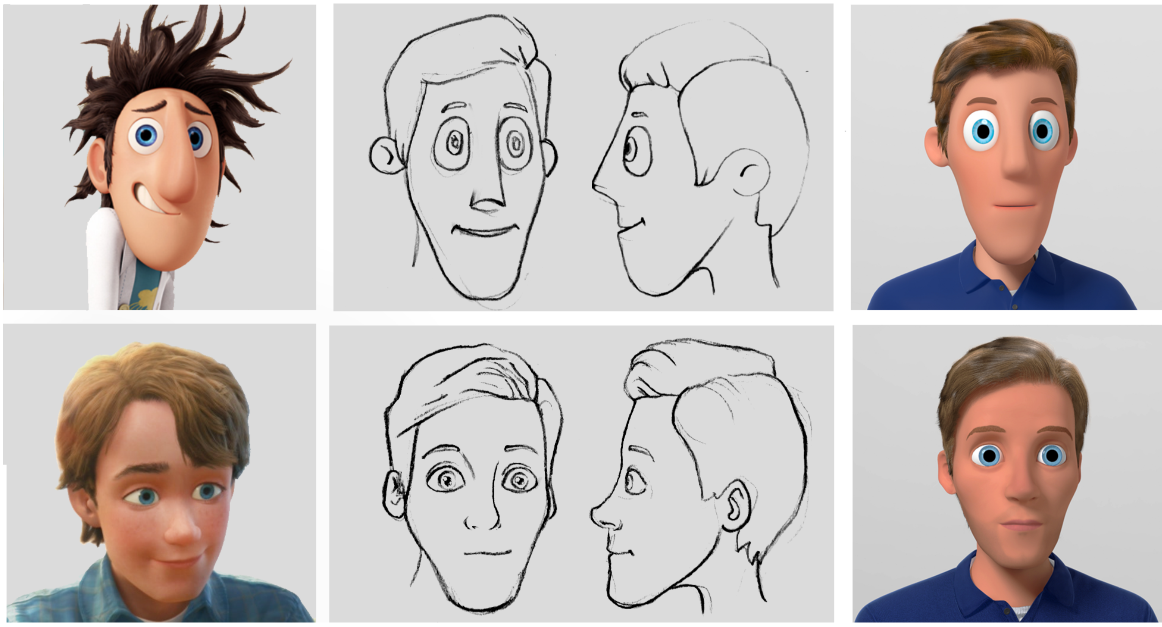


Figure 3.13: Two of the stylizations created for the study, showing the sketches provided to the artists on the left and their resulting stylised 3D models on the right. The designs are inspired by the films *Cloudy with a Chance of Meatballs* (2009) (top) and *Toy Story* (1995) (bottom).

high-resolution texture map. The shader parameters vary between the models in order to closely resemble the targeted render styles.

3.2.2 Experiment Design

The appearance of virtual humans is a function defined over a huge multi–dimensional space. While it is generally recognized that *shape* and *material* are the main contributors to the overall appearance of virtual characters, these two might be affected by several sub–dimensions. For example, material is the combination of shader, shader parameters, and textures, each of which having a potentially different influence on appearance. This makes the experiment design an extremely difficult task, given the large number of variables to explore.

Similar to previous work on rendering style (McDonnell et al., 2012; Ho & MacDorman, 2010), we want to analyze how different levels of stylization (e.g., shape and material) change the perception of a virtual character and we also employ similar scales for our experiments. The descriptions below are the ones given to the participants of the perception studies:

- *Extremely unappealing—Extremely appealing*: High appeal means that the character is one that is pleasant and you would like to watch more of. Unappealing means that you dislike to watch the character.
- *Extremely eerie—Extremely re-assuring*: Indicate if you find the character eerie, which means that they are gloomy and leave you with a sense of fear. Re-assuring means that the character restores a sense of security, confidence, calm in you.

- *Extremely abstract—Extremely realistic*: Indicate if you find the character's appearance to be highly stylised like in cartoons, or close to photo-realistic as in real pictures.
- *Extremely unfamiliar—Extremely familiar*: Indicate if you find the character's appearance familiar to you, in that you have seen something similar to it before, or if you find the character unfamiliar with an appearance that you haven't seen anything like before.
- *Extremely unattractive—Extremely attractive*: Indicate whether you find the character unattractive and ugly or beautiful and attractive.

We chose a seven-point scale in order to give participants more response options and to allow for comparison to previous studies. The Likert scales were numbered 1–7, with a description provided on both ends of the scale.

Since both the design and the analysis of our experiments share many similarities, we describe the general setup now and later only mention deviations. The user's task and the rating scales were explained on a written document to the participants before the experiment. Afterward, all stimuli were presented in a random order and shown for 3 seconds each. The display was calibrated, 20" wide and at about $50cm$ distance from the participants. The renderings have a resolution of 1024×768, corresponding to approximately $26.5cm \times 20.0cm$ on screen. After each stimulus presentation, participants were asked to rate it according to the above scales. In all experiments, the participants had normal or corrected-to-normal vision and were unaware of the final goal of the experiment. They were asked to report their 3D experience (how often they played video games, watched movies with visual effects, and how they would consider their knowledge of 3D graphics). We did not find any correlation between the reported 3D experience and the results of our tests, and thus omit this information for the rest of the paper.

All experiments reported in this study were conducted by the graphics group at Bielefield University and the participants were recruited mainly from this university as well.

3.2.3 Experiment 1a: Material and Shape

We first investigate the influence of shape and material, where we denote by material the combination of shader, shader parameters, and textures. The combination of each material with each shape style leads to a total of nine different versions of the character, times five different expressions, resulting in a set of 45 stimuli. Figure 3.14 shows a subset of the stimuli for the happy expression. We analyze the interaction between shape and material for the scales most frequently used in previous work: realism, appeal, reassurance, and familiarity. Twenty-two volunteers participated in this first experiment: 14 female, 8 male, with age from

19 to 30 years (average age: 24.5). Participants were asked to read and sign the informed consent and read through the information sheet, where they were familiarized with the details of the experiment.

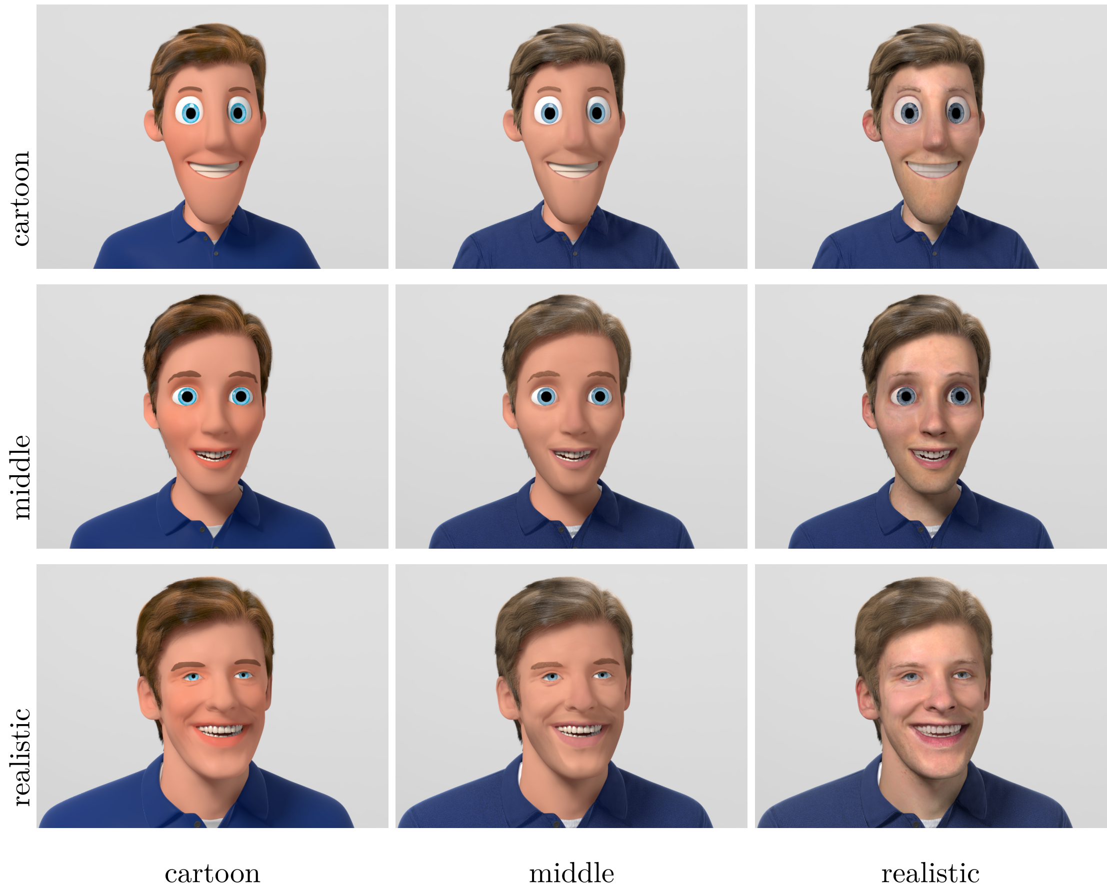


Figure 3.14: Stimuli used in Experiment 1: Material and Shape. Three levels of shape (*vertical axis*) and material (*horizontal axis*) stylization, shown here for the *happy* expression. The baseline stimuli are shown on the diagonal. Their textures have been transferred to the other shapes for producing the off-diagonal stimuli of mismatching stylization levels for shape and material.

Results

For statistical analysis of each rating scale we conducted an n-way repeated measures Analysis Of Variance (rm-ANOVA). We run Maulchly's test for validating sphericity of the data, and whenever it is significant we report results with Greenhouse-Geisser correction applied and marked with an asterisk (*). Whenever main interaction effects were found, we conducted a Tukey Honestly Significant Difference (HSD) test, suggested for multiple comparisons in an experimental design (Cunningham & Wallraven, 2011), for the comparison of means to further explore the results.

In this section, we analyze the effects of shape and material only. The results are presented on Figure 3.15, which shows the ratings averaged over all expressions. Despite a smaller

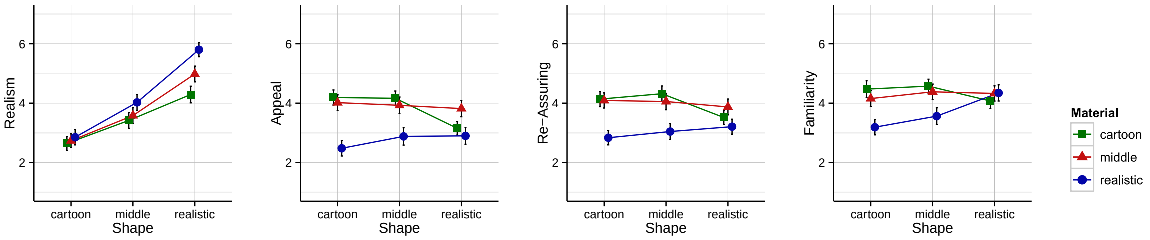


Figure 3.15: Results of Experiment 1: Material and Shape. Ratings for perceived Realism, Appeal, Reassurance/Eeriness, and Familiarity, for different Shape and Material stylisations. Error-bars denote 95% confidence levels.

offset and some noise, ratings for different expressions have been very consistent, which justifies averaging over all expressions as opposed to conducting the analysis on neutral expressions only. For statistical analysis, a rm-ANOVA with three factors (*Shape, Material,* and *Expression*) was used.

a) Realism

A main effect was found for Shape ($F(2, 42) = 113.18$, $p \approx 0$) and Material ($F^*(1.47, 30.82) = 23.15$, $p \approx 0$, $\epsilon = 0.73$), as well as for the interaction between Shape and Material ($F(4, 84) = 11.14$, $p \approx 0$). Post–hoc tests show that the cartoon shape was perceived as least realistic, no matter which material was used. Similarly, cartoon and middle materials did not make a difference for the middle shape, while the realistic material caused a more realistic perception for this shape ($p < 0.002$ for both comparisons). In contrast, all material levels differ significantly for the realistic shape ($p < 0.001$). These results indicate that *shape is the main contributor to the perceived realism.*

b) Appeal

We found a main effect of Material on the ratings of Appeal ($F^*(1.41, 29.67) = 42.69$, $p \approx 0$,$\epsilon = 0.71$), but no main effect of Shape was found. An interaction between Shape and Material ($F(4, 84) = 13.97$, $p \approx 0$) shows that a realistic material on a cartoon shape yields the least appealing combination, since a post–hoc analysis showed significantly lower ratings for this combination compared to all others ($p < 0.02$ in all cases). The realistic material is less favored on the middle shape as well, and the cartoon material on the realistic shape is similarly unappealing ($p < 0.02$ in all cases except the combinations mentioned above). The middle shape was rated as equally appealing regardless of material. These results suggest that *material contributes most to the perceived appeal of a CG character, and strong* *mismatches in the level of stylization of shape and material can result in very unappealing characters.*

c) Reassurance

Similar to the appeal ratings, we found a main effect of Material on the ratings of Reassurance ($F^*(1.51, 31.70) = 49.07$, $p \approx 0$, $\epsilon = 0.76$), but no main effect was found for Shape. An interaction between Shape and Material is present ($F(4, 84) = 12.02$, $p \approx 0$) and post–hoc analysis showed significantly lower ratings of Reassurance especially in shape-material combinations that reduce appeal the most: realistic materials on all shape levels and cartoon materials on the realistic shape ($p < 0.02$). The realistic material on the cartoon and middle shape was perceived most eerie. A Cronbach's alpha value of $\alpha = 0.88$ confirms high similarity between the Appeal and the Reassurance scale. We found that *participants rated a character on the reassurance scale similarly as they rated him on appeal.*

d) Familiarity

Again, a main effect has been found for Material ($F(2, 42) = 12.58$, $p \approx 0$), but not for Shape. Furthermore, there is also a significant interaction between Shape and Material ($F(4, 84) = 17.99$, $p \approx 0$). The results of the post–hoc test for Familiarity are less similar than between the Appeal and Reassurance ratings. Even though the combination of realistic material and realistic shape is unappealing and eerie, it was not rated significantly less familiar than other combinations. Realistic materials on cartoon and middle shapes result in the least familiar combinations ($p < 0.02$ in all cases). We conclude *the perceived familiarity of the character is independent from its perceived appeal or reassurance.*

3.2.4 Experiment 1b: Texture

One possible explanation of why the middle material was rated the most appealing for the realistic shape could be the reduced pigmentation variation as reported by Fink and Matts (2008), as discussed before. In order to analyze whether their findings on attractiveness can also explain our effects on appeal and reassurance, we designed a variation of the previous experiment. This experiment tests:

- whether it is possible to influence appeal or realism by changing only the albedo texture,
- if there is a possible correlation between attractiveness and appeal/reassurance, and

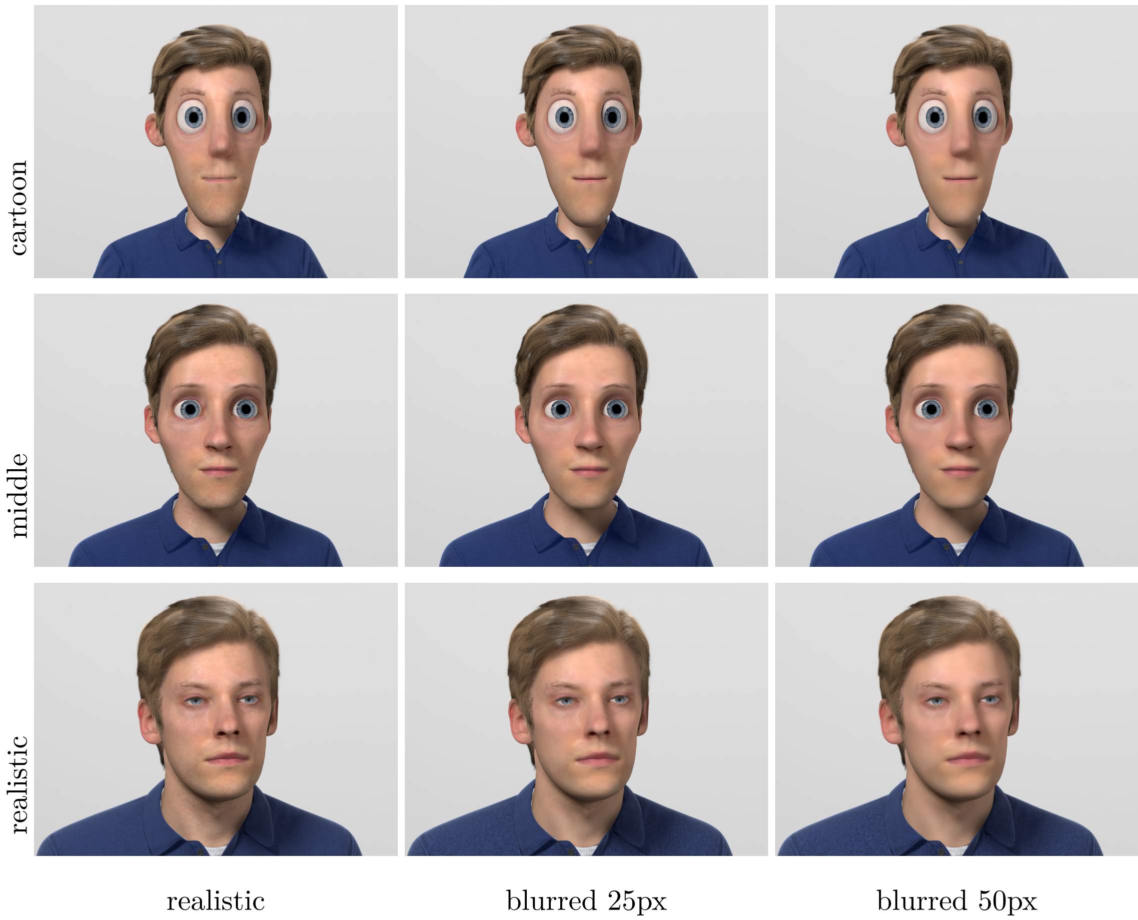


Figure 3.16: Stimuli for Experiment 1b: Texture. Realistic material with realistic texture and two variants with blurred textures (Gaussian kernels of 25 and 50 pixels, *horizontal axis*), for the three shape stylisations (*vertical axis*).

- whether appeal can be increased without sacrificing realism too much, simply by filtering a photo-realistic texture.

We created two additional textures with reduced skin details by applying uniform Gaussian blur of kernel sizes 25 and 50 pixels (for 4k textures), respectively. The 50px kernel covers barely 1 cm of the face, which translates into around four pixels in image-space. Lips and skin were filtered independently in order not to blur the boundary in between; eyebrows were not filtered. These three textures (realistic, blurred 25px, blurred 50px) were used in combination with the realistic material. To enable a comparison with the previous experiment, we also included the cartoon and middle materials (with their original textures only). This results in a set of 5 materials, which were also transferred to the middle and cartoon shapes, as shown in Figure 3.16.

For this experiment we tested these 5 materials on the 3 shape stylizations, but used the neutral expression only, leading to 15 stimuli in total. Note that the three realistic materials differ in their (blurred) texture only. The presentation of the stimuli was repeated three times with different random orderings. After each stimulus, participants were asked to rate it according to the previously described scales for appeal, reassurance, and realism, plus a new scale *attractiveness.*

Twenty–one new volunteers (13 female, 8 male; average age: 24.6) participated in the experiment. As before, participants were recruited from Bielefield University and were asked to read and sign the informed consent and read through the information sheet, where they were familiarized with the details of the experiment.

Results

For statistical analysis, a rm-ANOVA with three factors (*Shape, Material,* and *Expression*) was used. All results from the previous experiment were confirmed, and thus we only describe the main effects related to the added material levels.

a) Realism

Although a main effect was found for Shape ($F^*(1.29, 25.78) = 124.98$, $p \approx 0$, $\epsilon = 0.65$), Material ($F(4, 80) = 17.52$, $p \approx 0$) and an interaction between Shape and Material ($F(12, 240) = 6.42$, $p \approx 0$), the post–hoc shows that this is not related to the added textures. The ratings for the two blurred textures are between the realistic and the middle texture, but are not significantly different for any shape. This confirms our initial assumption that *blurring a realistic texture only slightly reduces the perceived realism of a character.*

b) Appeal and Attractiveness

Due to the high similarity between Appeal and Attractiveness (Cronbach's $\alpha = 0.87$) we report these results together. A main effect was found for Shape for Attractiveness ($F^*(1.33, 25.54) = 5.36$, $p = 0.021$, $\epsilon = 0.66$) but not for Appeal. Material was significant in both cases (Appeal: $F^*(1.68, 33.60) = 27.17$, $p \approx 0$, $\epsilon = 0.42$; Attractiveness: $F^*(1.56, 31.26) = 16.72$, $p \approx 0$, $\epsilon = 0.39$). The interaction between Shape and Material was significant (Appeal: $F^*(7.05, 94.03) = 4.99$, $p \approx 0$, $\epsilon = 0.59$; Attractiveness: $F(12, 240) = 2.88$, $p = 0.005$). As we hypothesized, the blurred textures were rated higher than the realistic texture. This effect is stronger for the cartoon and middle shapes and a significant difference between the realistic and 50px blurred version was found ($p < 0.003$ in all cases). For other comparisons between the blurred and realistic textures no significant difference was found. However, the graphs in Figure 3.17 show that the two blurred textures were rated equally appealing for the realistic shape. In contrast, a stronger blur is preferred for cartoon and middle shapes. Although

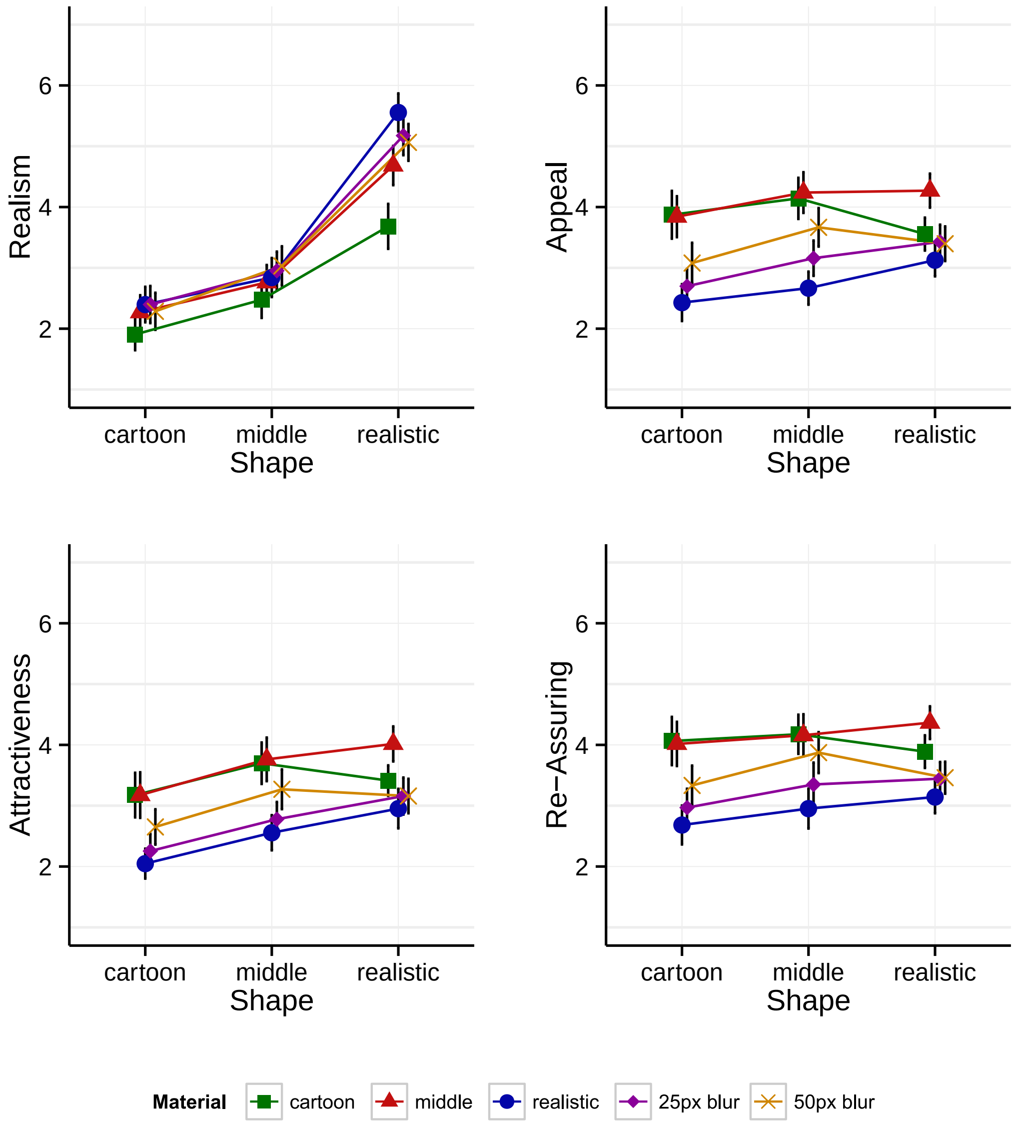


Figure 3.17: Results of Experiment 1b: Texture. While there is nearly no difference between the realistic and blurred textures for the realism scale, the blurred textures increase appeal and attractiveness and reduce eeriness.

the results of our tests are not significant in some cases, these findings are in line with research of Fink and Matts (2008). We generalize their findings to character shapes of different stylization levels. We therefore conclude that *blurring realistic skin textures is a reasonable approach for increasing appeal or attractiveness, without losing too much realism.*

c) Reassurance

Although the graphs of Reassurance and Appeal are similar (Figure 3.17; $\alpha = 0.89$), a main effect was found for Material only ($F^*(1.44, 28.72) = 24.55$, $p <\approx 0$,$\epsilon = 0.36$), but not for Shape. In addition, there is an interaction between Shape and Material ($F^*(7.128, 142.46) = 2.66$, $p = 0.029$, $\epsilon = 0.59$). The two blurred textures have been rated less eerie than the realistic version. Significant differences have been found between the realistic texture and the 50px blurred version for cartoon and middle shapes ($p < 0.001$). Thus, *blurring a texture does not only increase appeal, but also reduce eeriness.*

3.2.5 Conclusion

The two experiments described above allow us to draw the following main conclusions on the tested dimensions:

- Shape is the main descriptor for realism, while material is more important for perceived appeal, reassurance, and attractiveness. Strong mismatches in stylization between material and shape affect negatively the appeal and attractiveness of the characters and make them more eerie.
- Texture has stronger influence on appeal and attractiveness. Blurring a realistic texture does not significantly reduce realism but increases appeal and attractiveness.
- Ratings for appeal, reassurance, and attractiveness measure similar concepts ($\alpha > 0.87$ in all experiments), but do not correlate with the realism scale ($\alpha < 0.5$ in all experiments).

3.2.6 Experiment 2a: Material and Shape Extended

Experiment 1 indicated that different stylization levels of material and shape have a big impact on perceived appeal or realism. However, our set of stimuli contained only a single character, and the realism scale was not densely sampled. A more stylised character might reveal that big mismatches between material and shape cause unappealing results, or a stylization level between middle and realistic might cause uncanny reactions. To allow for a more generalized conclusion about different stylization levels, further investigation is required.

In the following experiment we analyze the effect of varying stylisations on shape and material, including matching and mismatching levels of stylization, on a significantly extended set of stimuli. In particular, we seek answers to the following questions:

- Can our findings be observed on other characters as well?
- Does a strong mismatch between material and shape create unappealing results only for realistic shapes or for all shapes?

Stimuli

We extended our initial stimuli with another character of different gender, because this adds by design a clearly distinctive person. For each character, two additional stylisations were

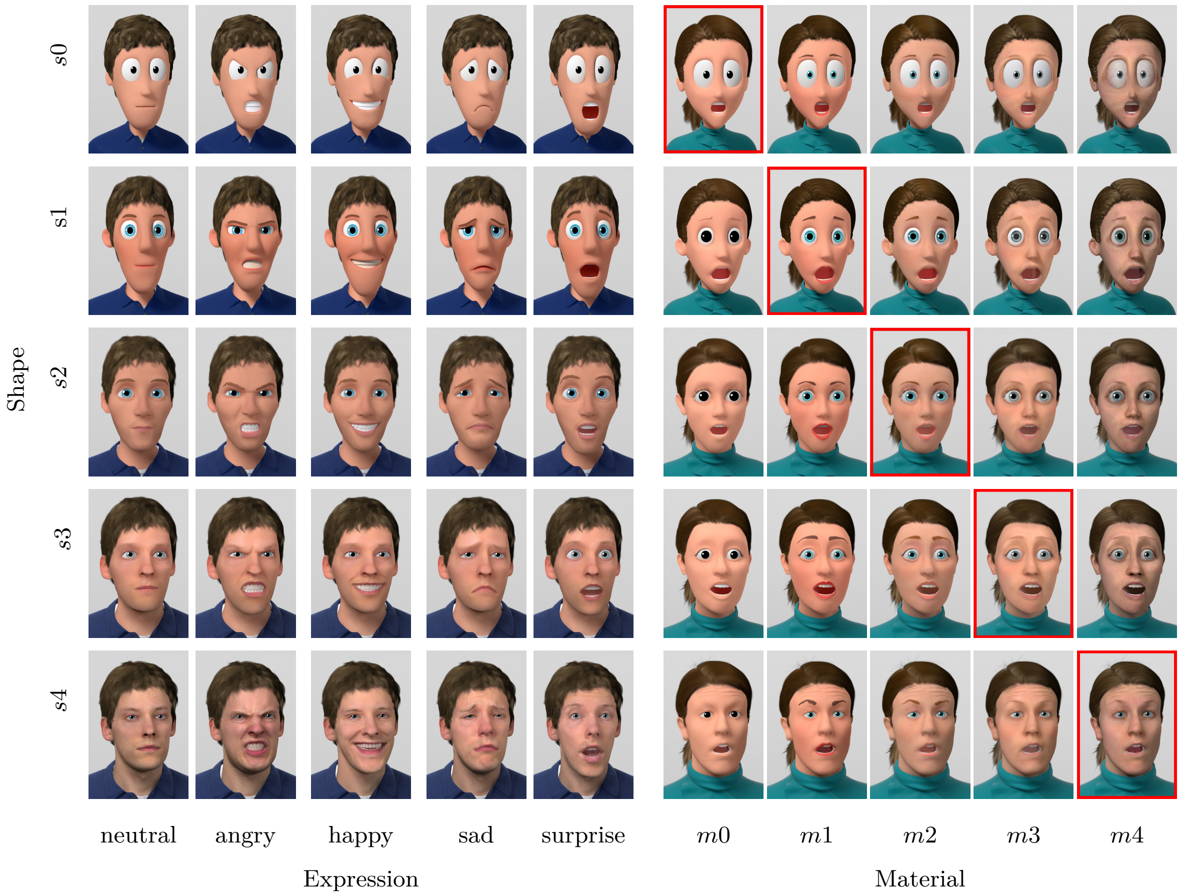


Figure 3.18: Stimuli for Experiment 2: *Left:* Renderings of the male character for different stylizations (rows) and basic emotions (columns). *Right:* Combinations of shape and material stylization for the female character (surprise expression), with baseline stimuli on the diagonal.

created, yielding five stylization levels from *level 0* (most stylised) to *level 4* (highly realistic). We distinguish between stylisations in material and shape by using the prefix m and s respectively. The new stylisations (level 0 and level 3) have been particularly designed by the artists to fill the gaps for perceived realism in the stylization scale. For these levels, our character designs are inspired by *Pocoyo* (2004) and *Tangled* (2010). We also changed the hairstyle of the virtual male character in order to allow a better comparison with a photograph of the actor. This provides us with baseline ratings on appeal and realism for the real person. The new set of stimuli is composed of 2 characters $\times$ 5 shape stylisations times 5 material levels $\times$ 5 expressions, leading to a total of 250 images. A representative subset of the stimuli is shown in Figure 3.18, for the five expressions and matching shape/material levels of the male character (left), and the 25 combinations of material and shape for the female character (right).

Procedure

The largely extended stimuli require a reduction of the scales in order to keep the experiment tractable. Given that the appeal, reassurance, and attractiveness scales measure similar concepts, and that the familiarity scale did not provide much information, we decided to keep only the realism and appeal scales for this experiment. Furthermore, we increased the display time of the stimuli to $4s$, and showed the neutral male and female baseline characters before the experiment, such that participants could better estimate the range of characters from the beginning on. At the end of the actual experiment, participants rated a photograph of the real characters in neutral expression. The rest of the experiment remains similar to the previous one. With all these changes, participants finished the experiment within 50 minutes or less. Twenty-one new different volunteers, recruited from Bielefield University took part (17 female, 4 male; average age: 23.4).

Results

Our results are summarized in Figure 3.19 and are mostly consistent across male and female. Repeated measures ANOVA with four factors (*Character, Shape, Material*, and *Expression*) was used for statistical analysis. Differences between the two characters were significant, but since they were rather small and/or inconsistent, we exclude them from further analysis. In the following we present an in–depth discussion of the realism and appeal ratings, and report the impact of emotion in Section 3.2.7.

a) Realism

ANOVA revealed a main effect of Shape ($F(1.98, 39.6) = 178.67$, $p \approx 0$) and Material ($F(1.33, 26.4) = 73.92$, $p \approx 0$) as well as an interaction between Shape and Material ($F(6.17, 134.1) = 11.59$, $p \approx 0$). Post–hoc analysis shows that all shapes ($p < 0.004$) and most of the materials ($p < 0.003$ except for level $m0$ and $m1$) differ significantly from each other. The 25 groups resulting from the combinations of Shape and Material differ also significantly in more than 80% of the cases. Most non-significant comparisons can be found for the Shape level $s0$. For example, increasing the Material from level $m1$ to $m2$ or from level $m2$ to $m3$ does not cause a significant difference. This contrasts with the case of the realistic Shape levels $s3$ and $s4$ ($p < 0.002$). This confirms that *as the shape becomes more realistic, the material stylisation becomes more dominant for perceived realism.*

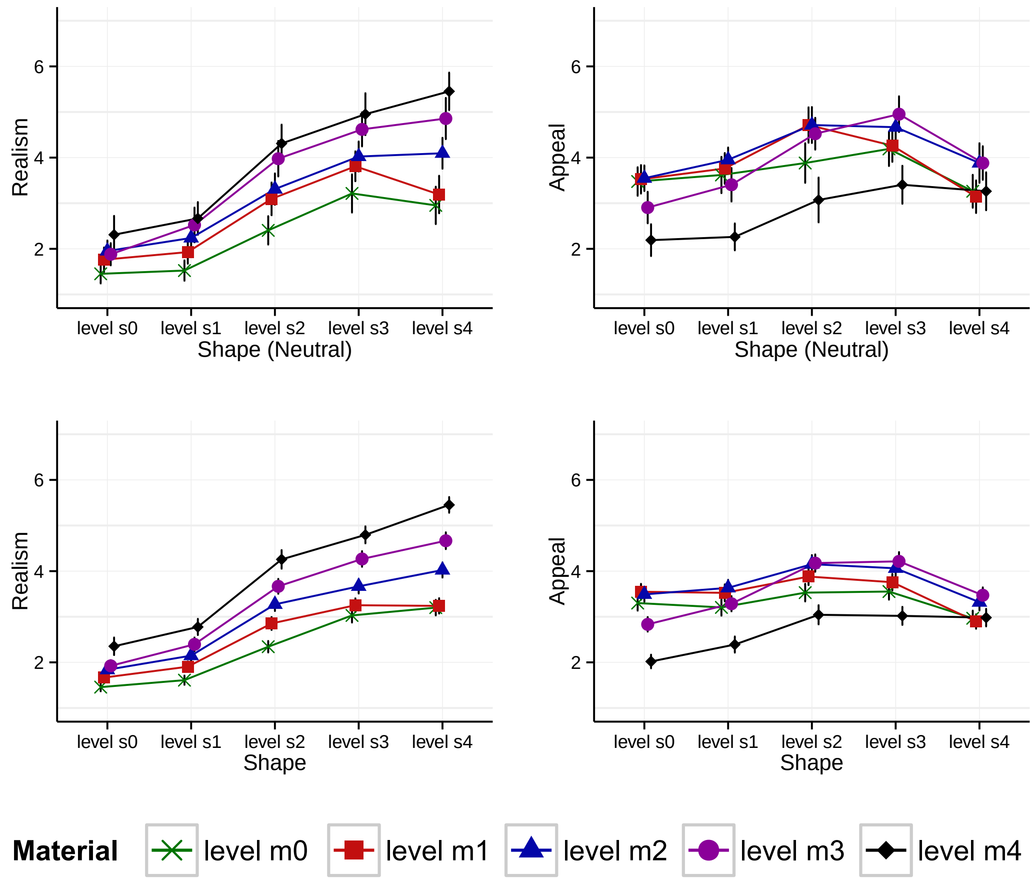


Figure 3.19: Results of Experiment 2a: Ratings for perceived realism and appeal for different shape and material stylisations. *Top row*: neutral expression averaged over male and female characters. *Bottom row*: averaged over all expressions and characters.

b) Appeal

The main effects of Shape ($F(2.58, 51.6) = 20.97$, $p \approx 0$) and Material ($F(1.88, 37.6) = 20.39$, $p \approx 0$) are comparable. There is a slightly weaker interaction between Shape and Material ($F(6.06, 121.3) = 14.29$, $p \approx 0$). Post–hoc analysis reveals that Shape levels $s2$ and $s3$ were perceived more appealing than the other Shape levels ($p < 0.001$ in all cases between the two groups).

For Material, only the most realistic version (level $m4$) was significantly less appealing than all other materials ($p < 0.001$). This supports our previous assumptions that smooth(ed) skin pigmentations are perceived as more appealing. For the abstract Shape $s0$, Material levels $m0$, $m1$ and $m2$ form a cluster without any significant difference; this cluster is found significantly more appealing than Material levels $m3$ and $m4$ ($p < 0.03$). On the other hand, Shape level $s3$ is rated significantly higher with matching Material levels ($m2$ and $m3$), with both more stylised ($m0$ and $m1$) and more realistic ($m4$) materials being rated significantly lower. These results support that *in all cases a strong mismatch between shape and material is perceived as unappealing.*

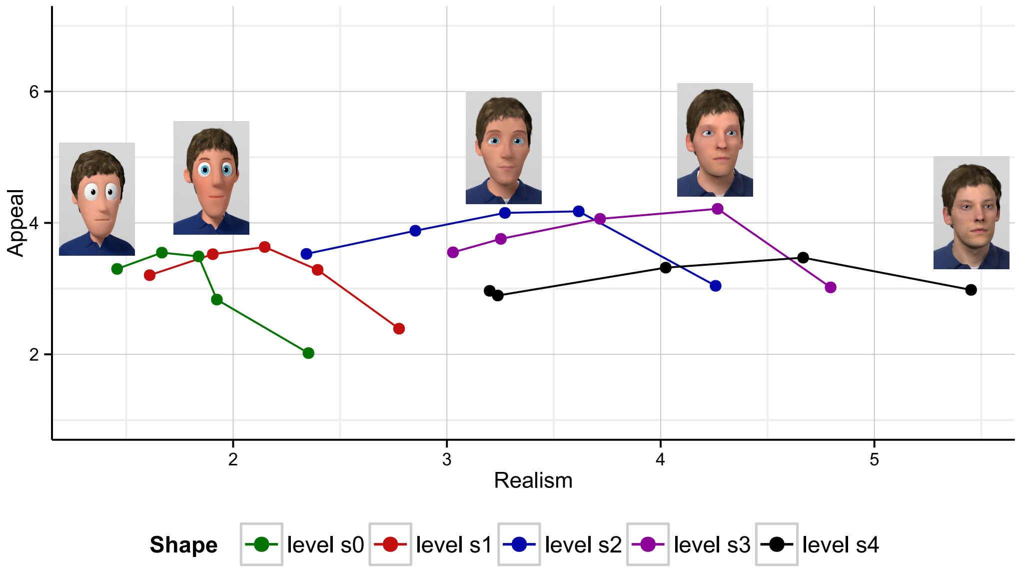


Figure 3.20: Participant ratings for our male stimuli plotted on a Realism-vs-Appeal diagram, similar to (MacDorman et al., 2009). Each graph corresponds to one Shape stylisation, while graph nodes correspond to Material levels. The icons are placed above the nodes of matching Shape/Material levels. The diagram reveals that perceived Realism is a bad predictor for Appeal. Instead, it is the compatible degree of realism of both shape and material that matters.

c) Photograph

At the end of the experiment, participants rated a photograph of the real actor in neutral pose. A statistical comparison could not be performed for this data, so we report the averages for observation purposes. As expected, the average Realism rating is very high ($\bar{x} = 6.98$, $SD = 0.15$). The average Appeal rating was $\bar{x} = 4.5$ ($SD = 1.40$), which is higher than the average ratings for the realistic *s4/m4* characters ($\bar{x} = 3.26$, $SD = 1.33$). This dip in Appeal rating for the *s4/m4* character is in agreement with the uncanny valley theory (Mori, 1970). However, appeal for stylisations *s2/m2 and s3/m3* (4.71, $SD = 1.25$ and 4.95, $SD = 1.25$) were rated highest. In addition, Figure 3.20 depicts that realism *alone* is a bad predictor for appeal. Our results show that *the compatibility of shape and material stylisations, i.e., their matching degrees of realism, has a stronger (and well predictable) influence on appeal.*

3.2.7 Experiment 2b: Effect of Emotions

In previous experiments, we have analyzed the overall effect that shape and material have on the perception of faces. Here, we first analyze whether different levels of stylisation in shape and material, including mismatches between them, affect the recognition and intensity of emotions (*anger*, *happy*, *neutral*, *sad* and *surprise*). We then discuss how ratings are affected by the particular emotion. This is interesting since previous findings suggest that the type of emotion affects the perceived uncanniness (Tinwell et al., 2011) and amygdala response (Calder, 1996), where particularly negative emotions (fear, sadness), induce more intense response from the perceiver. In particular, we seek answers to the following questions:

- Does the level of stylisation affect the intensity of emotions? Are they easier or more difficult to recognize?
- Do negative emotions affect the perceived appeal of characters? Is this influenced by stylisation of shape or material?

Intensity and Recognition of Emotions

As discussed previously, stylisation is a well-known tool for artists to enhance the expressivity of 3D characters, removing unnecessary details and enhancing specific features. In this experiment, we explore how the different stylisations of shape and material affect *recognition* and the perceived *intensity* of the emotions, and which of the two dimensions is dominant for emotion recognition.

We designed a different experiment and invited an alternate set of participants for the purposes of this study. The 250 stimuli set from the previous experiment was used, where each stimulus was presented for 4 seconds in random order; participants were first asked to classify the emotions according to the following options: anger, happy, neutral, sad, surprised. After each answer (except for neutral), a follow-up question asked to rate the intensity of the emotion with respect to a seven-point Likert scale bounded by *1 – Extremely Low* and *7 – Extremely High Intensity.*

Twenty–four new volunteers (16 female, 8 male, 23.6 years old on average) took part in this experiment. As before, the participants were recruited from Bielefield University.

Results are shown in Figure 3.21.

a) Recognition

We found a main effect of Emotion ($F^{*}(1.22, 28.04) = 74.00$, $p \approx 0$, $\epsilon = 0.31$), as well as several interaction effects between Emotion and Shape ($F^{*}(4.56, 104.9) = 41.30$, $p \approx 0$, $\epsilon = 0.285$), Texture ($F(16, 368) = 4.97$, $p \approx 0$) and Model ($F^{*}(2.3, 51.26) = 4.23$, $p = 0.016$, $\epsilon = 0.56$). The neutral expression is mainly responsible for all these effects; its recognition rate was lower ($p < 0.002$) than the other emotions, varying strongly across different shape

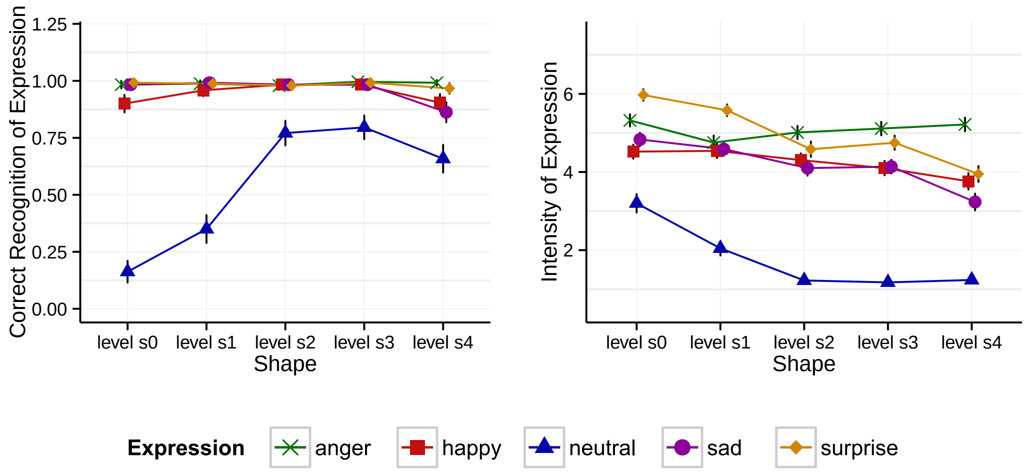


Figure 3.21: Effect of shape on the recognition and intensity of the emotions. Even though neutral condition was not easily identified, all expressions of emotion have been recognized well or independent of the shape. However, the intensity reduced continuously with higher shape stylisation levels.

levels. This neutral expression was in general poorly recognized for the more stylised shapes ($s0$ and $s1$); for instance, some participants reported that the big round eyes made them look surprised. This might be explained by the fact that cartoons are usually designed to enhance expressivity, not to be posed displaying a neutral expression.

We also found significant main effects for Shape ($F(4, 92) = 44.23$, $p \approx 0$), which is mainly determined by the neutral expression, as discussed above and a main effect for Material ($F(4, 92) = 10.09$, $p \approx 0$). The material level $m4$ reduced the recognition rate significantly ($p < 0.015$) but only by 2%.

Overall, we found that *the neutral expression of the characters was least recognised, mainly due to the fact that cartoon stylisation of shape produces expressive characters even when they are in a neutral pose. Realistic material can also slightly decrease recognition rates.*

b) Intensity

ANOVA reveals a main effect of Shape ($F^{*}(2.11, 48.61) = 91.40$, $p \approx 0$, $\epsilon = 0.53$) and Material ($F^{*}(2.47, 56.90) = 30.46$, $p \approx 0$, $\epsilon = 0.62$). The perceived intensity of emotions is continuously reduced with increasing shape levels ($p < 0.001$). Only in the case of shape levels $s2$ and $s3$ does the intensity remain constant. In the case of Material, the absolute difference was very small (0.5 between the lowest and highest mean) and only the material level $m4$ had a higher intensity ($p < 0.001$). This weak effect of realistic material on intensity of the emotion matches previous research (Wallraven et al., 2007, 2008), which found that details such as wrinkles increase the expressivity of realistic characters.

In addition, a main effect of Emotion ($F^*(2.57, 59.10) = 204.60$, $p \approx 0$, $\epsilon = 0.64$) and interactions between Shape and Emotion ($F(5.78, 132.94) = 19.00$, $p \approx 0$, $\epsilon = 0.36$), Material and Emotion ($F(16, 368) = 5.04$, $p \approx 0$), and Emotion and Model ($F(4, 92) = 19.55$, $p \approx 0$) were found. In particular, the happy, sad, and surprise emotions were perceived with lower intensity as the realism of shape increased. This difference was significant in the majority of cases for shape levels $s3$ and $s4$ ($p < 0.01$), but was less frequent for lower shape levels. The perception of the angry emotion, on the other hand, remained constant along shape abstractions.

Overall, we found that *emotions on cartoon shapes are perceived as more intense but that low intensity subtle emotions are harder to convey in abstract characters. Additionally, no or small impact of material on the intensity or emotion recognition was found, which indicates that shape is the dominant dimension when designing characters that are expressive.*

Effect of Emotion on Realism and Appeal

We focus here on the effect of emotion on realism and appeal with the extended stimuli set, the analysis which we conducted as part of the procedure described in Section 3.2.6. Figure 3.22 shows the results, which we analyze below.

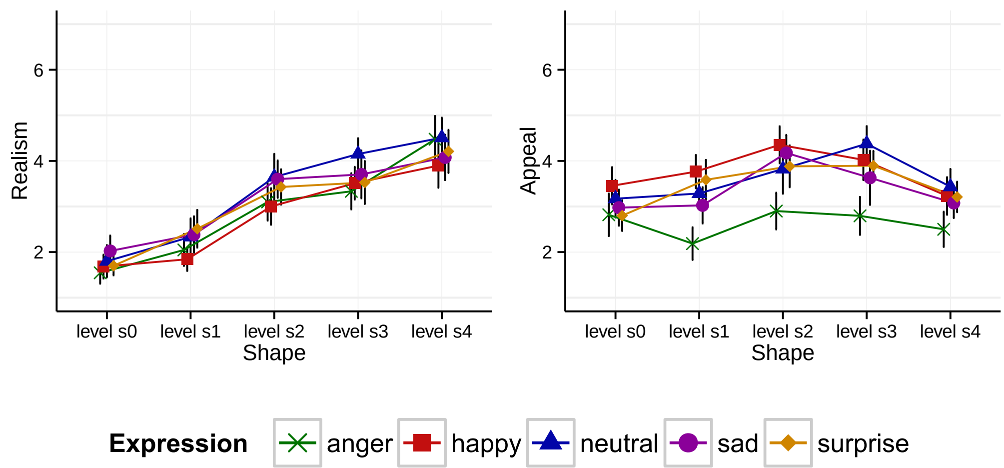


Figure 3.22: Ratings for realism and appeal for different emotions. While emotions do not differ in realism, anger was perceived as more eerie and unappealing for all stylisation levels.

a) Realism

A main effect of Emotion ($F(4, 80) = 10.38$, $p \approx 0$) was found, which could be mainly attributed to the neutral and sad emotions, which have been perceived as more realistic ($p < 0.006$). Because the means are located within a small range (± 0.16), we classify this effect as noise and omit similar examples for the rest of this section. *Nevertheless, equal realism ratings confirm that emotions were well designed by the artists.*

b) Appeal

A main effect was found for Emotion ($F(1.56, 31.36) = 19.34$, $p \approx 0$), which is primarily caused by the emotion of anger ($p < 0.001$). These results reveal that anger is rated much lower with respect to Appeal. Previous studies reported that negative emotions trigger unpleasant responses from the observers (Calder, 1996); our results confirm these studies. Moreover, this effect is maintained even in the presence of highly stylised and appealing characters, suggesting that *negative emotions are perceived as unappealing independent of stylisation level.*

Additionally, we observe that ratings are unsteady across different stylisation levels for the rest of the emotions. In many cases, interaction effects between Emotion and Shape or Material are found ($p < 0.001$). By using photographs, Zhu et al. (2014) showed that different instances of the same emotion do indeed vary in perceived appeal. We believe that this might also be the primary reason for the variations in our ratings. We rule out recognition as an error source, since all the emotions were recognized outstandingly well (see the previous subsection).

3.2.8 General Discussion

Shape and material are two of the main aspects that define the appearance of virtual characters, which in turn are crucial when defining the visual look of animated feature films. We have analysed the perceptual effects of different stylisations along these dimensions on computer-generated faces. In particular, we have studied five different stylisations of two virtual characters – male and female – ranging from very realistic to highly stylised, varying both the shape and the material.

Our results show that the main factor for the perceived *realism* is shape; material only affects how realistic a face looks as the realism of the shape increases. This interesting asymmetry means that while a realistic shape with a stylised material will look stylised, in contrast, a stylised shape with a realistic material will remain stylised. This implies that mismatches in material and shape impact the perceived realism less on abstract characters.

On the other hand, we have found that material is the main factor for perceived *appeal.* In general, appeal, attractiveness, and eerines are highly dependent on the material stylization. Matching levels of stylisation of geometry and material cause the highest ratings of appeal, while strong mismatches (e.g. very realistic material on a stylised shape) result in unappealing characters.

Interestingly, subtle stylisation of a realistic material, such as blurring, increases appeal without sacrificing realism. These stylisations reduce unwanted skin impurities, pores, and wrinkles. These results relate with previous findings on face perception showing that smooth homogeneous skin is generally rated more attractive, since it is a good estimate of a young and healthy subject (Fink & Matts, 2008). However, this trend is only observed for mild stylisations, and stronger ones quickly reduce realism.

Our results are consistent across all tested emotions, except for anger, which was consistently rated less appealing and more eerie. This can be explained by negative or aggressive emotions triggering a defense response and a negative reaction of the viewer (Calder, 1996). Our results are also consistent between different characters. Although small differences between the characters exists, all reported trends are consistent and well visible.

Realism alone was shown to be a bad predictor for appeal, which is not well aligned with the theory of the uncanny valley, although a similar finding was reported for rendering style (McDonnell et al., 2012). One possible explanation is that some of our characters were difficult to categorize by the participants, due to their mismatched appearance parameters (Saygin et al., 2012).

Finally, our experiments show how stylisation affects the intensity of emotion, and that shape is the main factor in this case, whereas material has no significant influence for stylised shapes. This confirms previous knowledge on modeling or drawing expressive stylised characters, where expression of an emotion is mainly determined by the global shape of the character. However, for realistic shapes, we have observed that material stylisation slightly, but significantly, reduces the perceived intensity of emotions.

As in all user studies, our results are only strictly valid for our particular set of stimuli. We have focused on a specific set of stylisations for two realistic characters, varying shape and material following typical designs used in feature animation. This of course limits the universality of the conclusions, which may not generalize if the character styles differ greatly from ours. However, since our design space was densely sampled and the observed trends are consistent between the different characters, we believe that our observations can be used as valid guidelines for creating digital characters within a reasonable range of styles.

3.3 Summary of Chapter 3

This chapter analysed various aspects of perception that are influenced by the mismatch in appearance elements of virtual characters and how their manipulation could affect the perception of gender, emotion, and appeal. Both studies show that mismatches will produce unappealing results and/or change the perception of information that the character is expressing. For motion and appearance mismatch, this meant that the underlying sex of the actor is detectable and will be even more obvious when it is shown on the virtual character of the opposite sex due to the contrast effect. Therefore, matching actor's and virtual character's sex is important for accurate portrayal of a character. The second study confirms that extreme mismatches in shape and material realism produce the least appealing results, therefore these combinations should be avoided.

However, by introducing these mismatches, we were able to get a deeper insight into some components of appearance and how they affect overall perception of the character. The motion and shape mismatch showed how stereotypes can additionally orient our perception and that some motions are richer in gender cues than others, therefore gender perception can be more easily biased by stereotypes on motions with fewer gender cues. The shape and material mismatch study indicated that materials are more crucial for the perceived appeal, while the shape is a predictor of perceived realism.

Furthermore, both studies reveal that the perception of virtual characters is complex and that additional definitions of factors should be considered when designing or interpreting the results of an experiment. For example, when we are talking about how realism affects appeal, what kind of realism do we have in mind – realism of shape or material? When we make claims of gender being recognised from body motion, what is the motion expressing in a social context?

When introducing obvious mismatches of the character's visual components, we found the information the characters are conveying will be compromised. However, by introducing obvious mistakes in the appearance, one can expect the appeal ratings to be lower. However, is it possible that different visual styles of characters would be perceived differently regardless of appeal? This research question, along with others, is explored in the following chapter.

Chapter 4

The Effect of Render Style on the Perception of Virtual Characters

Our previous study indicated that appeal is closely related to material properties of the character while the shape contributes to realism. Similarly, a previous study of McDonnell et al. (2012) showed how render style can change appeal ratings regardless of the perceived realism, which is not supportive to the naïve uncanny valley hypothesis. The same study also showed the render style could have other perceptual effects, such as how trustworthy people find the character (McDonnell et al., 2012), while other studies found changes in social behaviour (Bailenson et al., 2005) and emotional response (Volante et al., 2016) due to varying realism levels of characters. Another study found that the outcomes of higher realism levels might not always be associated with a negative reaction from the viewer (Cheetham et al., 2014). Render style can also affect emotion recognition and intensity (Wallraven et al., 2007).

The main focus of this chapter is therefore not mismatches of visual elements, but rather investigating the importance of different render styles of a character model. We are particularly interested in the distinction between realistic as opposed to stylised characters, in order to study the effect of realism and its potential changes in social cognition. We use the term *render style*, which signifies the visual qualities of the character other than shape: including material properties, lighting setup and post-processing effects.

The studies presented in this chapter continue to explore the effects of character's appearance not only on perceived appeal and realism but other potential factors as well. To test that, we created scenarios for the characters which were typically not considered before in relation to studying the effects of render style: *agency* and *personality* of the virtual characters.

The Agency Experiments, explore the interactivity of characters and how the render style affects it. We were interested in control over the character's motions (agency) since studies found that ownership positively affects the perception of the avatar itself. Specifically, if the motion of the observer's body is tracked and applied to an avatar (body ownership), it will be perceived as more appealing (Bailenson et al., 2005; Kokkinara & McDonnell, 2015) than without body ownership. Our study expands this idea by hypothesising that a level of control over the character, conducted by using the gamepad controls could also have a similar effect and that this might further be mediated with the character's appearance. We implement this in a typical screen–based game environment and compare it with the responses virtual reality to compare the effect of the platform.

The other, Personality Experiments, focus on the personality the character is exhibiting and asks the question, would render style affect the way we respond to a certain personality type, displayed on a character. We use a set of styles which vary in appeal (McDonnell et al., 2012) in order to explore this idea. We are particularly interested if a realistic style would induce different responses to other non-realistic styles. We build upon the mentioned difference between the way we perceive realistic as opposed to non-realistic characters, where we possibly develop automatic social cognitive response when faced with a realistic character (see our theoretical discussion in Section 2.3.4). By creating a longer exposure to the character with observable personality traits, we expect a stronger social response from the observer and possible differences depending on the render style of the character.

We use a combination of different measures for the studies but a notable addition to previous work is the inclusion of indirect measures, applied in social cognitive research: personal distance (proximity) and attribution bias. We justified the use of indirect measures in Section 2.3.4 since the difficulty of processing virtual characters, typically related to uncanny valley, could be attributed to fast and slow mechanisms of brain processing (Chaminade et al., 2007; Macrae et al., 1999; Albrecht & Carbon, 2014). While proximity has been used in the past to study virtual characters, it has never been used to assess render style before. To our knowledge, attribution style has never been used to study perception of virtual characters either.

Another addition is that we choose to analyse the effect of character's appearance in the context of interactive (games) and immersive environments (virtual reality). A wider range of games and interactive applications are reaching high levels of realism, therefore the need to investigate the effect of character's appearance in this environment is important. As for virtual reality, it is already becoming more consumer available and a wider range of its use is expected in the future, yet it is unclear whether high realism will be more appealing to audiences in VR.

We also collaborated with a professional actor in order to record a set of identifiable personality traits exhibited through motion and speech. By doing so, we target a more naturalistic setting where characters are used in order to assess the effect that the appearance could have in a practical implementation, where the character is created to exhibit a personality (animations, games, etc.).

4.0.1 Stimuli creation

For both studies we used similar render styles of one male model, seen in Figure 4.1. We used a realistic mesh to which we applied different materials in order to create distinct stylizations of the model. The render styles were created using two softwares. Autodesk 3ds Max 2015® was used to generate the mesh of the characters, applying basic texture maps (diffuse, specular and animation (motion captured from an actor and imported into 3ds Max imported as the BVH format). When only non-interactive screen–based animation sequences were needed for the experiment, then final videos were rendered using basic three-point light setup and the 3ds Max inbuilt Mental Ray render. For the interactive environment, the animations and the mesh were exported as the FBX format into the Unreal Engine® version 4.9.2. There, additional settings for the materials were used to achieve a desired render style. The specific individual differences are described below. Characters, rendered in 3ds Max are marked as "3dsMax" and characters created for the interactive environment "UE4". Toon CG was rendered for both interactive and non-interactive experiment, therefore has both markings.

- *Realistic* render style (UE4) used textures which were taken from a real actor and is similar to the *Human* render style in the study of McDonnell et al. (2012). The material setup was executed in Unreal Engine, where we used advanced skin shaders (subsurface profile shading model) and a typical material setup to reach a high level of realism based on the work of Jorge Jimenez et al. (Jimenez & Gutierrez, 2010). Special care was also taken in the construction of the eyes, such as including refractions on the cornea, since the eyes carry an important weight on realism assessment of the character.

- *Toon CG* render style (3dsMax, UE4) used a simpler material setup with textures created by an artist to resemble a style commonly seen in animated characters. The aim was to reach a highly appealing style, similar to that used in computer graphics cartoons in the industry. The style uses painted textures to create a soft skin. In appearance, this render style resembles *Toon CG* render style in McDonnell et al. (2012) study. The Unreal Engine version of the render style did not use any additional changes to the basic texture maps created in 3ds Max.

- *Toon Shaded* render style (UE4), was created using a post processing Cel Shader[1] effect, a plugin provided by Unreal Engine, which turned a version of *Toon CG* render style into a drawing style cartoon. The reason we used Toon CG as the base was due to more colour contrast in its texture which produced nicer effects with the mentioned Cel Shader. The components of the shader were set to produce the most appealing effect as well. The shadow effect was reduced as much as possible since it produced artifacts when observed at a distance. This render style was made to resemble the *Toon Shaded* render style in the study of McDonnell et al. (2012) and was created with the aim to be least realistic (but still appealing) representation among other render styles.

- The aim of the *Creepy* render style (UE4) was to induce a sense of discomfort by mismatching realistic appearance with a creepy component – grossly enlarged eyes, since unnatural modifications around the eye area were found to induce the uncanny effect (Seyama & Nagayama, 2007; Tinwell et al., 2013). For this render style, only simple maps of the Realistic render style were used (texture, normal and specular map) and subsurface scattering effect was removed. In addition, the mesh of the virtual character was edited in 3Ds Max 2015 to enlarge the eyes out of proportion.

- *Zombie* render style (UE4) was created to be obviously less realistic and unappealing, by additionally changing the diffuse map for the skin and the eyes of the *Creepy* render style by adding an unnatural colour (green for skin and red for the eyes). These changes resulted in an appearance of a "zombie" character, seen in movies and games. The aim of this render style was to be obviously less appealing, eerie and also non realistic.

- *HumanIll* render style (3dsMax) represented the most realistic version of the three eerie render styles but was modified to appear ill (pale shiny skin, slightly red eyes). To achieve this effect, the diffuse map was desaturated and added a yellow hue was used along with shading the skin of the virtual character using a waxy material. The eyes were glazed over using a semitransparent glass shader on the cornea. This render style is similar to the one used in the study of McDonnell et al. (2012).

The styles Realistic, Toon CG and Toon Shaded, were created to resemble characters in games and animated movies. The other three styles, Creepy, Zombie and HumanIll, were created intentionally to be unsettling and induce a negative response from the viewer.

The lighting setup in the immersive environments was created to enhance the perception of depth and, especially for the Realistic render style, to intensify the subsurface scattering effect on the skin. Specifically, there were two point lights on the front and one spot light

[1]http://www.skull.co.nz/, Retrieved 11 October 2016

Figure 4.1: Render styles for the male model used in the studies presented in this chapter. The images are taken from the experiment conditions, therefore the *HumanIll* render style is shown only to the waist since this camera view was used for the non-interactive experiment.

Figure 4.2: Actor during the capture session while portraying personality traits (*left*) and his motion applied to a virtual character (*right*).

pointing on the back of the character; a commonly used technique in the industry. Real–time shadows by the moving characters and baked direct lighting and shadows from the environment were also incorporated in order to enhance the overall realism and immersion. The character's mesh and skeleton, lighting conditions and the virtual environment were kept constant throughout. The only difference was the rendering in the screen–based personality experiment (Section 4.2.1), which was the only experiment fully constructed in 3Ds Max 2015® and therefore used the specifically assigned render to produce desirable results (Mental Ray).

Motion Capture For the characters' animations, we motion-captured two male actors. For both actors, a 21-camera Vicon optical system was used, where 52 markers were placed on the body, and where facial animation was used, 36 markers on the face (see example on Figure 4.2). We did not capture finger or eye motion. Similar to the approach used in our previous work, the captured body motion was mapped onto a skeleton where joint angles were computed and used to drive the virtual character in Autodesk 3ds Max 2014®. The facial motion was directly applied to the bones of the character's face.

4.0.2 Overview of the Experiment Designs

For clarity, the structure of the studies on the effects of render style are summerised in Table 4.1. We separated the studies to *Agency* and *Personality* groups, where each group consists of two experiments, varying in the types of render styles used, platform, and response measures. All experiments, except for Experiment 1 in the Personality group, have an interactive component, i.e. the participant can explore the environment and trigger events by using gamepad controls. In the first Personality experiment, the participants are passive observers to the character and press the assigned keys for the answer.

Table 4.1: Summary of the experiment design for studies in this chapter. Each experiment uses a subset of render styles, is presented on either screen–based or virtual reality platform and uses a combination of subjective Likert-scale responses and indirect measures.

Experiment	Render Style	Platform	Subjective Responses	Indirect Measures
Agency *Experiment 1*	Realistic Toon CG Toon Shaded	Screen-based	Agency (3 scales), Uncanny Valley (6 scales), Co-presence (3 scales)	Proximity
Agency *Experiment 2*	Realistic Toon CG Toon Shaded	Virtual Reality	Agency (3 scales), Uncanny Valley (6 scales), Co-presence (3 scales)	Proximity
Personality *Experiment 1*	Toon CG HumanIII	Screen-based	TIPI (Ten Item Personality Inventory)	X
Personality *Experiment 2*	Realistic Toon CG Toon Shaded Creepy Zombie	Virtual Reality	Empathy (3 scales), Uncanny Valley (6 scales), Co-presence (1 scale)	Proximity, Attribution Bias

Subjective Measures 5–point Likert scales were used to measure subjective responses of the participants (see Table 4.2). Agency questions (see *Agency, Own Will, External Force*) were designed by us, aiming to assess the sense of control participants felt over the virtual character, versus the sense that the character was moving as a sentient human being or by an external force (e.g. a computer algorithm). The Affinity and Realism questions are based on previously used indices for describing the uncanny valley (Ho & MacDorman, 2010; McDonnell et al., 2012) but we further separated the realism questions based on motion and appearance. We added questions about co-presence (Bailenson et al., 2004) to measure how immersive the environment was. The last three questions measuring empathy were based on the understanding of empathy as experiencing the emotions of others (Davis, 1983). We used the *Concerned* scale, which was used by Davis et al. (1983), while the other two describe emotions, which would in our opinion be most likely induced in the participant: *Excited*, which is an emotion of positive valence signaling engagement with the character, taken from the PANAS questionnaire (Watson et al., 1988) and a negative emotion taken from CAM battery (Golan & Hill, 2006): *Uneasy*). All questions could be answered on a 5–point scale, ranging from *1 – Not at all* to *5 – Extremely*. We chose to use the 5-point scale to match it with the PANAS questionnaire scales.

4.1 Agency and Render Style

This study, which was done in collaboration with fellow researcher Elena Kokkinara (Trinity College Dublin), explores the interactivity with the character in the form of control over the character's motion or the lack of it (agency). Since studies found that ownership positively affects the perception of an avatar (Bailenson et al., 2005; Kokkinara & McDonnell, 2015), we decided to explore this association further by changing the appearance (render style) of the character and observe how this interaction could affect the perceiver's engagement with the character. We tested if a different kind of control over the characters movements (using the gamepad) would make the character more appealing as well. We hypothesise, that the character being controlled cannot produce any unexpected behaviours, as the user is controlling it's movements, which will make it less threatening. This aspect is closely related to the term co-presence, where one of the definitions by Blascovich and his colleagues (2002) is the extent to which individuals treat embodied agents as if they were other real human beings. We aim to examine how people would react to characters that they control as opposed to the ones they only observe and if this is further mediated by appearance.

We did so by creating two experiments. In Agency Experiment 1, we implemented a screen–based task, where we investigate the effects of agency and render style in a typical interaction game design. Here, the participant was able to navigate through the environment and trigger events (and depending on the agency condition also the character's movements). This design was used to test how the character's agency and appearance impacts the response of the user in a typical game environment, since it is the most common interactive platform. However, we also use a combination of both subjective and behavioural measures of co-presence (Bailenson et al., 2005), which is a measure commonly used in virtual environments. We rely on the above mentioned definition of co-presence of Blascovich and his colleagues (2002) which is not limited to a particular platform. Even so, the virtual reality platform could significantly change the level of reported and behavioural co-presence, therefore we repeated the same experiment in a virtual reality platform (Agency Experiment 2). Apart from potentially increased co-presence, virtual reality could have other effects on people's response to virtual characters which would be of interest. For example, appearance realism was not reported to be as important as character's behaviour realism on the response

Table 4.2: Likert-scale subjective responses to explore the effect of render style on the perception of virtual characters. Agency questions were only used in the first (Agency) study, while the empathy questions were only used in the second (Personality) study.

Group	***Question***	***Statement***
Agency	Agency	*"Overall, I felt I was causing the movements of the virtual character in the room to happen."*
	Own Will	*"Overall, I felt I was causing the movements of the virtual character in the room to happen."*
	External Force	*"Overall, I felt as if the character was moved by an external force."*
Realism	Appearance Realism	*"I found the character's appearance realistic ("Not at all" = the character's appearance is highly stylised like in cartoons, "Extremely" = it could almost be mistaken for a photograph)."*
	Movement Realism	*"I found the character's movements realistic."*
	Overall Realism	*"I found the character realistic overall."*
Affinity	Appeal	*"I found the character appealing ("Extremely" = the character is one that you would like to watch more of and would be captivated by a game with that character as the lead").*
	Eerie	*"I found the character eerie ("Not at all" = character restores a sense of security, confidence, calm in me. "Extremely" = character is gloomy and leaves me with a sense of fear)."*
	Familiar	*"I found the character familiar ("Extremely" = I have seen something similar to it before)."*
Co-presence	Co-presence	*"I perceived that I was in the presence of another person in the virtual room with me."*
	Alive	*"The character appeared to be sentient, conscious, and alive to me."*
	Computerised	*"I perceived the character as being only a computerised image, not as a real person."*
Empathy	Concerned	*"I feel concerned about the character:"*
	Excited	*"I feel excited after watching the character:"*
	Uneasy	*"I feel uneasy after watching the character:"*

of people in virtual reality (Garau, 2003), but would the effect of realism in virtual reality be similar in comparison to the screen–based platform? Also, some studies which did not report the effect of appearance realism (Bailenson et al., 2001; Garau, 2003) used examples of realistic characters which would not be plausibly realistic any more due to significant advancements in the technology and software used to create (e.g. Unreal Engine) and display (e.g. Oculus Rift) high levels of realism. We argue therefore, that the need to reevaluate the effect of realism is justified.

4.1.1 Agency Experiment 1: Screen–based

For this experiment, we developed a screen-based game-like environment, where participants could explore and interact with virtual characters. For this purpose, we used characters of different render styles and placed them in a virtual environment, where participants could either observe or trigger character's motions.

Stimuli and Environment Creation for Experiment 1

The characters we used for this experiment were animated using motion capture of one actor, who we instructed to do a set of exercising motions. Facial motion was not used, apart from the blinking which was present in the idle condition. The actor was instructed to do a range of eight different gym exercises on the spot (push-up, squat, left and right leg lunges, forward back-bend, jump, jog on spot and arm stretching), as well as one standing (idle) posture (shifting the weight from one foot to the other, blinking, etc.). We judged gym exercises to be a reasonable set of motions that a person can do on the spot, since we wanted to avoid the characters to move in space. This way we could keep constant and controlled the distance from the viewer. Moreover, we showed a character exercising without acknowledging the presence of the viewer, which was a plausible scenario and did not require additional context that could have distracted participants.

The motions were exported to Unreal Engine and were assigned to the characters. Animations would trigger in a random order during the experiment through an algorithm, or trigger through button presses of the participants. All animation would smoothly blend in and out with the idle animation. Jogging and arm stretching animation would play with the trigger buttons of the controller and would loop until the triggers were released.

The virtual scene for the screen–based experiment was realised using Unreal Engine 4.9.2. The virtual environment consisted of a simplistic space with one virtual room for the training session and three virtual rooms for the experiment session (see Figure 4.3). Each room contained one of the virtual characters, rendered in a particular style: training character with no textures, *Realistic*, *Toon CG* or *Toon Shaded* (see Figure 4.4), and a red spot on

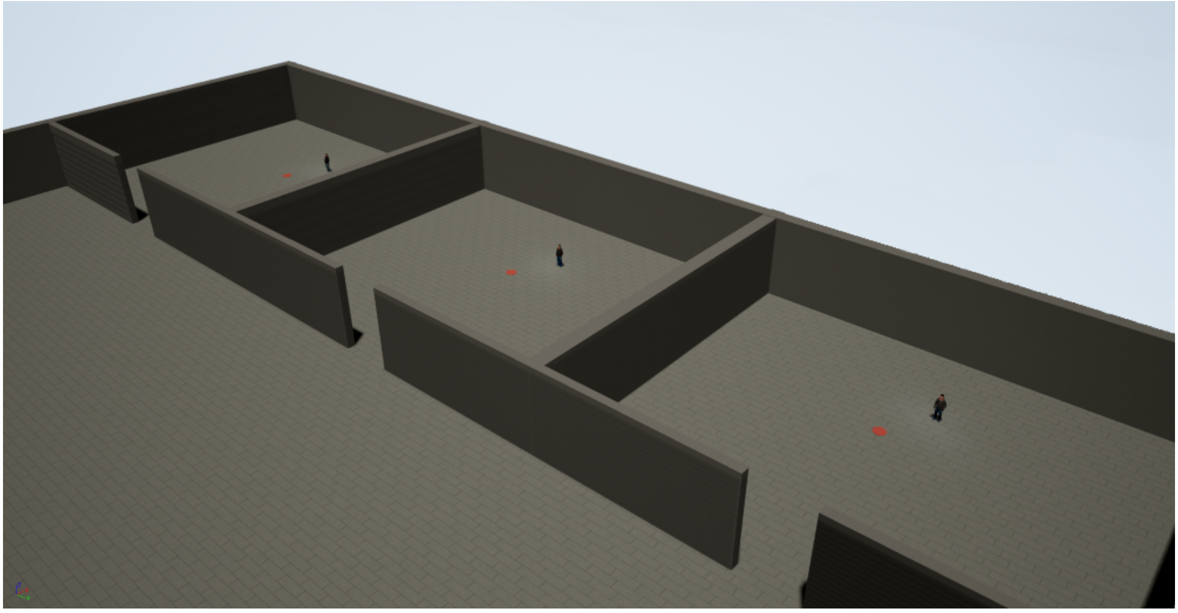


Figure 4.3: Virtual environment for the Agency experiments, viewed from above. Training room was shown in a separate environment, using the same design with the only one room.

the floor, that was placed $4m$ meters away from the virtual character (Figure 4.5). This position was chosen as an optimal observation distance, that would give some personal space to the viewer, but at the same time details of the character could still be visible on the TV screen due to the large screen size (55” Ultra HD screen). We chose a simplistic environment setup with no decoration, in order to avoid distraction of participants view from the virtual characters.

We created a first person camera view which tracked the movement of the participant. It was initially placed outside the first virtual room. Participants were able to navigate (translation and rotation of the camera) using the Xbox controller. We create this step, in order to increase the feeling that the participant was entering a game environment. In the virtual room, we placed a red spot which served as the observation distance. Access closer than $\sim$4 meters to the character was initially restricted with an invisible sphere collider that was placed around the character. Once the first person camera was navigated to the red spot, the camera translation was locked for 100s. This amount of time was necessary for playing each exercise animation two times, which we assumed to be enough exposure to the stimuli. During the exercises session, feedback of the character’s progress would appear on the screen. This element was added to the framework to decrease boredom and give feedback to the participants on which exercises their character had or had not already executed.

Figure 4.4: Render styles used for the Agency Experiment 1: 0. Training, 1. Realistic, 2. Toon CG, 3. Toon Shaded.

Experiment Design

We recruited 42 participants (24 males and 17 females, average age 24.1). Participants were students from different disciplinary backgrounds as well as employed individuals from various fields, recruited mainly via university mailing list and advertisement in the university campus. Most participants played games occasionally or were passionate gamers, only 3 participants never played games before. All participants were naïve to the purpose of the experiment. University ethics approval was granted for the experiment, and participants received a €5 book voucher to compensate for their time.

To measure participants responses to virtual characters, we used behavioural and subjective measures. For the behavioural response, we used proximity which is commonly used as an indication of co-presence and behavioural realism of the virtual character (Blascovich, 2002; Bailenson et al., 2003). We use it additionally as a measure of agency which we predicted to be related to positive evaluation of the character (Kokkinara & McDonnell, 2015) and could potentially signal social attractiveness as observed in real human interactions (McCroskey & McCain, 1974). We used a label reading task that was presented in the study of Bailenson et al. (2003), that was administered in order to ensure that the participant approached the character, for close inspection. For the subjective response, the questionnaire was administered, consisting of three question groups: Agency, Realism, Affinity, and Co-presence (see Table 4.2).

The participants were divided in two groups - Observing and Controlling, where each group saw characters in three render styles with an additional training example (*Realistic,*

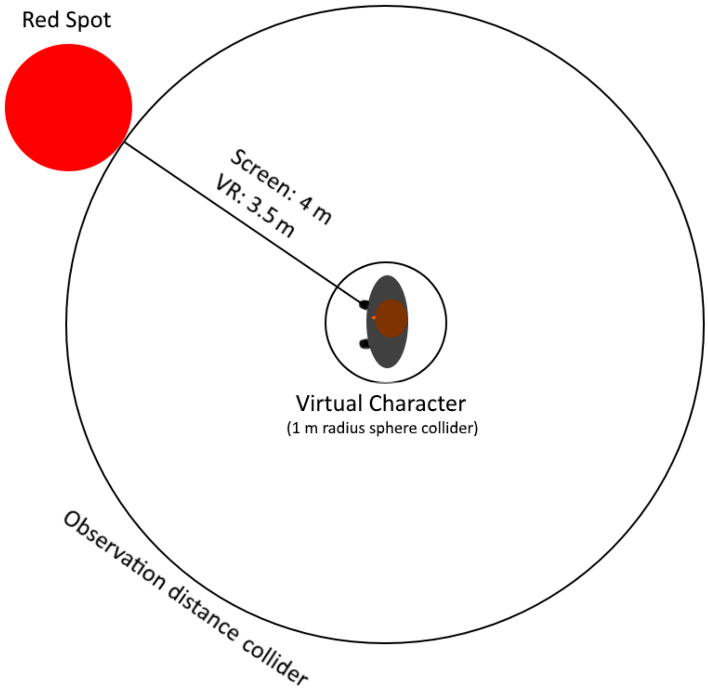


Figure 4.5: Sketch of the top-down view of the room with the red spot and observation distance collider. The observation distance was set according to the platform the experiment was set in; screen-based: $4m$, virtual reality: $3.5m$.

Toon CG, Toon Shaded, no texture). The three render styles were placed in the rooms for participants in a counter-balanced order.

Procedure All participants first signed a consent form and read instructions and then filled a form with demographics information. Participants were seated $1.6m$ away from the TV screen (recommended minimum distance for a 55" Ultra HD screen) and were given the Xbox controller for the interaction with the experiment. The experimenter repeated the instructions verbally. Participants then started with a training session in order to familiarize with the controls, the tasks of the experiment and the questionnaire, with the help of the experimenter. Participants used the controller to navigate (translation and rotation of the camera) the game environment. They were instructed to move inside the virtual room and place themselves (the camera) on top of the red spot. Once the participant reached the red spot, the camera translation locked and the participant was instructed to observe or use the gamepad (depending on which condition they were in). In the observing condition, each exercise animation was shown two times, while in the controlling condition the participant was instructed to explore the buttons on the controller and observe what will happen to the character. During the exercises session, feedback of the character's progress would appear on the screen.

When the observation period was over, the proximity task started. An instruction to look for the text labels around the character appeared on the scene as a 3D text (see Figure 4.6). We used two word labels – one at the front and one at the back of the character, so participants would have to walk around the character. The words were simple and short (chalk, table, pencil, black, red, green). Participants were instructed to read and remember the labels and then move outside the virtual room. As the participant was moving around the character, trajectory data of the camera movement was saved. Participants were free to move as close to the character as they wanted, however intersecting the character was restricted using colliders around the character and the camera. The $1m$ radius sphere collider was chosen because we wanted to prevent participant's camera view to intersect with the character as he is moving on spot. The text itself was already visible from $\sim$3.5 meters distance. Trajectory data of the camera movement was saved during the proximity task, i.e. starting from the moment that the camera control was unlocked, until the moment that the participant exited the virtual room.

On exiting the room, a screen with instructions appeared and participants wrote down the two words on a piece of paper. This task was designed in a way that would minimise additional attentional load of a memory task, hence the words chosen for the labels were very simple. All participant correctly reported all the words. After writing down the two words, we administered a questionnaire regarding the virtual character which was presented in the room in order to additionally collect the participant's direct opinions. We created an 12-item questionnaire, where the items were take from groups Agency, Affinity, Realism and Co-presence (see Table 4.2). The questionnaire appeared on the screen and participants were asked to move a virtual slider and select their answer by using the controller. When participants had visited all three rooms and completed all the tasks they were allowed to leave comments on an open text form.

Results

Data from 1 participant in the observe-only condition was excluded due to misunderstanding of the instructions (the participant forgot to perform the proximity task in at least one room). We did not consider data from training periods.

Each questionnaire item was analyzed separately. We performed repeated measures ANalysis Of VAriance (ANOVA) with between factors *Condition* (Controlling and Observing) and *Participant Sex* and within factor *Render Style*. Due to a high amount of scales, we present them grouped in four categories, outlined in 4.2: *Agency, Affinity, Realism* and *Co-presence*). Tukey Honestly Significant Difference (HSD) test was used as a post-hoc test to analyse significant interactions. The results are presented in Figure 4.7.

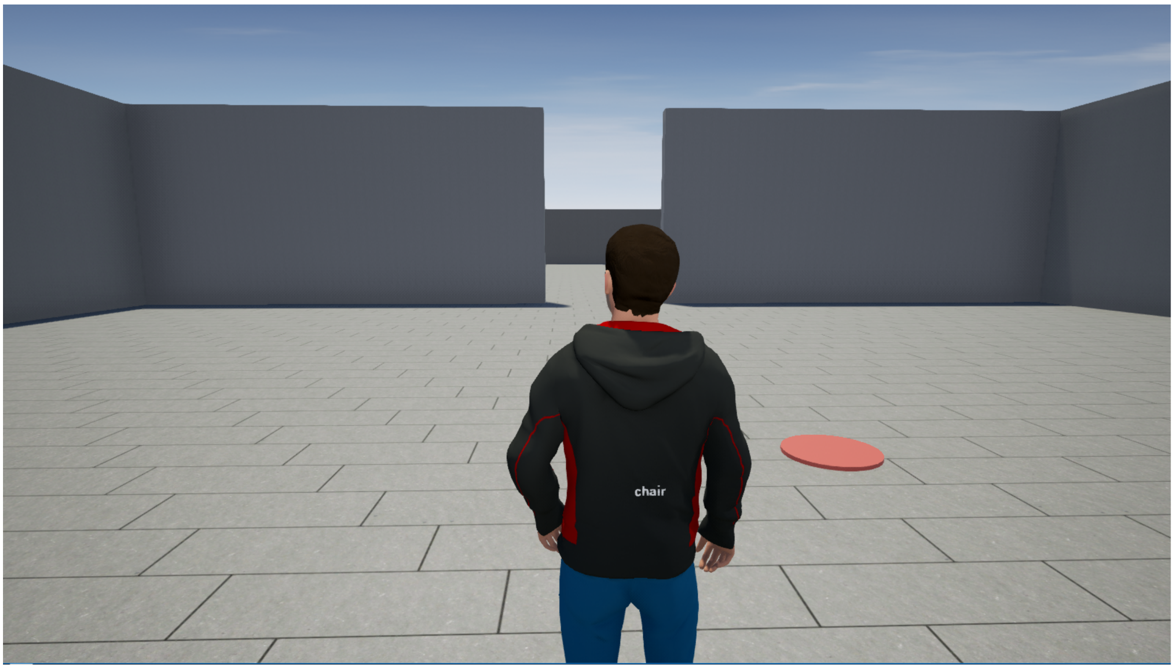


Figure 4.6: The virtual room. Character is shown from the back view where the second label was presented for the proximity task. Participants were instructed to navigate around the character, find both labels and exit the room (opening in the wall in front of the character).

a) Agency

As expected, participants reported high levels of Agency towards all virtual characters only in the Controlling group, as shown by the significant main effect of Condition ($F(1, 37) = 257.27, p \approx 0$). There was also a significant interaction with the Render Style ($F(2, 74) = 3.89, p = 0.024$) and Participant Sex ($F(2, 74) = 3.50, p = 0.035$). Post–hoc revealed however, that there was a significant difference only in the case of the Toon Shaded render style, for which female participants reported higher levels of Agency than male participants ($p = 0.018$). No other effects were found for other variables of Agency – Own Will and External Force. We see that *our two groups, Observing and Controlling, differed in the level of perceived agency towards the character, as expected.*

b) Affinity

We found no main effects for the Appeal or Eeriness scales. There was a main effect of Render Style, which made a difference on Familiarity ratings ($F(2, 74) = 10.22, p = 0.001$). Post–hoc revealed that the Realistic render style was considered significantly more familiar than the other render styles ($p < 0.003$). There was also a main effect of Participant Sex on Familiarity, where females gave lower Familiarity scores overall ($F(1, 37) = 4.37, p = 0.044$) and a further interaction with the Condition ($F(1, 37) = 8.52, p = 0.006$) showed that this

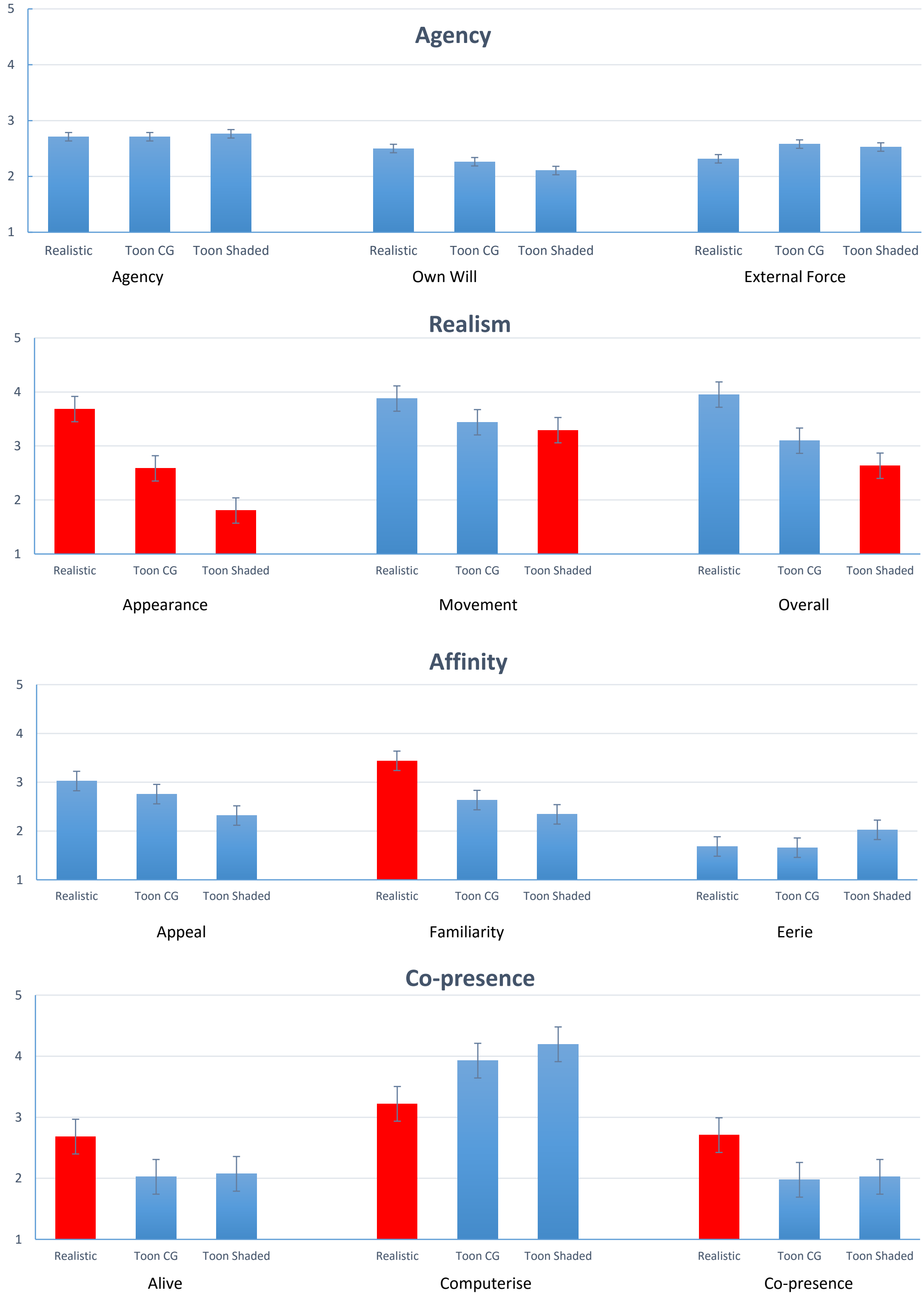


Figure 4.7: Experiment 1: Questionnaire Ratings per Render Style. The red bars label significant difference with at least one more variable.

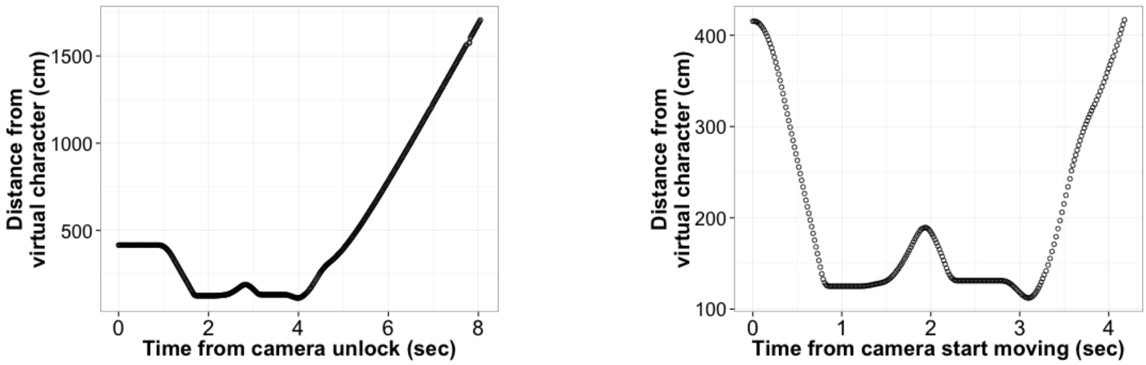
was particularly the case when they were in the Observing group ($p < 0.009$). One more interaction was found for Eerie scale with the Condition ($F(2, 74) = 4.09, p = 0.021$), where participants in the Observing group found the Realistic render style less Eerie than the Toon Shaded render style, post–hoc ($p = 0.012$). We conclude *that the Realistic render style is more reassuring than the other two styles, indicated by higher levels of familiarity and slightly lower eeriness level when this render style is observed.*

c) Realism

We found a significant main effect of Render Style on Appearance Realism ($F(2, 74) = 31.25, p \approx 0$), with all styles significantly different from each other and higher scores in the order Realistic (post–hoc: $p \approx 0$ for all), Toon CG, Toon Shaded (post–hoc: $p < 0.003$ between both non-realistic styles). Interestingly, a main effect for Render Style was found for Movement Realism ($F(2, 74) = 3.79, p = 0.027$), where Realistic render style was judged to have the most realistic motion, but post-hoc showed this was only true in comparison with Toon Shaded render style and the same was found for Overall Realism ($F(2, 74) = 25.43, p \approx 0$). Overall, *Realistic render style was recognised as most realistic in appearance, as expected, followed by Toon CG and Toon Shaded.* In some cases the Realistic render style even changed the *perception of movement to be perceived as more realistic.*

d) Co-presence Questions

Again, the Render Style had an effect on the Co-presence ratings. Participants perceived the characters to be more Alive and felt more Co-presence and less like a Computerised in the case of the Realistic render style (Alive: $F(2, 74) = 7.23, p = 0.001$, post–hoc: $p < 0.005$; Co-presence: $F(2, 74) = 8.29, p = 0.001$, post-hoc: $p < 0.002$ for all; Computerised: $F(2, 74) = 17.42, p \approx 0$, post–hoc: $p < 0.001$). There was also an interaction between Render Style and Participant Sex for the Computerised ($F(2, 74) = 3.24, p = 0.045$), where male participants judged the Realistic render style to be less like a computerised image even in comparison to both styles ($p < 0.008$ for all), whereas female participants judged only the Toon Shaded render style to be more like a computerised image in comparison to Realistic render style ($p < 0.02$). We conclude that *the Realistic render style raised the level of perceived co-presence, a tendency which was slightly more pronounced when rated by male participants.*



(a) *(b)*

Figure 4.8: Distance of of a randomly chosen participant (Controlling condition) from the Realistic render style in time during the proximity task; a) Entire trajectory from the moment that camera unlocks from the red spot until participant exit the room; b) used trajectory for analysis: from the moment participant started moving until exited the $4m$ radius circle.

e) Proximity

During the proximity task (see Section 4.1.1), the Euclidian distance of the current camera (user) position and the virtual character position in the virtual space was calculated for each moment (Figure 4.8, *a*). In order to calculate a representative value for the proximity task, the trajectory data needed to be cleaned. First, we discarded the part of the trajectory where the participant remains static at the beginning of the proximity task. We also discarded the part of the trajectory where the participant departs from the 5–meter circle ($4m$ from the $1m$ character collider) that surrounds the virtual character, heading towards the exit of the current virtual room. Some participants moved outside the circle but returned close to the character before heading towards the exit. We only discarded the data from the last time that they exited the circle.

Cleaned trajectory data was not found to be normally distributed (see Figure 1 in Supplementary material). Hence, we chose to use the median of the distances for each participant as the response variable for the proximity task.

Table 4.3 shows the mean and standard errors (removing the between-subject variability (Cousineau, 2005; Morey, 2008)) of the median distance for participants on each controlling group and rendering style. Median distances from the characters were smaller for participants in Controlling condition, especially for the Toon Shaded style. However, variations were especially large.

To formally test for possible effects of our factors, we conducted a mixed-design ANOVA. Rendering style was the random effects factor, while character control factor was the fixed

Table 4.3: Descriptive statistics for medians of distances in Proximity task (Screen-based Experiment): Means, standard deviations (SD) and standard error (SE).

Observing

	N	Mean (cm)	SD	SE
Realistic	20	185.58	34.99	7.82
Toon CG	20	193.24	24.18	5.41
Toon Shaded	20	194.8	35	7.83

Controlling

	N	Mean (cm)	SD	SE
Realistic	21	182.31	36.76	8.02
Toon CG	21	180.77	34.62	7.55
Toon Shaded	21	173.91	31.39	6.85

effects factor. Data was tested for normality. No main effect or interactions of Render Style or any of the Condition groups (Controlling and Observing) were found. However, we found some interesting correlations with the questionnaires (see below).

Correlations We analyzed the possible relationships between behavioural measures of Proximity and the subjective responses to the questionnaire to assess if people's perception of the character is reflected in their behaviour. Non-parametric Sperman's rank correlation tests were conducted on the results of both groups (Condition) of participants separately to analyze the data.

For the Observing group, a correlation between Proximity and Appeal was found. Participants from this group came slightly closer when they perceived the character to be more appealing ($r_s = -0.27, p = 0.048$). Even though the participants from this group did not have control over the character, they reported various levels of control and we found that Agency was related to Proximity ($r_s = -0.32, p = 0.014$). Participants also came closer when they perceived the character as more Alive ($r_s = -0.26, p = 0.044$).

For the Controlling group, similar patterns were not observed. We found no correlations between Proximity for perceived Agency or Appeal. However, participants came closer to characters, which they rated higher on the Eerie scale ($r_s = -0.33, p = 0.008$).

Overall, we conclude that *control over the character, but only perceived control, will make people move closer to the virtual character.*

Discussion

Our questionnaire results indicate that the realistic render style was perceived as generally more familiar and was given higher ratings of co-presence, which argues against some assumptions that realism should be avoided when creating virtual characters (see for example (Schneider et al., 2007)). It also shows that our realistic render style was constructed well and did not exhibit any off-putting mismatches of visual elements (Seyama & Nagayama, 2007; MacDorman et al., 2009; Saygin et al., 2012). However, we applied realistic motions to all render styles, and this mismatch in appearance and motion could have lowered their appeal. In the least realistic style, the appearance even lowered the perceived realism of motion.

We didn't find significant effects of character control or render style on proximity, but we did observe some interesting relationships between how close people moved to the character and how they rated the characters afterward. Specifically, participants moved closer when they perceived the character to be more appealing and more alive, but only when they were observing as opposed to controlling. This is in line with our initial hypothesis and studies in social psychology where people go closer to the ones they have a positive attitude towards (Evans & Howard, 1973) and are socially attracted to (McCroskey & McCain, 1974).

Participants also moved closer to characters, which they rated more eerie, but only when in controlling condition. This result could be explained by our prediction that a controlled character is less threatening, while participants might have felt more comfortable to explore this eeriness out of curiosity. Another possible explanation could be that those participants who moved closer to the characters, did so to observe more details in the face which is where eeriness is generally the most strong.

Participants in our study also came closer to characters, when they felt that the characters were more alive. This may be due to the fact that they had more interest in observing an entity that appeared alive to them. Again, this would point more to the interest of observation and less to social behaviour.

Control over the character affected the observed co-presence, which was expected. Not having control over the character increased the sense that the character is a separate entity and increased a sense of aliveness of this character.

4.1.2 Agency Experiment 2: Virtual Reality

To determine if the higher immersion and sense of co-presence with virtual characters would alter our results on proximity and affinity from the Screen–based Experiment, we implemented the same environment in the virtual reality setting.

The experiment framework was similar to Experiment 1. However, some modifications were made to maximise the performance on the current Oculus Rift system and to minimize the risk of simulation sickness. Specifically, light intensity in the scene was reduced, while we also removed one of the three lights on the characters' scenes (spot light behind the character). This permitted a frame rate above 75 fps, which is recommenced for Oculus Rift DK2. Although the added value of the third light would be an enhanced perception of depth on a flat TV-screen, we assumed that this would have minimal impact on a immersive VR setup, where stereoscopic view and motion parallax ensure perception of depth.

Moreover, using the same observation point (see the red spot on Figure 4.5) as in the Screen-based Experiment, many details on the virtual character were inevitably lost in the VR setup, due to the screen resolution (1080x1200 per eye). We decided to move the observation point closer to the virtual character ($3.5m$), in order to allow a similar view of the rendered details (the recommended distance for objects in focus in DK2 is 0.7-3.5m).

Navigation was different in VR, since we used head tracking for camera rotations, and the Xbox controller for translation. We adjusted the speed of the camera and restricted side movements and camera rotations using the controller, in order to prevent simulation sickness. These changes significantly differentiated the way the proximity text was performed, hence, we avoided direct comparison of results from the Screen-based Experiment. Additionally, underestimation of distance has often been shown in VR (e.g., (Loomis et al., 1992), etc.) so a direct comparison would not be appropriate.

Finally, the questionnaires were created as flat billboards inside the environment in order to allow stereoscopic view and to avoid interrupting the immersion.

We recruited 38 participants (28 males and 10 females, average age: 27.9) with the university department ethical approval and compensation of €5 for their time. Participants were recruited as before with similar demographics. One participant was deaf, but did not show any difference on the task performance from the rest of the sample, so his data was not excluded. Participants were informed about the risks of VR and signed a consent form. They also had to initially adjust and calibrate the HMD. All other methods and procedures were kept the same as in the Screen-based Experiment.

Results

The same methods of statistical analysis were used as in the previous experiment. The results per individual scales are presented, as before, grouped in 4 categories 4.2. Summary of results is seen in Figure 4.9.

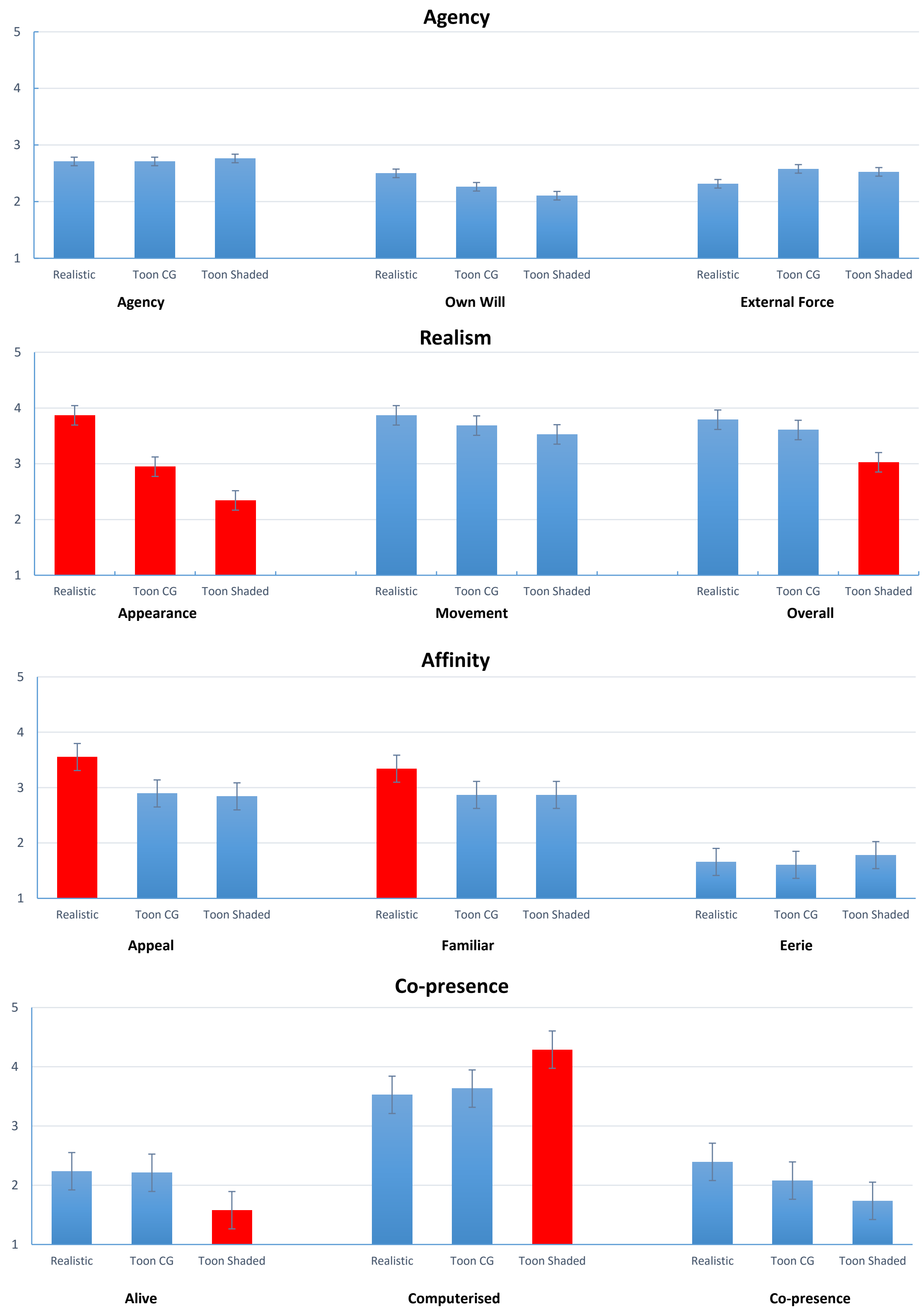


Figure 4.9: Experiment 2: Questionnaire Ratings per Render Style. The red bars label significant difference with at least one more variable.

a) Agency

Condition, as expected, had a main effect on the perceived Agency ($F(1,34) = 106.94, p \approx 0$), where controlling the character would induce a higher level of agency. There was only one more main effect in this category of questions – regardless of which character participants saw or if they were controlling him or not, Participant Sex had an effect on Own Will ($F(1,34) = 4.58, p = 0.040$), where males would perceive to a greater extent that the characters possessed their own will. There was no main effect of Render Style or Condition for the External Force question. As in Experiemtn 1, *agency was detected in the way it was intended by the experiment design, and we also observe some gender differences in the level of perceived control.*

b) Affinity

The Realistic render style was judged higher on Appeal overall ($F(2,68) = 4.19, p = 0.019$; post-hoc: $p < 0.03$, for all). Similarly, Familiarity ratings were significantly higher for a Realistic render style as well ($F(2,68) = 3.48, p = 0.036$), post–hoc however did not find significant differences for the Realistic render style in comparison to Toon Shaded and Toon CG. No effects were found for Eerie. We conclude that *the Realistic style in virtual reality was found to be more appealing and slightly more familiar than less realistic styles.*

c) Realism

Appearance Realism was judged with higher scores in the expected direction for all types of Render Style ($F(2,68) = 17.88, p \approx 0$; post–hoc: $p < 0.02$, for all). Movement Realism was not affected by Condition, while Overall Realism was again influenced by Render Style ($F(2,68) = 5.49, p = 0.006$). Post–hoc revealed that Toon Shaded had a significantly lower realism overall than the other render style ($p < 0.015$ for all). Similarly as Experiment 1, *the realism of the Realistic render style was perceived to be higher than other styles while Toon Shaded was least realistic overall.*

d) Co-presence

Opposite to our expectations and regardless of the fact that the experiment was conducted in an immersive environment, Co-presence was rated low for all conditions. There was an effect of Render Style in the case of Alive, where Toon Shaded render style was perceived significantly less alive than other two styles ($F(2,68) = 4.08, p = 0.021$; $p < 0.016$

Table 4.4: Descriptive statistics for medians of distances in Proximity task (VR Experiment) according to the Render Style: Means, standard deviations (SD), standard error (SE) corrected for within variability.

Observing

	N	Mean (cm)	SD	SE
Realistic	19	175.66	27.05	6.21
Toon CG	19	192.90	30.62	7.02
Toon Shaded	19	184.34	35.91	8.24

Controlling

	N	Mean (cm)	SD	SE
Realistic	19	181.37	20.16	4.62
Toon CG	19	182.61	21.4	4.91
Toon Shaded	19	186.31	27.37	6.28

for all) and the same style was considered significantly more as a Computerised image ($F(2,68) = 4.64, p = 0.013$; post–hoc: $p < 0.015$ for all). Therefore, *even though this experiment was designed to be more immersive, the co-presence was low, especially with the Toon Shaded render style.*

e) Proximity

Table 4.4 reports the mean and standard errors of the median distance for participants in each Condition. Average median distances from the characters were smaller for participants in the Controlling group, especially for the Toon CG style. However, as with Experiment 1, variations were especially large. The mixed effects ANOVA results did not show any main effects of the factors.

Correlations We computed Sperman's rank correlation test between behavioural measures Proximity and questions about perceived agency, co-presence and affinity and realism towards the character, for reasons already described in section 4.1.1, paragraph Correlations. This was done for both groups of participants separately.

In the Observing group, only one significant correlation was found, where participants kept more distance between them and the character when the perceived aliveness (Alive) was higher ($r_s = 0.30, p = 0.024$).

In the Controlling group, participants got closer to the characters when they perceived to have a higher level of control (Agency) over the character ($r_s = -0.29, p = 0.024$), as we predicted. Participants also got closer if they rated Movement Realism higher ($r_s = -0.27, p = 0.044$) and when they perceived the character to be highly Computerised ($r_s = -0.43, p = 0.001$).

Again, we see that *control over the character reduces the proximity distance, possibly indicating comfort when approaching the character.* The threat appeared to be present when observing the character and perceiving him to be alive, a combination which increased the distance between the observer and the character.

Discussion

In this experiment, the realistic style was perceived significantly more appealing and familiar that the other two styles, which suggests that realism is a positive choice for virtual characters in immersive environments. Our results on realism perception also indicate that observed characters are perceived as more real as opposed to the characters that are controlled by the user.

An interesting but unexpected result was that co-presence was low in all conditions. This could be due to the fact that the character was not engaging enough or because the character's behaviour was not 'natural', meaning that participants received no response from the character to their presence in the room, which is usually indicated by eye-contact (Bailenson et al., 2003).

The proximity correlations also confirmed our expectation that an increased feeling of co-presence is related to keeping more distance to the character, especially when the person is just an observer. In addition, participants that had control over the character and perceived them only as a computerized image, went closer to it. However, the same participants would also go closer if they perceived the movement to be more realistic, which could mean they were interested in it because it was realistic or they found it more realistic because they could observe it closely and see finer details of the motion.

4.1.3 General Discussion for Agency Experiment 1 and Experiment 2

An important finding of both experiments is that high level of appearance realism in interactive environments is preferred over other, more artistic stylisations. This is especially true for immersive environments, such as the one used in our virtual reality experiment. This finding is also supported by the fact that other render styles were not found to be particularly unappealing and eerie, therefore were not constructed inappropriately. It is also evidence that new advances in game engines (in our example Unreal Engine) are powerful enough to use realistic virtual characters (and possibly environments) for a positive interactive experience. To our knowledge, this is the first study that used high-level designed realistic characters to test these effects on both platforms.

The results also indicate that control over the character will make him less threatening but not necessarily more appealing, as found in previous studies (Bailenson et al., 2005; Kokkinara & McDonnell, 2015). However, the mentioned studies were examining the ownership over the character's motion, which is possibly a different experience than triggering the motions by using gamepad buttons and not mirroring the actual motions.

Overall, our results did not indicate any proximity differences according to render style, however, in certain circumstances, we were able to validate our initial hypothesis that lack of perceived control caused participants to stay further away from the character. In a screen-based setup, people went closer to more appealing render styles especially when they were in the observing group, which might mean they felt more comfortable with them. This could be a valuable insight for the design of player and non-player characters during the creative process of a game. However, the same was not true for the virtual reality platform, where appeal was not the mediator of the proximity behaviour, since we found that an increased "aliveness" of the character made people stay further away. This could mean virtual reality induces different behaviour responses to screen-based experience, possibly more related with co-presence or the feeling that the character is alive.

As part of our experiment design, we included only minimal, neutral facial cues as this was not a factor that we were testing in this experiment. However, previous work has found differences in brain activation for approaches to menacing or friendly appearances (Carter et al., 2013). Previous studies also found differences in proximity when the character's eye-gaze was focused on the participant (Bailenson et al., 2003).

In our experiment, we used just one male character model. It is possible that different effects would occur with different character appearances, or with more stylised geometry.

4.2 Personality and Render Style

This study explores the effect of render style of the character, expressing a set of personality traits, on the indices of uncanny valley and social cognition. The reasoning behind this design was primarily to introduce a more engaging behaviour of the character, who expresses personality traits. We also included facial animation which was shown to have an effect on the character's pleasantness (Carter et al., 2013).

This study stands out from our previous work in a number of ways. First, it was conducted on an extremely large number of participants from various backgrounds (1106 individuals), since the experiment was featured and supported by the Science Gallery Dublin exhibition titled *Seeing*[2]. This makes our results notably more reliable. Another aspect of this study was that its design was conducted as a set of between–subject trials, where each participant saw only one example of the render style and one personality situation, avoiding the negative effects of repeated measures, such as variance in responses due to comparison with previous conditions. A noticeable addition to the scenario was also sound and aesthetic design of the environment which created a pleasant atmosphere (see Figure 4.13) since we also wanted to create an experience which would be interesting to the pool of participant who were visitors to a gallery exhibition.

In addition to measuring how close they would get to the character as described in Agency Experiment 1, we introduced two more indirect assessments to study how render style affects perception: attribution bias (Gilbert et al., 1988) and empathy (Bouchard et al., 2013) towards the character. For this reason, we make a clearer distinction between realistic and other stylised characters, since we predict the realistic character will trigger social cognition related to real humans whereas stylised would not. In the case of stylised render styles, we do not expect a strong social response, since the social schema used for real humans should not be triggered. Some indications of this were found in the study of Bailenson et al. (Bailenson et al., 2005) where realistic characters were able to elicit embarrassment reactions from participants, whereas a recent study found a similar indication for empathy, where medical students reported higher levels of empathy for a realistic character (Volante et al., 2016). Therefore, we expect a higher empathetic response for realistic characters and differences in proximity behaviour for different render styles. However, we do not make predictions about attribution bias; we explore it to gain a deeper understanding on how the brain is processing information about the virtual character. According to our model, the difficulty of visual processing of the characters would be seen in the results as a tendency to attribute the origin of character's behaviour to his stable personality traits, since this requires less cognitive power than the analysis of situational context which would prompt attribution towards situation.

We created two experiments. Personality Experiment 1 served as a pilot test, where we tested the possibility that personality traits could be faithfully recognised by observers when expressed by a virtual character. We also wanted to test if render style could affect or bias the perceived personality traits. Then, for the main Experiment 2, we continued to created a large-scale, between groups experiment where, as discussed earlier, we analysed the peoples' emotional responses to the character (empathy), as well as perceived affinity, realism, co-presence, proximity and attribution bias.

[2] https://dublin.sciencegallery.com/, Retrieved 13 October 2016

4.2.1 Personality Experiment 1: Screen–based

Although psychology literature is rich with research on personality, there is a lack of resources providing situation descriptors or dialogue to convey a range of complex personalities for virtual characters. Our first task was to use the available literature in order to build such a resource that could be used for future studies on personality perception.

For this experiment, we were further interested if personality we created for the character could be altered with an 'uncanny' or unappealing appearance. Personality traits are related to appeal, particularly as one study showed Tinwell et al. (2013) that particular negative personality traits are attributed to characters with an uncanny appearance. Studies from psychology showed that some personality types are more socially desirable than others (Musek, 2007) and that appearance of an individual influences the way his or her personality will be perceived (Borkenau & Liebler, 1992; Naumann et al., 2009). We used this information to hypothesise that we would attribute negative characteristics of behaviour to unappealing render styles. For this, we applied the motion from the actor portraying personality traits to a character rendered in two different styles, where one style was deliberately created to be appealing and the other unappealing.

Stimuli Creation

For the creation of stimuli, we invited a professional actor to our motion capture studio, where both facial and body motion of the actor was used. Audio was captured simultaneously using a Behringer C-2 studio condenser microphone placed in front of the actor. 20 animation clips were created for Experiment 1 (10 situations x 2 render styles) using Autodesk 3ds Max 2014®. The average length of a clip was 37 seconds (min: 30 sec, max: 47 sec).

The actor was instructed to react to different emotional situations which we created in order to bring out the Big Five traits and their polar opposites (10 different scenarios in total). The descriptions are based on the Big Five facets (Goldberg, 1990; Costa & McCrae, 1992; John et al., 2008), zero-acquaintance observations (Mehl et al., 2006) and language analysis (Gill & Oberlander, 2003; Dewaele & Furnham, 2000). The actor was instructed to perform responses to the 10 situations (see Appendix, Table 6.1 and Table 6.2) by taking into account personality and situation descriptions. For example, the actor was told he would portray an extraverted personality in a situation where he received a new television as a birthday present from his friends. He was instructed to express gratitude, be humorous, show positive emotions and mention how important his friends are. He was also given information on the type of body language (e.g., expressive, physically animated) and language (e.g., informal, use plural "we") he should use. The length of each performance depended on the

Figure 4.10: Two render styles of the model: Toon CG style (*left*) and HumanIll style (*right*).

actor's free interpretation of the character and the capture was stopped when all the main characteristics of the personality were expressed.

Experiment Design

A between-groups design was used for this experiment, where participants in the first group viewed the 10 animation clips rendered using the ToonCG style and participants in the second group viewed the 10 animation clips rendered using the HumanIll style (see Figure 4.10). This paradigm was used as it allowed us to indirectly measure the effect of render style, without participants having explicit knowledge of what was being tested (i.e., if a within-groups design was used, a contrast effect may have occurred where participants would be aware of the different styles being tested and might alter their personality ratings accordingly).

36 participants (18 per group, 15 females and 21 males), aged 19 – 44 (average age: 28.4), took part in the experiment. Ethical approval was granted from the university in order to proceed with the study. Participants were students from different disciplinary backgrounds as well as employed individuals from various fields, recruited mainly via university mailing list. They were naïve to the purpose of the experiment, and were given €5 book vouchers as a reward for participation.

Participants were first asked to rate their own personality on the Ten-Item Personality Inventory (TIPI) (Gosling et al., 2003). TIPI is an questionnaire which measures the Big Five factors as dimensions. Each scale describes one pole of the dimension with two markers taken from existing Big Five instruments (Goldberg, 1990). For example, *extraverted, enthusiastic* are markers of the positive pole of the Extraversion dimension (*reserved, quiet* are markers of the negative pole).

After the self-ratings, the 10 animation clips were shown per participant in random order, and they were informed that the character possessed different personality traits in each clip. After each trial, they were asked: "How appealing do you find this virtual human? Use all cues: appearance, motion and audio", on a scale from 1 (Not Appealing At All) to 7 (Extremely Appealing) by pressing the number keys on the keyboard. Following this, participants rated the character's personality on the TIPI scales (for a full version of the questionnaire, see Appendix). They did this by reporting their level of agreement that the character's personality matched that of each trait marker pair, from 1 (Strongly disagree) to 7 (Strongly agree) and then continued to the next animation. The TIPI traits can been seen on the horizontal axis of Figure 4.11, presented as E - Extraversion, A - Agreeableness, C - Conscientiousness, O - Openness, ES - Emotional Stability.

For clarity, the 10 situations and the TIPI scales were named differently to avoid confusion. Therefore the situations are labeled Situation 1 to 10, where the first five situations describe personalities with high scores on the Big Five factors (S1 – Extraverted Personality, S2 – Agreeable Personality, S3 – Conscientious Personality, S4 – Open Personality and S5 – Emotionally Stable Personality) and the other five lower scores on the same factors (S6 – Low Extraverted Personality, S7 – Low Agreeable Personality, etc.). The scales the participants used to measure the personalities were: Extraversion, Agreeableness, Conscientiousness, Openness and Emotional Stability.

Results

We first recorded ratings of the 10 scales of the TIPI questionnaire, for all participants. As is common practice for analyzing TIPI ratings (Gosling et al., 2003), we then collapsed the 10 scales down to 5 by averaging the scores for the positive poles (e.g., Extraversion) with the reverse scores for the opposite pole (e.g., Low Extraversion). This gave us 5 Personality scales in total (Extraversion, Agreeableness, Conscientiousness, Openness, and Emotional Stability). For each of the scales, a three-way repeated measures ANalysis Of VAriance (ANOVA) was conducted with within-groups factor *Situation* (S1 – S10) and between-groups factors *Participant Sex* (2) and *Render Style* (2). For all scales, there was no main effect of Participant Sex, implying that in general, the ratings of personality were consistent across male and female participants. Furthermore, we tested for a correlation between the participant's own personality ratings and their ratings of character on the same scales and found no correlation, implying that in our study *the sex and the personality of the participant did not affect their ratings of character personality.*

a) Personality Scales

A main effect of Situation occurred for all scales (Extraversion: $F(9,288) = 45.26, p \approx 0$, Agreeableness: $F(9,288) = 36.69, p \approx 0$, Conscientiousness: $F(9,288) = 29.62, p \approx 0$, Open-

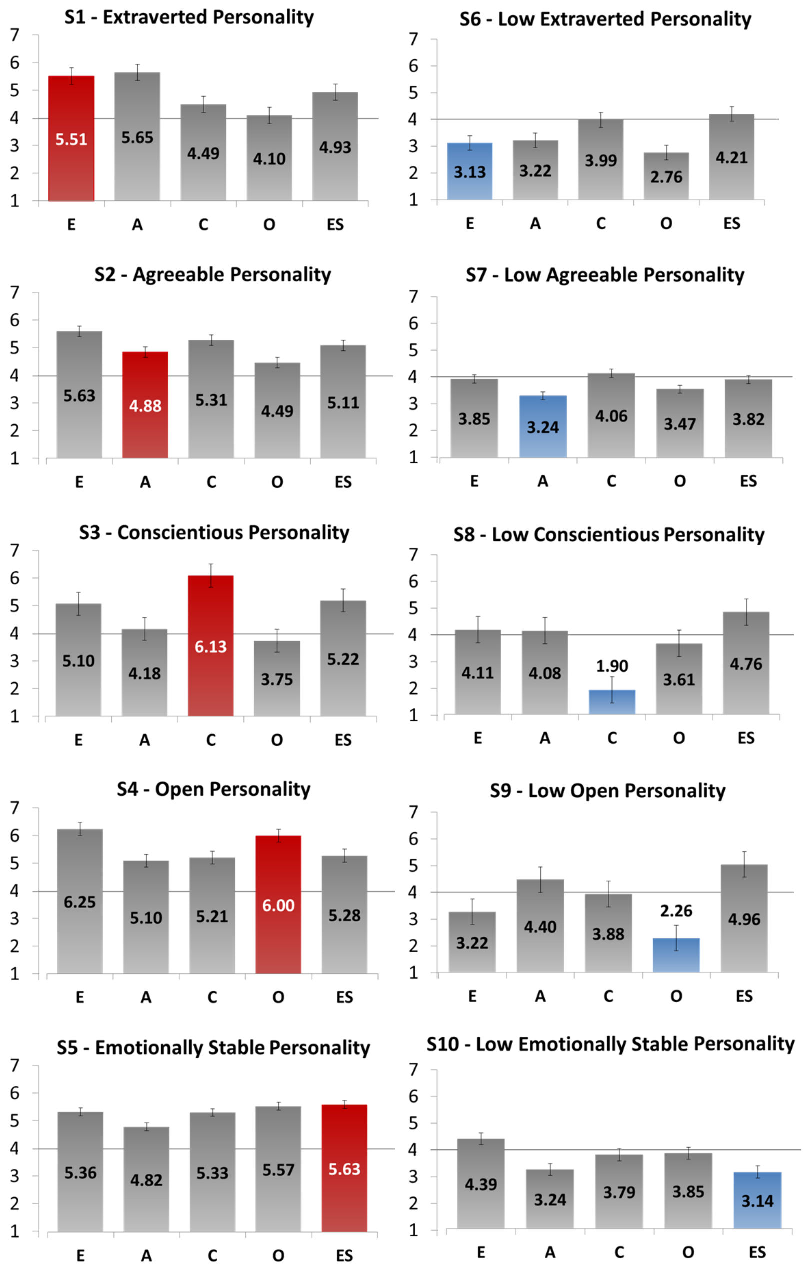


Figure 4.11: Personality scales (TIPI ratings on traits): E - Extraversion, A - Agreeableness, C - Conscientiousness, O - Openness, ES - Emotional Stability, averaged over Render Style and Participant Sex, for the independent variable Situation (10 examples) which portray positive (graphs on the left) and negative (graphs on the right) poles of the Big Five traits. Coloured bars represent the average ratings of the intended trait, for each Situation (e.g., red bar shows an average rating of 5.42 for the extraverted trait in S1 – Situation intended to display an Extraverted Personality). Each bar is labeled with the average mean value and standard error bars.

ness: $F(9,288) = 61.31, p \approx 0$, Emotional Stability: $F(9,288) = 29.62, p \approx 0$). *This shows that the situation affected ratings of personality.*

Figure 4.11 shows the ratings of the Personality scales for each Situation. On observation of this graph and using one-sample t-tests with a constant 4 (middle of the rating scale from 1 – 7), it can be seen that the intended traits were conveyed for each situation above chance (red bars for the situations that expected high ratings on the intended trait scale, and blue for the situations that expected low ratings, $p \approx 0$ in all cases). We also observe that the intended trait was not the only trait that was strongly identified for each Situation (e.g. S1 received high ratings for the intended Extraversion trait, but also for the Agreeableness trait). *This implies that the constructed situations were sufficiently recognised but also elicited other traits that were not specifically intended.*

A main effect of Render Style occurred for the Agreeable scale, where ToonCG was judged as significantly higher on this scale than HumanIll ($F(1,32) = 4.95, p < 0.04$). Means = 4.5 and 4.07, respectively. For this trait, a three way interaction also occurred between Participant Sex, Situation, and Render Style ($F(9,288) = 2.38, p < 0.02$) where female participants rated the HumanIll style as significantly lower on Agreeableness than the ToonCG style in Situation S3 ($p < 0.02$), while there was no significant difference between the styles for Situation S3 for male participants. No other main effect of Render Style or interaction with Render Style was found for the other four traits. These results suggest that *the style used to render a character can alter the interpretation of its personality.*

b) Appeal

Finally, we present the results of the ratings of Appeal (see Figure 4.12). No effect of Render Style or interaction with Situation was found, however, a main effect of Situation was found ($F(9,306) = 15.48, p \approx 0$). Post–hoc analysis showed that each of the five positive situations (S1 – S5) were rated as significantly higher on Appeal than each of the negative situations (S6 – S10) ($p < 0.01$ in all cases). Furthermore, all of the positive situations were rated equally on Appeal and all of the negative situations were rated as equally lower on Appeal, except for Situation S10 which was rated as significantly higher on Appeal than S6 ($p < 0.05$). Our results imply that *the appeal of a scenario is highly dependent on the personality content and the render style does not directly influence the overall appeal.*

Discussion

Our most significant finding is that the perception of personality can be indirectly altered by changing the style used to render the character. In our experiment, the ill-looking render

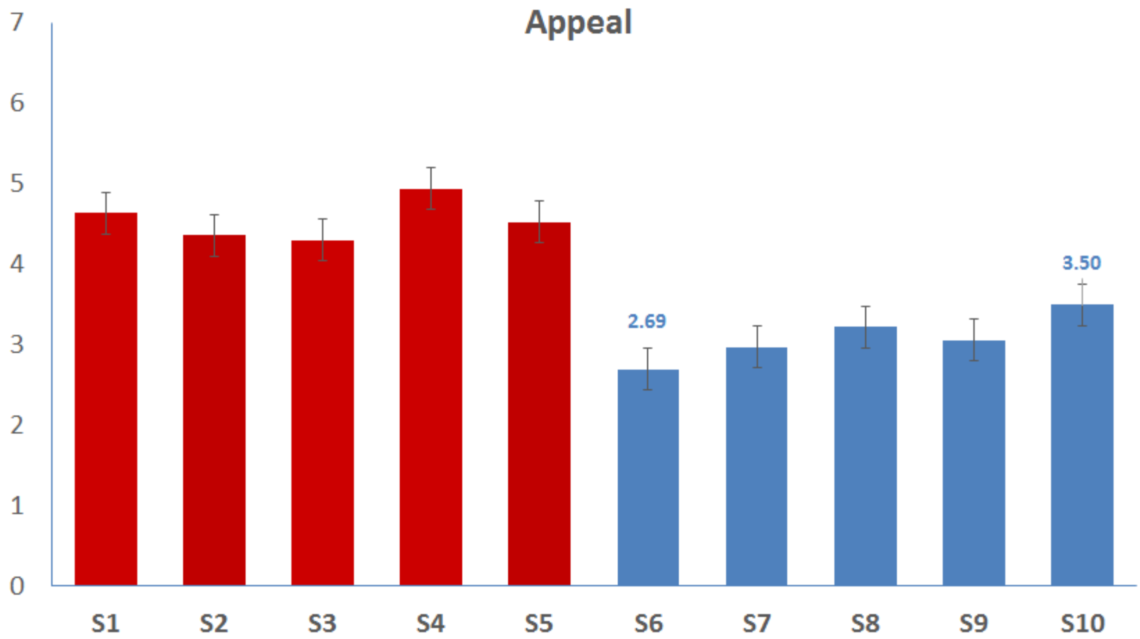


Figure 4.12: Average ratings of Appeal per Situation (S1 – S10). The most notable and significant difference is between the positive (S1 – S5) and negative (S6 – S10) situations. Also, from the negative situations, S6 was rated significantly lower than situation S10.

style evoked less desirable personality traits than the cartoon style. This was apparent in the agreeableness personality trait, where the ill style was judged overall as less agreeable than the cartoon style. This implies that the ill character was seen to be more critical and quarrelsome, as well as less sympathetic and warm than the cartoon.

The appeal of the characters varied more according to personality and not render style, which could be a limitation of our experiment design. Participants were asked to rate appeal after every situation and may have neglected the appearance of the character, since it was kept constant across all situations. Another limitation of our current study is the fact that certain motion was not included, such as eye and finger motion, which could have an effect on the perception of personality. Also, the performance of only one actor was captured, who was let to interpret the reactions to the situations in his own way. In addition, we recognize the importance of sound on the perception of personality. However, since we only had one actor, we cannot account for any changes in the perception to be due to the sound of his voice. Finally, even though all created personalities were recognized, our results show that certain situations conveyed personality traits we did not specifically intend for the character to have.

The results from the screen–based experiment have shown that the constructed situations were able to convey the intended personality traits, especially the negative poles of the Big Five. This helped us to continue with our next experiment, where we presented the personality traits in an interactive environment (Personality Experiment 2).

4.2.2 Personality Experiment 2: Virtual Reality

Based on the results from Personality Experiment 1, render style was shown to bias the perceived personality of the virtual character in such a way, that certain less desirable personality traits are attributed to less appealing render styles, which prompted us to investigate the effect of render style on other scales, such as empathy and proximity behaviour.

Secondly, the previous experiment showed that some personalities we built were more appealing than others. The following experiment uses a between-group experiment design, where each participant sees only one personality (and one example of a render style). This will determine if the differences in appeal were found previously due to the fact that all participants saw all personality situations and were comparing them between each other. Here we are interested to know how would a more (or less) appealing personality further influence peoples' subjective responses and proximity behaviour? Could certain personalities induce more empathy, were more appealing and result in higher co-presence depending on which render style presented them?

And lastly, similar to the Agency experiment, the realistic render style in the following experiment served the investigation of potential effects that realism could have on the perceived personality. This was done to extend the uncanny valley theory by assessing other attributes and effects of realism on people's perception of the character exhibiting personality traits. In order to study this, we not only use measures of empathy, co-presence and proximity, but also add the indirect measure of attribution bias as an indicator of the ease of cognitive processing. This is based on our theoretical model of social perception (see Section 2.3.4) where realistic characters are compared to real humans but are recognised as not real, therefore are processed in a non–automatic way. This should lead to cognitive difficulty ("slow process") and on a response level, the answer would reflect the fundamental attribution error (Ross, 1977). This error is associated with the tendency to associate behaviour to the person and not situational factors which could provoke the person's behaviour. We would therefore like to use the same measure of attribution bias used by Gilbert et al. (1998) to study the potential cognitive busyness while observing virtual characters of different styles.

Stimuli Creation

The stimuli creation and experiment design was similar to the Agency experiment we conducted in virtual reality (Agency Experiment 2), however, there were a few changes.

Figure 4.13: Aesthetic design of the virtual environment intended to provide context to the experience, induce a higher level of immersion and increase motivation to participate in the experiment.

For the stimuli, we used five different render styles, shown on Figure 4.14. We chose these styles to have three examples used in the Agency study, which are examples of appealing render styles while the two obviously eerie render styles served as a reliability measure – we predicted these styles would increase the level of eeriness, lower appeal and increase personal distance between them and participants. The reason behind using intentionally creepy render styles was to induce a more extreme reaction from the participant, since our previous Agency experiment did not show a strong behavioural response to characters of different render style. Apart from affecting the subjective measures, we anticipated that intentionally eerie render styles would increase the viewing distance for the label task as opposed to more reassuring styles.

From the previous experiment (Personality Experiment 1), we used only a subset of personality situations, which were more suitable for the investigation of empathy and social cognition (attribution bias, proximity and co-presence). Therefore, Open and Conscientious personality along with their polar opposites (See Appendix for descriptions) were not included in this experiment, since the character is mainly talking about his preferences and who he is, which would make assessment on some scales in this experiment difficult or even invalid. The chosen set of 6 personalities – Agreeable/Non Agreeable, Extraverted/Introverted and Emotionally Stable/Neurotic represented suitable situations as well as enough variation to our stimuli. We also renamed the negative poles (Low Agreeable Personality into *Non Agreeable*, Low Extraverted Personality into *Introverted*, and Low Emotionally Stable Personality into

Figure 4.14: Render styles of the character used in the experiment (from *left* to *right*): Realistic, Toon CG, Toon Shaded, Zombie and Creepy render style. The last two render styles were created purposefully to induce a feeling of eeriness in participants.

Neurotic) for practical purposes since the clear distinction that we were expecting low ratings on a particular TIPI scale for a portrayed personality was not needed here.

While we used the same environment setup for the virtual reality as described in Section 4.1.2, a noticeable addition was sound and aesthetic design of the environment (see Figure 4.13) which created a pleasant atmosphere since we also wanted to create an experience which would be interesting to the pool of participant who were visitors to a gallery exhibition (Science Gallery Dublin). For this experiment, we also included facial animation which was shown to have an effect on the character's pleasantness (Carter et al., 2013).

A notable difference was also the fact that the experiment was programmed to run automatically from start to finish, with all instructions of the tasks being presented in the environment as part of the experience. This was a demanding task which required extensive pre–testing to ensure the participants will not be stuck in one area and would not know how to progress with the tasks. Some guides were implemented as colliders which prevented the motion of the participant to continue, e.g. through the window. The presentation of tasks however, was implemented in a form of pop–up information text boards, which triggered when the participant came close to a blue information symbol in the environment. All data was saved at the end of the experiment, therefore the participant had to finish all tasks in order for the data to be collected. This avoided cluttering of redundant data.

Experiment Design

We used the same behavioural measure (proximity) which was measured when the participant left the red spot to find the word tag near the character. Similar to before, the questionnaire was put on a virtual board in the environment. The questions (apart from the first one) were rated on a 5–point Likert scale, ranging from *1– Not at all* to *5–Extremely*.

The first question was designed to measure attribution bias, which requires the participant to make a decision about the locus of the character's reaction to the situation. We constructed this question similarly to the study of Gilbert et al. (1988) after observing the character portraying a personality. We posed the following question:

- "You witnessed the character reacting to a situation. In your opinion, what is the MAIN cause of the character's behaviour? (1 – character's unique way of behaving; 2 – situation made the character behave in this way):

1	2
Character	Situation

The following three questions were related to the measures of empathy – Concerned, Excited and Uneasy, the selection of which is explained in section 4.0.2. The questions which follow are the uncanny valley questions (Affinity and Realism groups, see Table 4.2) in the Agency experiment. The co-presence question was reduced to only one statement ("I perceived that I was in the presence of another person in the virtual room with me.") since the Agency experiment showed little indication on the three measures of co-presence but we wanted to have at least one example to compare to.

This experiment was featured at the Science Gallery Dublin under the title *Virtual Empathy* as part of the Seeing exhibition taking place from 24th June 2016 until 25th September 2016. The exhibit included a monitor, Xbox controller, headphones, Oculus DK2, and a rotating chair (sample picture of the setup is shown on Figure 4.15). Visitors could participate in the experiment voluntarily. They were introduced to the task with the help of the mediator employed by the gallery who was trained to follow the experiment protocol. The protocol included important points about hardware maintenance, security issues, and instructions regarding working with participants. For example, no children under 13 years of age were allowed, participants over 13 but under 18 years of age were informed their data would not be collected or included in the analysis, visitors who had trouble understanding English language needed to have a translator; if no translator was present, they could not continue with the experiment. In case of nausea, the experiment had to be terminated and the participants remained seated for a couple of minutes and were given a glass of water.

Due to an overwhelming response from the visitors of the Science Gallery, we were able to collect data from over 2000 participants, from which 1106 (429 females, 677 males; average

Figure 4.15: The setup at the Science Gallery exhibition titled *Seeing*. Participants for this experiment were visitors of the gallery who volunteered to participate.

age: 28.7) were included in the analysis. The other participants were excluded due to failing to fill out the demographics, or have taken the experiment before. We also programmed the experiment in such a way, that the subjective responses from the participants were saved only if they finished the whole experiment. This way, if someone decided to finish the experiment before, their data was not included in the analysis.

The experiment had three stages: *waiting room*, where the participant would get used to the environment and controls. They were also introduced to the guidance through the environment by triggering information boxes (see Figure 4.16) and it was also where they first encountered the moving virtual character, which was purposefully created to have the highest level of visual realism. The reasoning behind exposing a realistic render style at the beginning was for the participants to formulate which level of realism was possible to achieve in this particular environment. This way, if they encountered a less realistic character further on in the experiment, they would not assume it is not possible to exceed the level of realism they were presented with. The next stage was *training room*, where the participants were observing a box which was changing colour for 5 seconds. After this time, the label appeared on the box to train people on how to execute the label reading task. The last stage was the actual *experiment room*, where we presented the animated character and collected participants' responses to him.

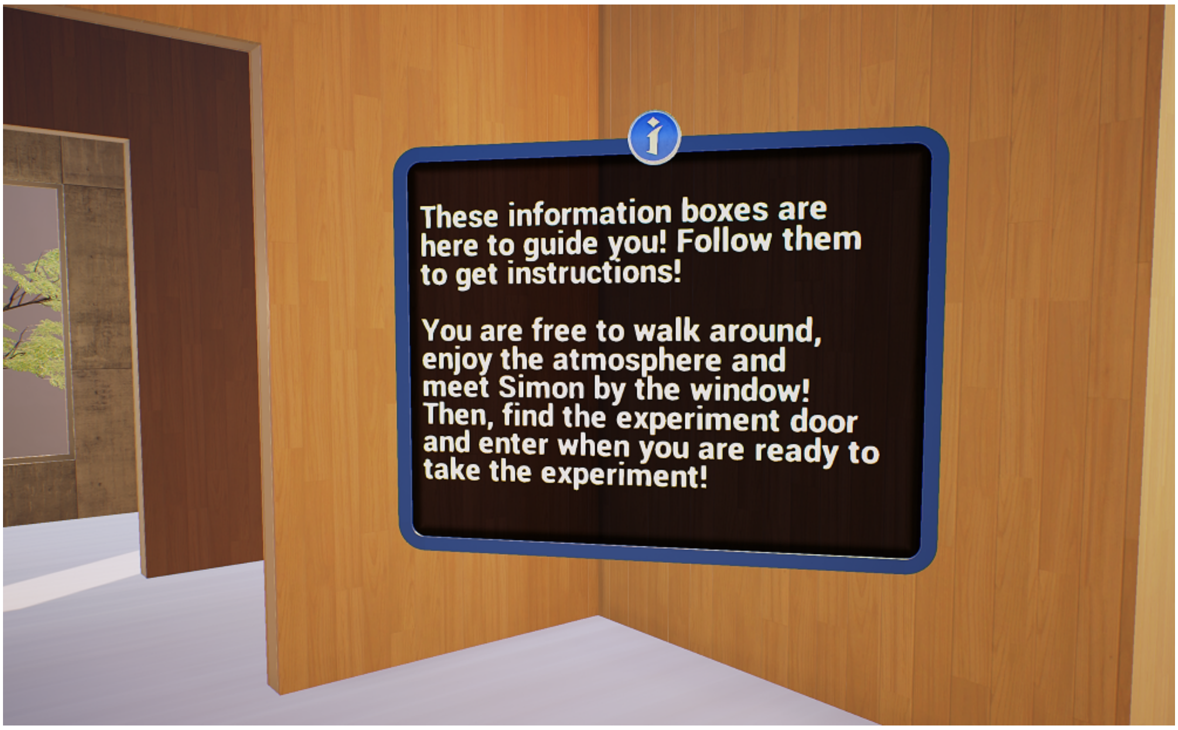


Figure 4.16: An example of the information box with guidelines for the participant. The box was presented as a blue dot in the environment and the text box only expanded when participant moved closer to it. This interaction was created so the boxes would not occlude the environment when in the extended mode.

The participants were asked to sit in front of the monitor where they were first presented with an electronic consent form and information sheet, approved by the University department ethics committee. The participants were asked to fill out the demographics: age, sex, game experience, experience with virtual characters, native English speaker or not and a question if the have taken the experiment before. With the help of the mediator, they were given the headset and a set of headphones to put on their head.

They were first presented with the waiting room where they learned how to use the controls and move around the environment. The instructions were given to them in the form of pop-up windows which guided them through the environment. The first task was to go closer to the virtual character which was standing near the window. When they got close enough to the character, they unlocked the next level which was accessible to the door with a sign “Experiment Room”. The door opened when they navigated close to it and they were teleported to the next level, where they were presented with two rooms – training and experiment. The experiment room only unlocked when the training room was completed. The instructions guided the participant to the training room with the box and when they reached the red spot they were locked in position and were instructed to observe the box. The box was changing colours for a dedicated amount of time, after which instructions to go closer and find the word tag appeared. When moving closer and away from the box, their position trajectory was recorded. After exiting the room, a multiple choice questionnaire appeared on the virtual board in front of them, asking them to choose the word tag they found in the room. After, the experiment room unlocked and they were able to walk towards the red spot again, this time locking while observing the character in a randomly assigned render style, expressing a particular personality, also randomly chosen from a set of 6 personalities (due to the between–groups design, where each participant saw only one render style of a character, exhibiting one personality). After the animation stopped playing, they had the same task to memorise the word tag and exit the room. Upon exiting, the questionnaire board appeared with the empathy, uncanny valley and co-presence questions. After they answered all the questions, the experiment self-terminated.

Results

Even though the experiment design was similar to the one in the study of Agency (see Section 4.1.2), the main difference is that each participant saw only one example of the render style and one personality expression. This design was a lot more complex than the one in the Agency experiment, therefore we used different approaches for the analysis of our data, based on our predictions where we expected differences to arise. For clarity, we present the results in separate sections.

Section a) compares the Realistic render style pairwise to others, therefore creating a 2×6 factor design (2 *Render Style*, 6 *Personality*) which we analysed for 4 different paired comparisons (Realistic – Toon CG, Realistic – Toon Shaded, Realistic – Creepy, Realistic – Zombie). The reasoning behind such an approach was to explore differences specifically between realistic as opposed to non–realistic styles. Our dependent variables were grouped in categories Empathy (*Concerned, Excited, Uneasy*), Affinity (*Eerie, Familiar, Appealing*) and Realism (*Appearance Realism, Movement Realism and Overall Realism*). Others which were not grouped included *Co-presence*, *Attribution* and *Proximity.*

Section b) where we analyse how the type of exhibited personality affected our data. We were not specifically interested in Personality effects, however we report them to aid the discussion. We are, however, interested in interactions between Personality and Render Style by looking into individual personalities and how different render styles might effect the dependent variables.

Section c) provides the analysis of results on other comparisons, which were also not part of our main interest but could have significant differences. Therefore, we report overall main and interaction effects between all styles and personalities. These effects are reported to evaluate the strength of overall Personality and Render Style on the dependent variables.

Section d) reports the results of the dependent variable Attribution. Here, we analyse significant differences in how participants perceived the behaviour of the character – as coming from his own personality or as a result of the situation.

Section e) provides additional information on how some aspects of participants could influence our data. We did not include participant sex or other demographic information into the main analysis since we had a large number of variables and potential effects. Therefore, we analyse potential effects of demographic variables by using generalised linear models to discuss possible effects on the results we have collected in our study.

a) The effect of Realistic render style

We conducted a two-way ANOVA (ANalysis of VAriance) to assess the main and interaction effects of *Render Style* (2 examples) and *Personality* (6 examples) on the participants' responses. ANOVA is considered robust for the deviations from normality. However, we used Levene's test to check for homogeneity of variance and for variables, where the assumption was breached, we did a log transformation on the data and repeated the ANOVA process on the transformed data. Tukey Honestly Significant Difference (HSD) test was used for post–hoc comparisons. For the analysis of the effects of *Render Style* inside each *Personality*, we conducted a One-Way ANOVA separately for the 6 personalities.

Correlations between Proximity and questionnaire items were calculated using Sperman's Rank Order test, since we found *Proximity* variable was not normally distributed (Kolmogorow-Smirnov test). Since none of the correlations were found to be significant, this analysis is not discussed further.

Realistic vs. Toon CG comparison When analysing the main effects for the Realism variable group, we confirm that the Realistic render style was considered significantly more realistic in appearance than the Toon CG render style (Appearance Realism: $F(1, 437) = 4.94$, $p = 0.027$), as expected (see Figure 4.17, *left*). However, the difference in the Overall Realism did not appear to be significant, even though it was close to being that way. No differences were observed for the Movement Realism. No interactions between Render Style and Personality were found.

In the Affinity variable group, no main effects for the independent variables were found, but there was an interaction for the Appeal variable (Appeal: $F(5, 437) = 3.38$, $p = 0.005$). Neurotic Realistic character was found to be more appealing than Neurotic Toon CG character ($p < 0.05$). However, under closer inspection, Levene's test showed non–homogenous variance distribution and under logarithmic transformed data, the interaction remained significant

Table 4.5: Descriptive statistics for minimal distances in Proximity task according to the Render Style: Means, standard deviations (SD), and standard error (SE).

	N	Mean (cm)	SD	SE
Realistic	222	162.49	94.90	6.37
Toon Shaded	227	181.09	99.73	6.62
Toon CG	227	185.66	107.20	7.12
Creepy	212	180.87	102.38	7.03
Zombie	218	157.32	91.76	6.22

(logarithmic transformation applied: $F(5, 437) = 2.49$, $p = 0.031$) but the post–hoc did not detect significant differences between the Neurotic characters anymore.

No main effects or interactions were found for Empathy variables when comparing Toon CG and Realistic render style.

We did get an interaction for the Co-presence question. Extraverted Toon CG character was perceived to induce a higher level of co-presence than the Extraverted Realistic character ($F(5, 437) = 4.96$, $p \approx 0$, post–hoc: $p < 0.003$).

The behavioural measure Proximity showed that participants moved significantly closer to the Realistic than Toon CG render style (Proximity: $F(1, 437) = 5.71$, $p = 0.017$) (see also descriptive statistics for Proximity varaible, Table 4.5).

Realistic vs. Toon Shaded comparison The main effects for the Realism variable group confirm again that the Realistic render style was considered significantly more realistic in appearance than the Toon Shaded render style (Appearance Realism: $F(1, 437) = 6.15$, $p = 0.013$). The difference was also significant for the Overall Realism ($F(1, 437) = 9.82$, $p = 0.002$) while the Movement Realism did not appear to be significant, as expected. No interactions between Render Style and Personality were found (see Figure 4.18).

In the Empathy variable group, for the variable Concerned, we found an interaction between Personality and Render Style. Participants' concern for the character changed according to Render Style but the post–hoc did not reveal specific differences. Observation of the graphs in Figure 4.20 shows concern for the character varied greatly according to the personality type, participants were more concerned for the Realistic character when he was portraying Neurotic and Agreeable personality, whereas Toon Shaded had higher concern ratings for the Non Agreeable personality. All the other personality types have similar level of concern according to the Render Style.

Figure 4.17: Main effects of Render Style and interactions with Personality for comparisons between Realistic : Toon CG (*left*)and Realistic : Zombie (*right*). Significant differences found by the post–hoc test are marked as means on individual bars.

A significant effect for Eerie in the Affinity variable group was found, where the Realistic render style scored significantly higher than the Toon Shaded (logarithmic transformation applied: $F(1, 437) = 7.77$, $p = 0.006$).

An interaction was found for Co-presence (Co-presence: $F(5, 437) = 2.80$, $p = 0.017$) but post-hoc could not find specific significant differences. There is a trend, however - Realistic render style induced more co-presence when on personalities which belong to the negative pole of the Big Five (Introverted, Neurotic, Non Agreeable) whereas Toon Shaded render style induced more co-presence when portraying positive personalities (Extraverted, Aggreable, Emotionally Stable).

No main effects or interactions were found for Proximity.

Realistic vs. Zombie comparison Again, in the Realism variable group, Realistic render style was rated significantly higher on Appearance Realism ($F(1, 428) = 6.32$, $p = 0.012$) and Overall Realism ($F(1, 436) = 6.52$, $p = 0.011$) with no interaction effects (see Figure 4.17, *right*) .

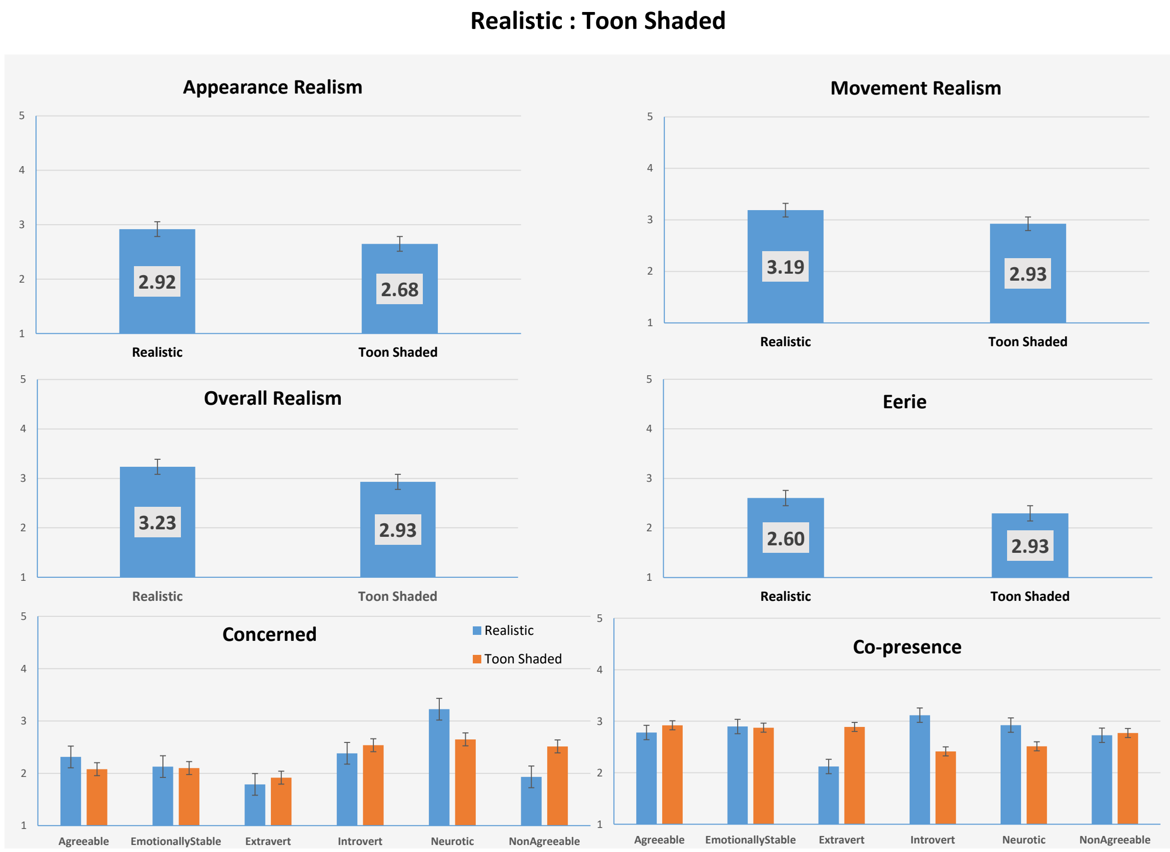


Figure 4.18: Main effects of Render Style and interactions with Personality for comparisons between Realistic : Toon Shaded. Significant differences found by the post–hoc test are marked as means on individual bars.

In the Affinity variable group, an interaction between Render Style and Personality was found for Appeal ($F(5, 428) = 2.60$, $p = 0.025$) and Eerie ($F(5, 428) = 3.38$, $p = 0.005$). However, post–hoc did not discover any significant difference between the characters expressing the same personality. On observation (see Figure 4.21), Realistic was rated more appealing than Zombie render style when portraying Agreeable and Neurotic Personality, and Neurotic and Introvert Realistic characters were considered more eerie than the same personalities seen on the Zombie render style.

No other effects or interactions (Empathy, Co-presence, Proximity) were found when comparing the Realistic and Zombie characters.

Realistic vs. Creepy comparison No significant main effects and interactions were found on any dependent variable for the two render styles.

Discussion on Section a) Results showed, that all the render styles, apart from Creepy, were perceived as less realistic than the Realistic one, therefore the design of the latter was appropriately conveying a more realistic appearance. The lack of difference with Creepy in realism and all other scales also shows that all the differences we got between Realistic character and other styles on dependent variables could be due to changes in realism. However, the lack of differences with the Creepy character also means that the excessive details on the Realistic character (subsurface scattering on the skin) did not add significantly to the realism of the character and the enlarged eyes of Creepy character, intended to induce a more eerie effect, did not skew the realism component. This indicates that, even though different, Realistic render style did not reach a high level of perceived realism, and indeed the mean values for appearance realism are relatively low, seldom going over middle value "3". We conclude that *the Realistic render style was accurately perceived as having a higher level of realism in comparison to other styles*, however perceived realism was low for all styles and *smaller changes to the materials did not add sufficiently to the perceived realism* of the Realistic style, signaled by the lack of difference compared to the Creepy character.

Other results signify that render style makes a difference in combination with certain personalities. For example, the neurotic personality appears to be more appealing on Realistic character compared to some less realistic styles (Toon CG, others styles show a tendency towards the same relationship), however it also seemed to be more eerie (Toon Shaded, others styles show a tendency towards the same relationship as well). This could signify that the eeriness did not necessarily induce a negative response from the participants, it could also have appeared interesting, signified also by the tendency to have a higher level of concern for the Realistic Neurotic character. Similarly, some personalities portrayed by the Realistic character induced more co-presence, especially on negative poles of the personality traits (Introverted, Neurotic, Non Agreeable), and Toon CG induced more co-presence when exhibiting one of the positive poles of personality (Extraverted). Again, the average co-presence was very low, therefore making strong conclusions about co-presence would not be valid. The next results focusing on main effects of personality and interactions with all the render styles (not only two by two comparisons) could help us understand the relationship better. Overall, we can conclude, that *the effect of the Realistic style on participant's responses differs according to the portrayed personality.*

The behavioural measure did not show many differences according to render style, except when compared to Toon CG character. Our research showed that participants moved closer to the Realistic character than Toon CG. This result is difficult to explain based on other results. There was no significant difference in appeal or co-presence between the styles, except when paired with a certain personality, as mentioned. We conclude *Realistic render style had a small influence on the proximity behaviour.*

b) The effect of Personality

We analyse the effects of the personalities further by including overall effect across all render styles to see how the personality type the character was portraying affected the dependent variables. We conducted 5×6 (5 *Render Style*, 6 *Personality*) two–way ANOVA which showed Personality had an effect on all dependent variables, especially for the ones belonging to the Empathy group (Concerned: $F(5, 1076) = 21.85, p \approx 0$; Excited: $F(5, 1076) = 9.77, p \approx 0$; Uneasy: $F(5, 1076) = 4.38, p = 0.001$). Other effects included: Appeal ($F(5, 1076) = 3.30, p = 0.006$), Appearance Realism ($F(5, 1076) = 2.46, p = 0.032$), Movement Realism ($F(5, 1076) = 6.91, p \approx 0$), Overall Realism ($F(5, 1076) = 3.21, p = 0.007$), Eerie ($F(5, 1076) = 4.24, p = 0.001$). Table 4.6 shows individual comparisons and significant differences. From the main effects Figure 4.19, we see the negative poles of the personality were affected mostly on the Empathy group. The Neurotic personality raised the highest levels of concern and participants felt most uneasy observing this personality as well but only compared with Emotionally Stable and Extravert. Introvert personality induced more concern compared to all personalities except Non Agreeable and Neurotic personality. Non Agreeable and Introvert also induced the least excitement in participants. In the Affinity group, Introvert stands out as least appealing (significantly different only when compared to Emotionally Stable and Extravert) and Extravert was least eerie, compared to negative poles of the personality. Main effects of Personality were found also for perceived realism where Emotionally Stable was rated the highest on Movement Realism (significantly more compared to Introvert and Extravert). Introvert was perceived to have the lowest Movement Realism and Overall Realism (compared to Agreeable and Emotionally Stable).

Table 4.6: Tukey test for individual Personality examples on the dependent variables for which effects were found. The numbers present the exact p-values.

Variable	Personality	Post-hoc (Tukey's Test), *p* values: *Agreeable*	*Emotionally Stable*	*Extravert*	*Introvert*	*Neurotic*	*Non Agreeable*
Concerned	*Introvert*	0.036	0.025	0.000	X	0.000	X
	Neurotic	0.000	0.000	0.000	0.000	X	0.000
Excited	*Introvert*	0.000	0.034	0.000	X	0.011	X
	Non Agreeable	0.007	X	0.000	X	X	X
Uneasy	*Neurotic*	X	0.003	0.002	X	X	X
Appeal	*Introvert*	X	0.014	0.012	X	X	X
Eerie	*Extravert*	X	X	X	0.015	0.004	0.006
Movement Realism	*Emotionally Stable*	X	X	0.009	0.000	X	0.050
	Introvert	0.000	0.000	X	X	X	X
Overall Realism	*Introvert*	0.044	0.009	X	X	X	X

Since the exhibited Personality has such a strong effect on our data, it makes sense to look at the effect of Render Style again, but individually inside each Personality. We did so

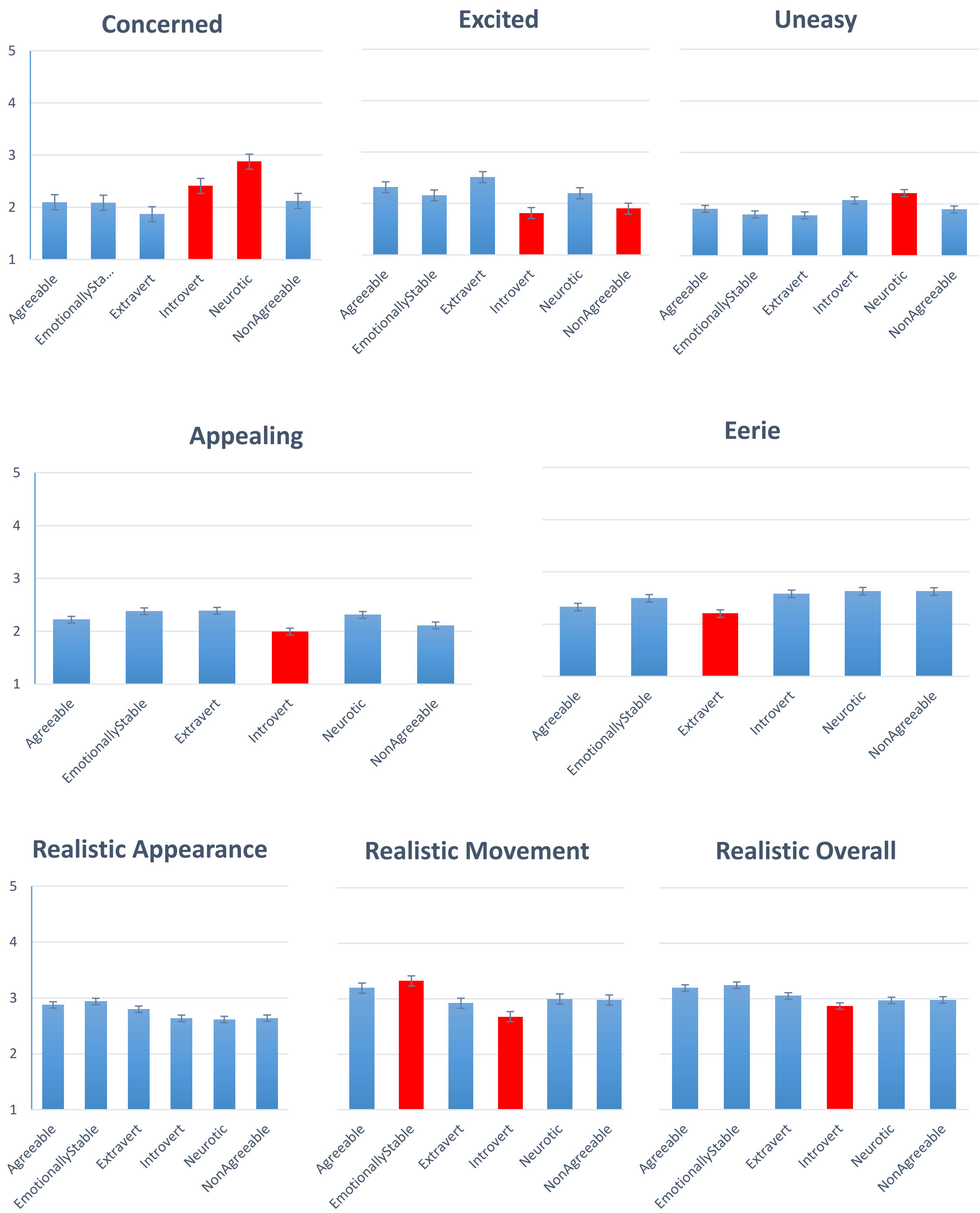


Figure 4.19: Main effects for Personality. While no effects for Familiarity, Proximity and Co-presence were found, other dependent variables were affected. For clarity, red coloured bars label significant difference with at least one more Personality type. Individual comparisons between Personality types are found in the post-hoc table.

by using one-way ANOVA for each Personality across Render Style. The labeled coloured means on interaction graphs (see Figures 4.20 to 4.23) show significant effects.

Agreeable: differences between Render Styles were found for the Concerned variable ($F(4,178) = 2.72, p = 0.032$), Eerie ($F(4,178) = 3.39, p = 0.011$) and Proximity ($F(4,178) = 3.09, p = 0.017$). A closer inspection with post–hoc showed that, even though we did not find a main effect on Uneasy, participants reported higher ratings on this scale for the Zombie character ($p = 0.020$), Zombie was rated higher on Eerie compared to Toon CG ($p = 0.008$) and Creepy character ($p = 0.026$). Interestingly, Proximity analysis showed that participants came closer to Zombie render style than any other style but significantly closer than Toon CG render style ($p = 0.007$).

Extraverted: this Personality portrayed by different render styles was perceived differently as well, where for Toon CG, participants expressed more Co-presence ($F(4,178) = 4.73, p = 0.001$) than Realistic and Zombie character ($p < 0.03$ for both).

Neurotic: Realistic render style portraying this personality raised higher scores on the Concerned variable ($F(4,180) = 2.51, p = 0.044$) and significantly more so than Toon CG ($p = 0.042$). Also, for the variable Appeal, differences were found ($F(4,180) = 2.68, p = 0.033$) where as before Realistic was found more appealing than Toon CG ($p < 0.015$).

Introverted: this Personality was affected by Render Style ($F(4,191) = 3.61, p = 0.007$) for the Familiar, especially for Toon Shaded, which was considered less familiar than Realistic and Toon CG ($p < 0.03$). Eerie was affected as well ($F(4,191) = 3.40, p = 0.010$), where Realistic was rated more eerie than Toon Shaded and Zombie ($p < 0.05$).

Discussion on Section b) Firstly, we see that the personality the character was exhibiting had an influence on our results, which was to be expected. The negative pole of the personalities (Introverted, Neurotic, Non Agreeable) induced more extreme scores for empathy, higher concern and unease, but also less excitement. Appeal was more positive also for positive pole and eeriness was higher for the negative pole. As expected, some personalities were rated lower on motion realism, since they were presented by different motion cues. It could also simply be that characters that were moving less were not rated highly on realism, which is shown by the higher perceived movement realism for emotionally stable personality in comparison to the introverted one. For the latter, we instructed the actor to have restricted movement which could have been perceived by the participant as less realistic. The only variable which was not affected by personality was the proximity measure. *The character's personality therefore contributed to the variety of responses on empathy, affinity and even realism, but did not affect perceiver's behaviour.*

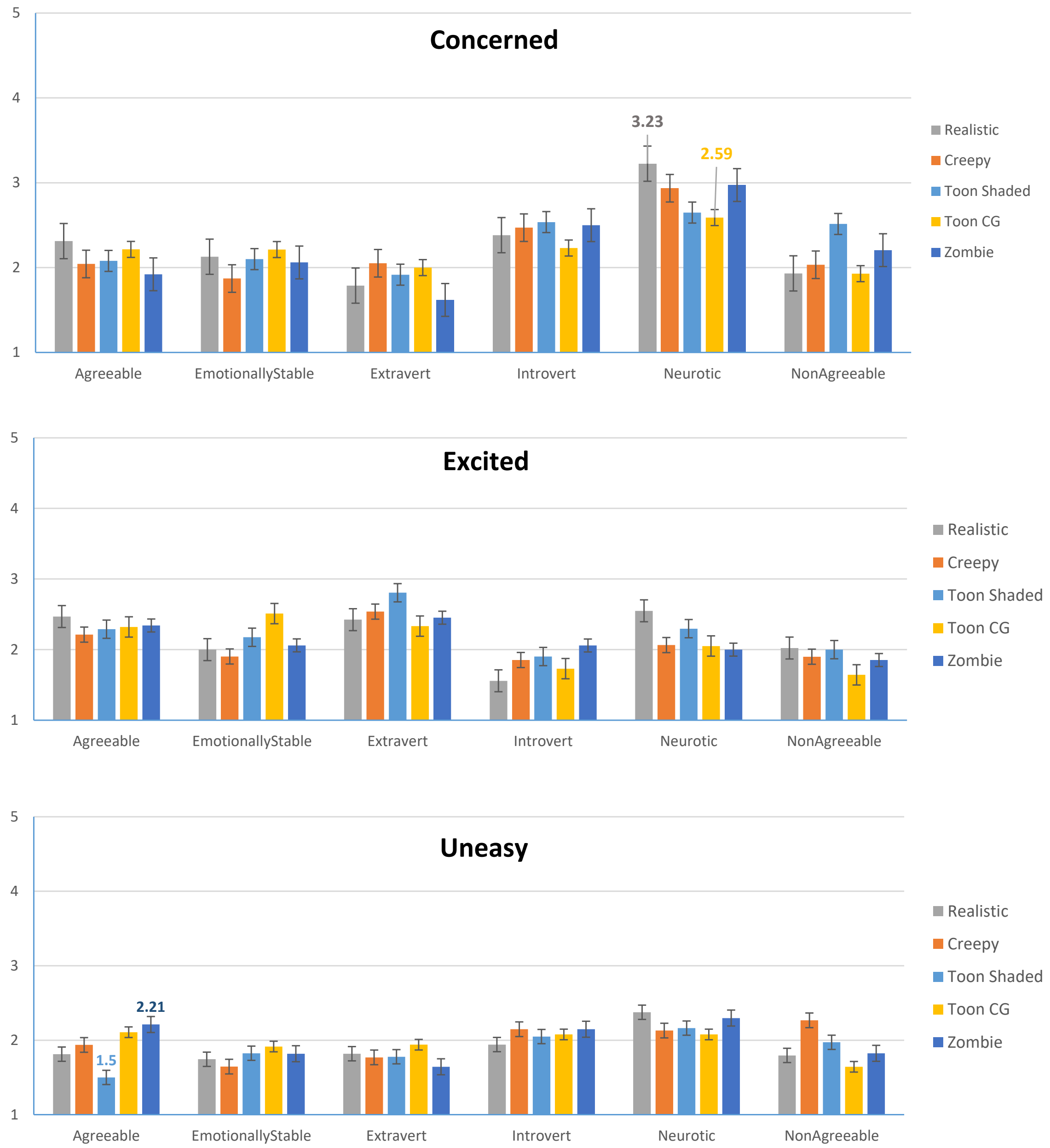


Figure 4.20: Mean scores for different render styles according to Personality for the items in the Empathy variable group. Labeled means present differences between means which were found to be significant within Personality. Main effects and interactions are discussed in the text.

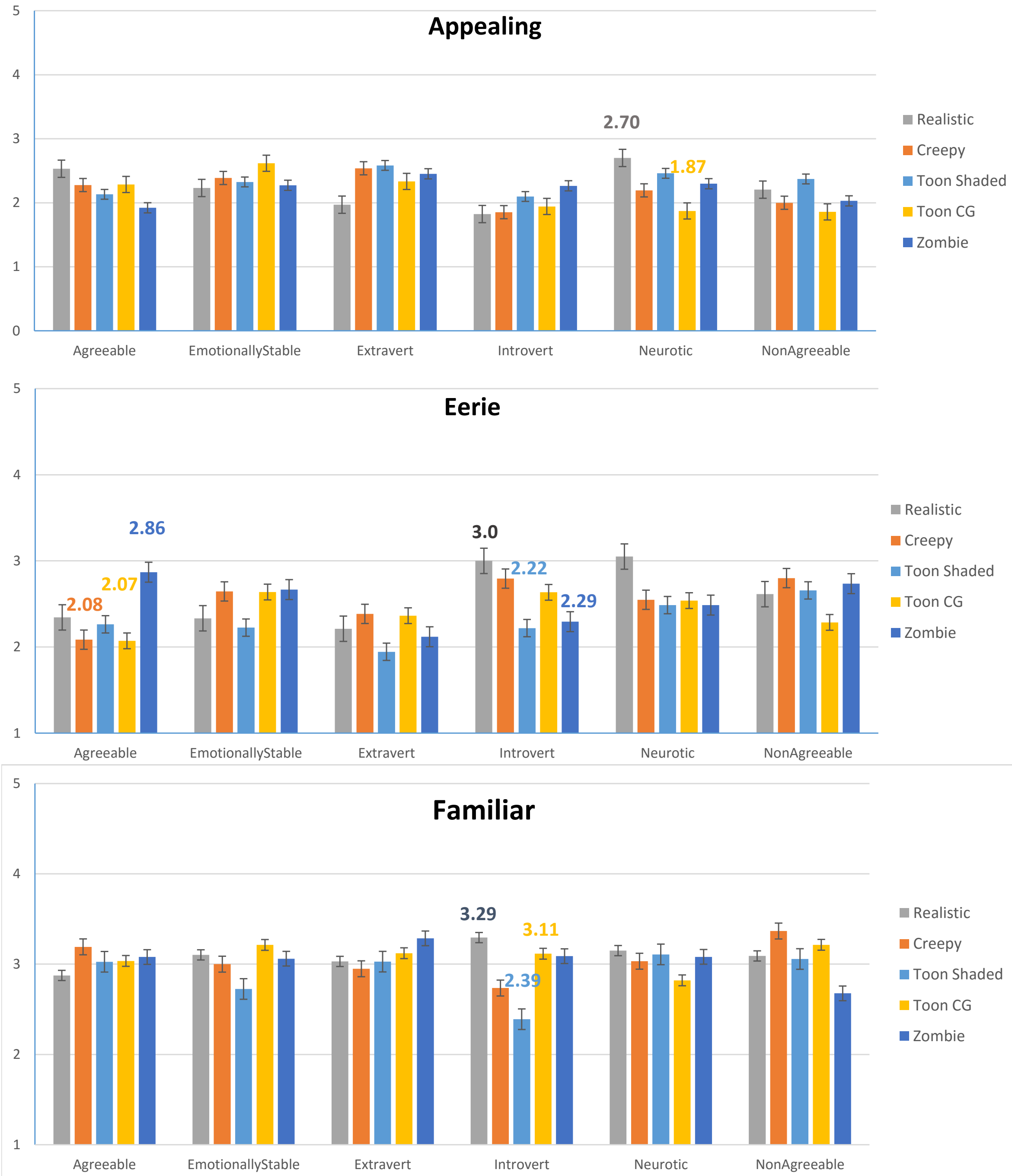


Figure 4.21: Mean scores for different render styles according to Personality for the items in the Affinity variable group. Labeled means present differences between means which were found to be significant within Personality. Main effects and interactions are discussed in the text.

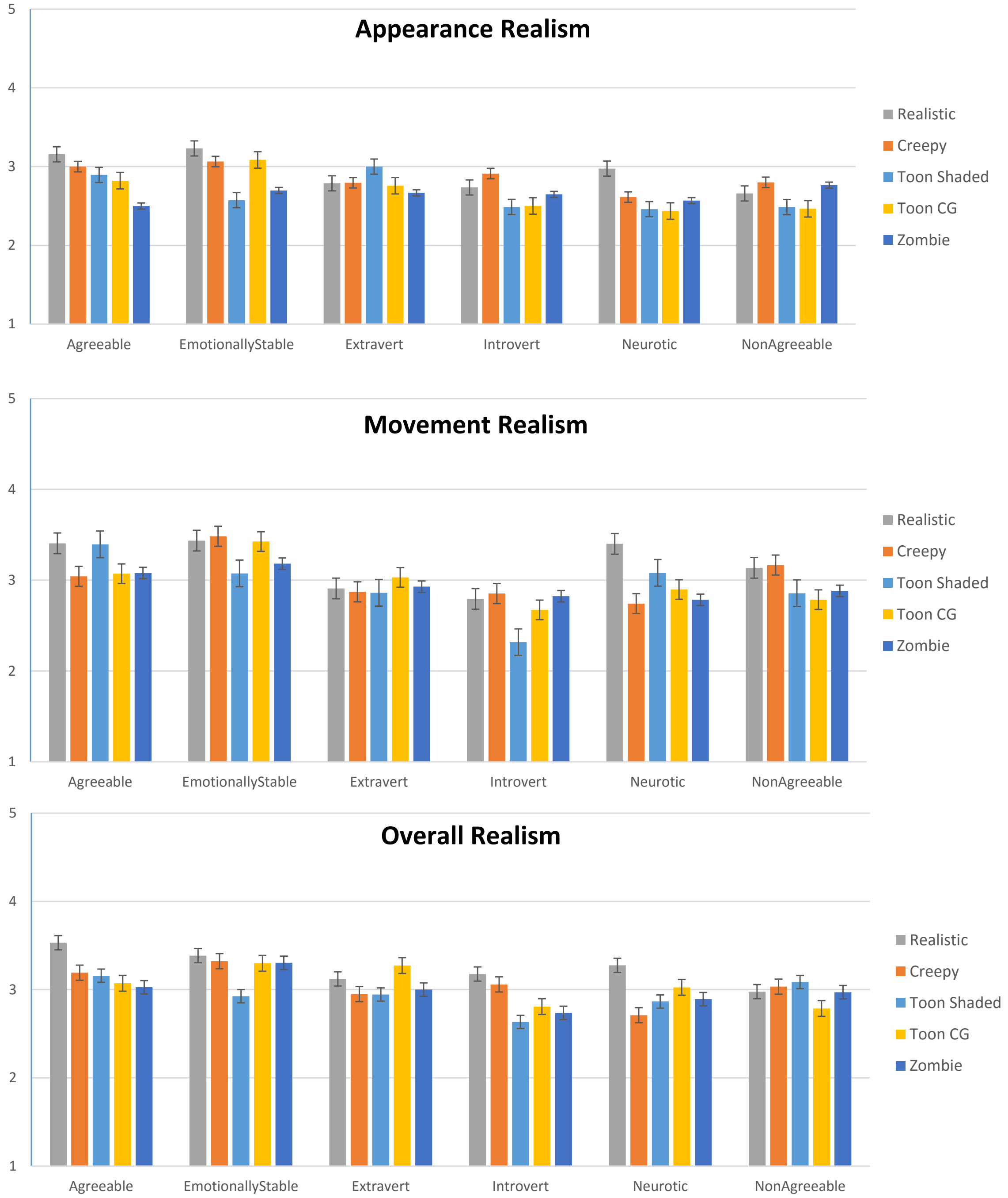


Figure 4.22: Mean scores for different render styles according to Personality for the items in the Realism variable group. No differences between means within Personality were found. Main effects and interactions are discussed in the text.

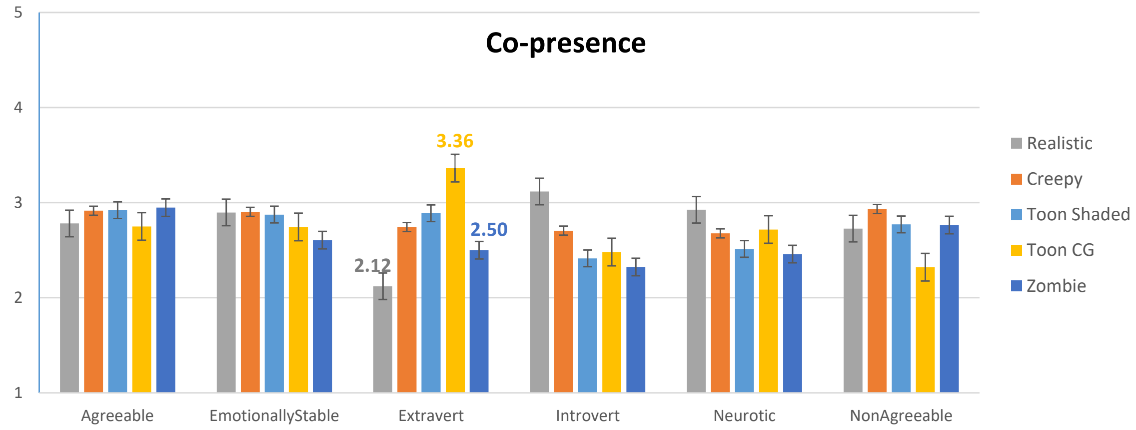


Figure 4.23: Mean scores for different render styles according to Personality for Co-presence. Labeled means present differences between means which were found to be significant within Personality. Main effects and interactions are discussed in the text.

By looking inside each personality and seeing the effect the render styles had on the participants' response, we get some additional explanation to the results we made in Section a). The results indicate that different render styles interact with the exhibited personality, which notably changes the response to that character. For example, the concern for a Realistic render style was high when he was portraying neurotic personality, but not when expressing non agreeable personality. Toon CG render style, on the other hand, improved co-presence when expressing extraverted personality (in comparison to Realistic) and was less eerie in the agreeable one (more so than the Zombie render style). However, as with Realistic render style, this was not a pattern which repeated for every exhibited personality. The reasons behind this are difficult to determine from our experiment. For example, we see that the Realistic style is a more appropriate choice if the intention is to reach a higher level of concern for the neurotic situation but the style will make no difference if the character is expressing non agreeable traits. We could also interpret that since extraverted personality was rated the least eerie personality overall, the increased co-presence when Toon CG render style was expressing this personality could indicate this combination was a good design choice. Similarly to the conclusion made for the Realistic render style in Section a), we conclude that *render style changes the perception of the personality that the character is exhibiting.*

Overall, proximity behaviour was not affected by personality. This is surprising, since research conducted with this measure in real life situations with humans found a difference in how people approach others according to how socially attractive the other is (McCroskey & McCain, 1974). It is possible our situations which exhibit certain personalities did not vary enough in appeal and eeriness to really show responses on a behavioural level. Furthermore, the only significant effect for proximity we found for the agreeable personality, was conflicting. When paired with a character's appearance, the most eerie combination of style and personality as rated by the participants (Agreeable Zombie character) was also the character participants came closest to. This could mean that participants came closer to the character because of the fact it was eerie, possibly out of interest to examine mistakes. *Proximity is slightly affected by the render style and personality combinations which might indicate interest to observe a particularly eerie character.*

Analysing results inside each personality can have negative effects as well, since lower number of cases for each variable can inflate the possibility of false positives. We continue now with the analysis over all render styles to examine which effects were strong enough to be detected over all personalities and render styles.

c) Overall Analysis

To explore the most robust results, we conducted the analysis across all render styles and personalities for dependent variables, therefore 5×6 two–way ANOVAs were conducted on the results and main effects of *Render Style* and interactions with *Personality* were analysed. Not many dependent variables were affected this way. We found expected main effect for Appearance Realism ($F(4, 1076) = 2.81, p = 0.024$) and Overall Realism ($F(4, 1076) = 2.55, p = 0.038$), as shown in Figure 4.24. Post–hoc revealed however, that the Realistic was significantly more realistic overall only when compared to Toon Shaded ($p = 0.025$), where for Appearance Realism there were no significant post–hoc values.

There were also two significant interactions for variables Eerie ($F(20, 1076) = 1.78, p = 0.020$) and Co-presence ($F(20, 1076) = 1.65, p = 0.035$). As already mentioned, co-presence was different between Extraverted Toon CG character and Extraverted Realistic character, and this effect was detected in the current post–hoc analysis as well ($p = 0.019$). As for eeriness, no specific significant differences were found.

Interestingly, Render Style had a main effect on the Proximity variable ($F(4, 1076) = 3.266, p = 0.011$), where post–hoc showed participants went more close to the Zombie but only when compared to Toon CG ($p = 0.022$) (Figure 4.24).

We also conducted non–parametric correlations between Proximity and other dependent variables (questionnaire items) using Spearman's Rank Test, but found no significant correlations.

Discussion to Section c) The results concerning realism confirm our previous concern that the Realistic render style was not considered highly realistic, since the significant difference in overall realism was only in comparison to the Toon Shaded render style. A strong

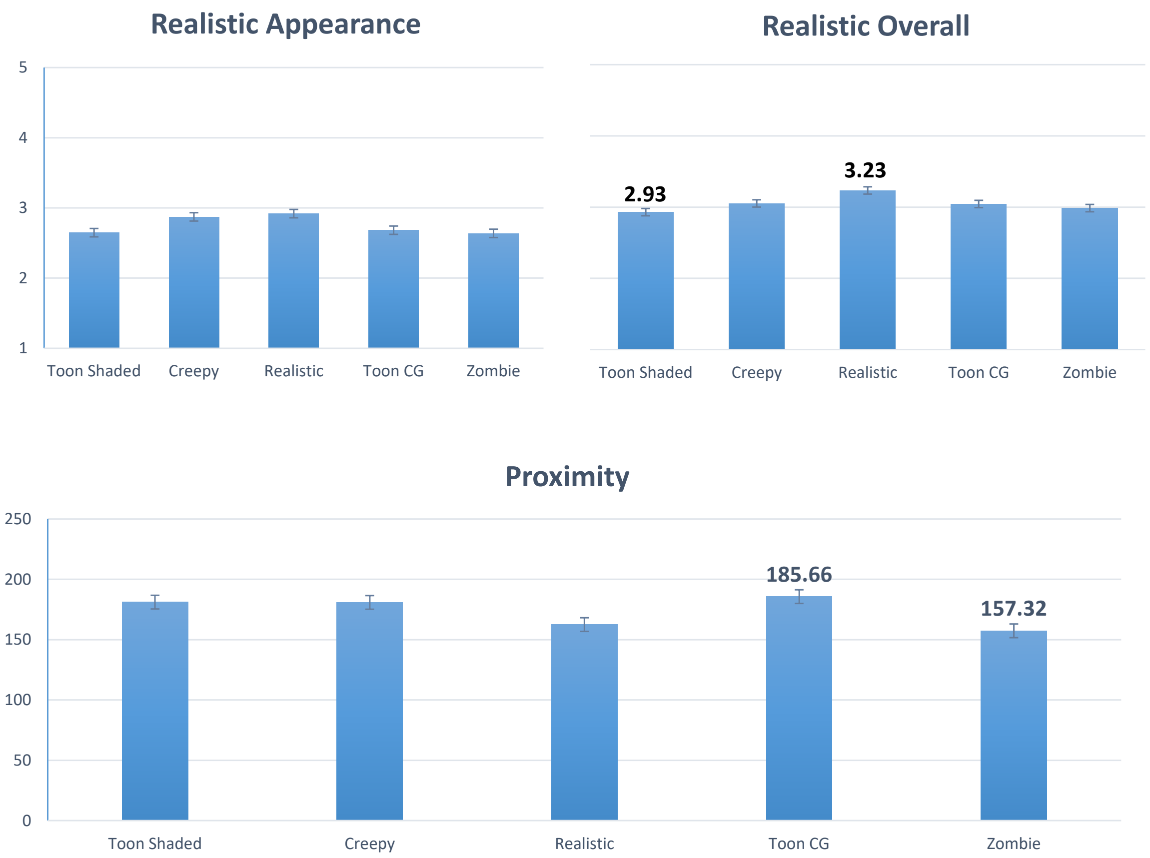


Figure 4.24: Main effect of *Render Style* for two of the Realism items and Proximity. Labeled means present differences between means which were found to be significant in the post–hoc test. Realistic render style was rated most realistic but significantly more so compared to the Toon Shaded render style. Participants also went closer to the Zombie render style, especially compared to the Toon CG render style.

result of higher co-presence difference between extraverted Toon CG and Realistic character is interesting, since it stands out from other data, but is hard to interpret. Also, the difference between proximity towards Zombie render style could mean the investigation of details, however it could also show the lack of interest for the Toon CG render style. This could possibly be also because Toon CG render style does not have a lot of details on its texture and no visible (or intentional) flaws. In addition to not finding any correlations between proximity and other responses from the participant, these results could indicate that our measure of proximity was not related to how comfortable people were to getting close to the character but rather served the examination of details of the render style. We conclude that the observation of overall effects of our data gave us some additional information on how to interpret the already discussed results, however some other results need further exploration.

d) Attribution Bias

We analysed the frequencies for Attribution of character's behaviour to either Character or Situation. Due to the variables being categorical, we approached this by using crosstabulations and computed significant differences in attribution bias across *Render Style* and *Personality* using Pearson Chi-Squared test. We found both independent variables changed the frequencies of attribution significantly (Render Style: $\chi^2 : 10.28, df = 4, p = 0.036$; Personality: $\chi^2 = 88.71, df = 5, p = 0.000$).

By partitioning the Chi-Squared into independent 2–way component tables, we further investigated the interactions. We found that certain render styles changed the attribution bias even further. Figure 4.25 shows the significant differences which were obtained by comparing the effects or Render Style inside Personality. We found that Realistic render style significantly changed the attribution towards Situation in both Agreeable and Extravert conditions, while Creepy render style did so in the Neurotic condition. However, the most extreme and significant differences are only compared to certain render styles (Toon Shaded and Toon CG).

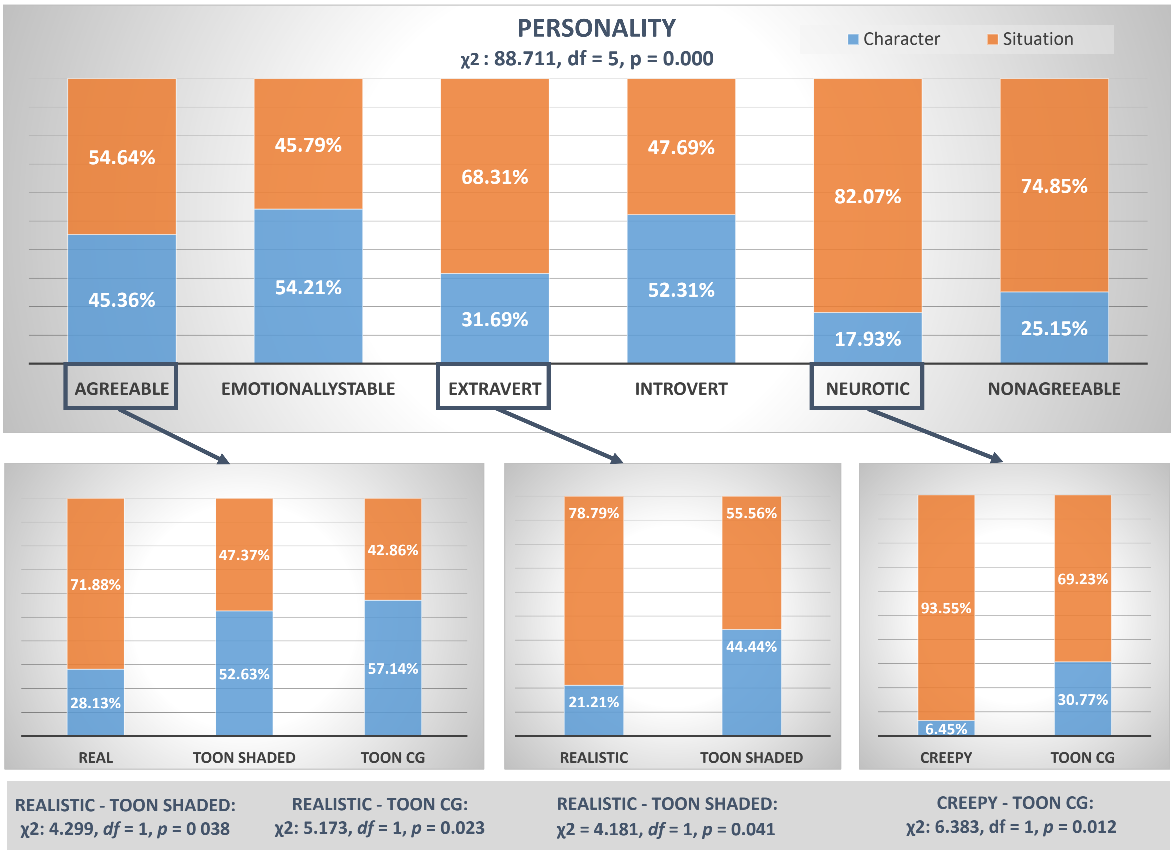


Figure 4.25: Percentage of Attribution for Situation and Personality across Render Style. Render style of the character significantly changes the attribution in the cases of Agreeable, Extraverted and Neurotic. The Chi-Squared test marks the size of significance for each comparison.

Out of interest, we conducted a non–parametric Sperman's Rank Test correlation between Attribution and all the dependent variables. There were no significant correlations except a weak correlation with the variable Concerned, where higher concern was associated with attributing the character's behaviour to Situation ($r = 0.08, p < 0.05$) .

Discussion to Section d) Main differences were found for the render styles which were perceived to be more realistic (Realistic and Creepy) which influenced some behaviour to be attributed more to the situation. However, this happened only for particular personalities and not others. We included the attribution of behaviour from personality to situation to explore automatic (or less automatic) processes, but it is hard to conclude if this was the source of difference in our results, since we did not achieve high co-presence and the proximity results do not seem to indicate the behaviour of personal space which people exhibit in social situations. Therefore, it would not be strictly valid to speak about social cognitive effects of the attribution bias since other measures did not show our virtual characters induced socially cognitive processes. If our results were to be analysed in this fashion, they would indicate that Toon Shaded and Toon CG render styles were most difficult to process in certain situations since the attribution of a behaviour to situation was lower compared to other styles, but particularly to the realistic ones (Realistic and Creepy).

However, the fact that render style could change the attribution of behaviour at all, is an interesting result which prompts further investigation. For example, the result that higher concern for the character while attributing his behaviour to situation, could imply that empathy is higher for a character if we perceive him not being responsible for a difficult situation. Since this is further mediated by the character's appearance, it is possible that attribution bias could be used as a measure of empathy towards the character. But this can only be true for negative situations, since the attribution of behaviour to situation in positive context (such as our result for extraverted and agreeable personality) was not associated with higher excitement, for example. We conclude that *attribution bias is an interesting measure to investigate response to virtual characters but needs further exploration.*

e) Analysis of Demographics

We did not expect any specific effects of which would be related to participants' background: *Sex*, *Age*, *Native* (if the participant is a native English speaker or not), *Game Experience* and *Character Experience* but we conducted a generalized linear model analysis to screen for possible effects on our data. This analysis is a form of linear regression, which is robust for violations of normality and finds systematic effects that can be made linear through a suitable transformation (Nelder & Baker, 1972). We chose this model because it can predict effects of multiple variables on our data without specific requirements for normality or homoscedasticity of variance.

Table 4.7: Parameter estimates table for significant predictors. B is the coefficient of interest and the intercept value is the reference parameter. For the Age covariate, positive (negative) coefficients indicate positive (inverse) relationships between predictors and outcome. For factors, a factor level with a greater coefficient indicates higher ratings on the dependent variable. For example, for the Concerned variable, non–native speakers gave higher values than predicted (signified by the positive B), while Age was negatively associated with Appeal, which means that the older the participants, the less appealing they rated the characters. B = 0 indicates that the parameter is similar to the intercept coefficient.

				95% Wald Confidence Interval		Hypothesis Test		
Variable	**Category**	**B**	**SE**	*Lower*	*Upper*	*x2*	*df*	*Sig.*
Concerned	(Intercept)	2.099	0.1867	1.733	2.465	126.402	1	0.000
	Native = No	0.451	0.0620	0.330	0.573	53.024	1	0.000
	Native = Yes	0^a						
Excited	(Intercept)	2.189	0.261	1.784	2.594	112.308	1	0.000
	Native = No	0.287	0.069	0.152	0.421	17.481	1	0.000
	Native = Yes	0^a						
	Age	.011	0.0027	-0.016	-0.006	16.887	1	0.000
Uneasy	(Intercept)	1.952	0.199	1.562	2.343	95.909	1	0.000
	Native = No	0.167	0.066	0.037	0.297	6.365	1	0.012
	Native = Yes	0^a						
Appeal	(Intercept)	2.251	0.213	1.834	2.667	112.144	1	0.000
	Native = No	0.345	0.071	0.207	0.484	23.983	1	0.000
	Native = Yes	0^a						
	Age	-0.010	0.003	-0.016	-0.004	9.510	1	0.002
Realistic Appearance	(Intercept)	2.343	0.217	1.918	2.768	116.698	1	0.000
	Native = No	0.149	0.072	0.008	0.290	4.299	1	0.038
	Native = Yes	0^a						
Eerie	(Intercept)	2.943	0.212	2.530	3.356	194.876	1	0.000
	Game Experience = High (passionate gamer)	-0.368	0.136	-0.635	-0.101	7.282	1	0.007
	Game Experience = None	0^a						
	Native = No	0.193	0.070	0.055	0.330	7.580	1	0.006
	Native = Yes	0^a						

a. Set to zero because this parameter is redundant.

We put the 4 categorical factors in our analysis as predictor factors while *Age* was the covariate. We performed separate estimations for the individual dependent variables. Table 4.7 contains significant results of the analysis.

We found effects for Native on variables Concerned, Excited, Uneasy, Appeal, Realistic Appearance and Eerie, where non–native speakers were rating all scales significantly higher than the native speakers. Game Experience had an effect on how participants rated Eerie, where passionate gamers where giving lower ratings of eeriness overall. Age made a difference in how participants were rating Appeal and Excited. The older the participants, the more excited they felt after watching the character but they also put lower markings for appeal. Other factors (Participant Sex, Character Experience) did not significantly influence our data.

Discussion to Section e) Adding the demographic details in the analysis provided with some interesting findings, not related to our expected predictions. For example, being a native speaker or not had a difference in most of our results measured with the questionnaire. This could possibly indicate differences in understanding the meaning of the words or cultural differences in perception, which were not the focus of our study but should be investigated further. Results such as game experience and age are interesting as well. It would be expected that a passionate gamer would be more critical about the created characters. It seems however, that the effect was reverse, since they rated characters as less eerie, probably indicating that gamers are more familiar about the limitations of the gaming platform and were more forgiving of mistakes on the characters. It also shows the care taken to create the virtual characters was met with appreciation from those participants, frequently engaged with game characters. Older participants appeared to have enjoyed the experience more but the lower appeal ratings could indicate they would not be motivated to watch the character again. We conclude that *some demographic information, such as age, country of origin and game experience of the participant should be taken into account when analysing responses to the virtual characters.*

4.2.3 Additional Experiment: Close-up

Overall, our results provided with some insights into how render style could affect our perception of the character in the virtual environment. However, the proximity measure did not give much indication of behavioural differences according to the factors. Furthermore, because of this measure, the participants had to be placed at an appropriate distance from the character, therefore small details in the facial animation were obscured. We decided to test if a closer view of the character would make a difference in how they rated him on the questionnaire items.

Experiment Design

This experiment was conducted in the last two weeks of the *Seeing* exhibition, therefore all the protocols regarding participants were the same. We tested 72 participants (31 female, 41 male; average age: 26 ± 12). Every other aspect of the task also remained the same, only that the red spot was put approximately $1m$ from the character on the very border of his collision sphere as opposed to having to observe the character $4m$ away, as in the previous experiment (see Figure 4.26). Proximity had to be omitted from this experiment.

We changed the number of presented characters to get only the most relevant differences. We reduced the number of Render Style factors to 3 (Toon CG, Realistic, Zombie). We picked these three styles because results from the previous experiment gave the most prominent

Figure 4.26: *Left:* initial viewing distance for the character (Distance condition: Far). *Right:* the chosen distance for the Close-up experiment (Distance condition: Close); this distance was chosen to be as close to the character's sphere collider without intersecting it.

differences for these styles. Also, we reduced the number of personality types to 2, one from the negative and one from the positive spectrum of the personality (Neurotic and Extravert) in order to get a matching number for both types of personality in our data set.

Results

In order to assess if the distance made a difference in how participants rated the characters, we examined differences between how participants rated the individual personalities (*Extravert, Neurotic*) and particular render style (*Toon CG, Realistic, Zombie*) on the dependent variables (*Appeal, Realism, Empathy, Co-presence, Attribution*) depending on the condition Distance (*Close, Far*). If data for the individual comparisons was not distributed normally, we analysed the differences by using Mann–Whitney U test, in case of normal distributions we chose independent sample t-test. For Attribution, as before, Pearson Chi-Squared test was used on the categorical data.

None of the comparisons between corresponding factors were different based on how close or far participants were standing, except for the Zombie render style expressing Neurotic personality. This character was rated more Eerie in the Close condition ($U(60) = 4, Z = -2.18, p = 0.037$). The Appearance Realism was also rated more realistic when participants were standing closer to the Neurotic Zombie character ($U(60) = 11, Z = 2.17, p = 0.030$).

Conclusion: the distance towards the character did not influence the subjective ratings except for the Neurotic Zombie character. The closer view in this condition increased the feeling of eeriness and appearance realism. Since no other differences were detected, we conclude overall, *that the first chosen distance (4m from the character) gave sufficient details of the character's appearance.*

4.2.4 General Discussion for Personality Experiment 1 and 2

In Experiment 1, we showed that different personalities can be successfully constructed and applied to the character and they will not only vary in affinity but will change the levels of empathy, and perceived realism, as shown in the Experiment 2. However, some of the personalities have not exhibited intended traits, which was especially notable on the positive poles of the personality types: extraversion, emotional stability and agreeableness. These traits were assessed positively by the participants but the exact description of the personality was not matched by the responses from the observers. This puts a limitation in the explanation of results in the direction of a particular personality type, e.g. extraversion as the least eerie personality type. This is not a valid conclusion, since it is only our situation, which was intended to portray extraversion, which was perceived as less eerie. More examples of a personality type would need to be included to assess the effect of the personality type. The negative poles (neurotic, introverted and non agreeable) of the personality types were more distinctive, as shown in Experiment 1. However, as before, generalising that it was only the particular personality type which affected participants' responses to the character would not be valid. More examples of the same personality type would be needed to test this. In our experiments, we only found an indication that the relation between personality and peoples' responses is a possibility.

Another observed effect, which was interesting was that in Experiment 2, we did not get as many differences in appeal between personalities as we did in Experiment 1, which confirmed our assumption that a choice in experiment design, i.e. repeated measures as opposed to between group design, might have influenced the personality (and possibly other variable) ratings. The only similar result was observed for introverted personality, which was rated significantly lower on appeal than other personality situations in both experiments. We conclude that experiment design should be carefully considered when measuring the response to virtual characters.

The most notable result is that render style affects the perception of characters, and this is further related to a particular behaviour the character is exhibiting. While we found that the Toon CG was perceived more Agreeable than the HumanIll render style in Experiment 1, Experiment 2 showed some more differences which were associated with social perception. Possibly the most striking result in comparison with previous studies on the uncanny valley is that, overall, the render styles did not vary significantly in appeal and that appeal is not associated with low eeriness. Render style only changed the appeal in combination with a personality, potentially a personality which was not “suitable” for a particular appearance. For example, realistic render style was found more appealing when the character in this style was portraying a neurotic personality. However, it was also seen as more eerie. An extraverted cartoon–looking character could have been a better choice for our extraverted personality, where the level of co-presence was found to be higher. Also, Zombie render style, even though intentionally created to be eerie, was only seen as explicitly eerie when portraying an agreeable personality, possibly again since a positive personality would not be associated with such a character. Our results therefore show that the question of affinity towards virtual characters is complex and depends on the context but can still follow a particular logic – realism increases empathy in negative situations, whereas unappealing characters will not be associated with positive scenarios. There are some limitations to this conclusion. Namely, our results do not show a concise effect for all situations on all render styles. For example, it does not explain why Zombie render style was particularly eerie only when expressing Agreeable personality and not Extraverted or Emotionally stable.

Some interesting relationships between dependent variables we used to study the effects were found as well. As mentioned, appeal was not related to lower eeriness according to our results. Also, proximity was not associated with higher co-presence with the character or any other sign of affinity. In our study, it seemed to have been an indication of interest in visual details, while we were not able to recreate the effect of personal space which can be observed in social situations. This could have been because we lacked cues which make the co-presence higher, e.g. eye-gaze (Bailenson et al., 2003).

The indirect measure of attribution proved to be an exciting new addition to the measurements as well, which extends the previous study of McDonnell (2012) which used an indirect measure of trustworthiness but did not find an effect of render style. Our explanation was that the lie detection task in that study was too difficult for the participants and therefore the response showed more cognitive effort than perceptual effects. Even though it is hard to explain where the results from our study comes from and if it actually indicates the ease of processing (Gilbert et al., 1988), it showed some differences according to which render style is exhibiting a certain personality and the attribution of behaviour towards situation seems to be associated with empathy as well. This is not unexpected, since higher empathy is evoked when we feel the person is not responsible for the negative outcome but was subjected to a difficult situation. However, based on our experiment, stronger conclusions cannot be made on this topic.

4.3 Summary of Chapter 4

The most important result in this chapter is that realistic render style was not only considered appealing, but also prefered in certain platforms (virtual reality) and has increased empathy for situations, which are typically associated with negative affect (neurotic). There was also an indication that the cartoon render style could be more readily associated with certain positive situations, since it would make the character seem more “alive”. This shows that the affinity towards virtual characters based on their appearance is one component, but the interaction between the appearance and character’s behaviour will add additional information, which if appropriate, will produce an engaging result. This can also be seen when comparing the Personality Experiment 2 and Agency Experiment 2, which were conducted in virtual reality. The realistic render style in the Agency experiment was significantly more appealing than other styles, while in the Personality experiment only an interaction with a particular personality. Similarly, toon shaded render style was least favoured in the Agency experiment (more eerie, less appealing, less alive), whereas we did not observe the same trend in the Personality experiment. Toon shaded was actually rated less eerie than the realistic render style in some personality situations (e.g. introverted). This is an important difference, since it puts emphasis on what the character is expressing when judging appeal. It is possible that since participants in the Agency experiment only observed the character exercising, the appeal was based mostly on appearance while in the Personality experiment, appearance interacted with the portrayed personality and influenced the perceived appeal into direction of the already mentioned appropriate or expected way.

The Personality experiments also showed that appeal does not imply lower eeriness and the latter could actually indicate increased interest in the character. Our styles which were intentionally created to be unsettling, therefore raised the eerie effect (especially in positive situations) but were not necessarily considered less appealing, and were approached closer, and in some cases, even rated least eerie (neurotic personality for the zombie render style). The Agency experiment did not show a similar pattern: a more appealing style was usually rated as less eerie as well. According to Ho and MacDorman (Ho & MacDorman, 2010) eeriness is not related to agreeableness of the character, therefore this result of the Personality experiment is not surprising.

An interesting finding is also that the render style could actually change the perception of motion realism (result from both Agency and Personality experiment). It seems that certain styles, since they are less realistic in appearance, could also influence how motion is interpreted. However, the same motion could potentially be different on the character, if certain render style hides details of the motion, or does not provide enough depth cues, etc.

Unfortunately, our measure of personal distance did not describe behaviour which people exhibit in real life situations. It appeared to have been related more to interest in visual details, where people would come closer to a character to observe the details of his appearance. This was the case in both Agency (particularly in the controlling condition) and Personality experiment (when observing Zombie render style). However, in the Agency experiment, participants moved further away from the character when they perceived to have less control over him. This shows that changes in the interaction with the character influence the proximity measure to a greater extent. Since we did not measure perceived agency in the Personality experiment, we cannot directly compare the results or the actual differences in average minimal distance from the character. But we observe that the participants did not approach the character closer if they perceived him to be more alive in the observation condition of the Agency experiment, while we did not get any correlation between co-presence and proximity in the Personality experiment. Since measures of co-presence were relatively low for both experiments, we conclude that a higher level of induced co-presence would show a stronger effect in the proximity data. Even so, the observation of details which changed the proximity behaviour in our experiments is an interesting measure which signals motivation to move closer in order to observe the character.

The other indirect measure (attribution bias) which signifies the ease of processing also showed some changes according to character realism which prompts further investigation.

The notable addition to our experiments was also including virtual reality as a more immersive experience, which allows measuring the behaviour of the participants. However, apart from already mentioned preference for realism in virtual reality, we did not get much difference when comparing the results between platforms, particularly in the Agency experiment, where the experiment had a similar design. We were especially expecting differences in co-presence, which did not prove to be the case. This is not a surprising result however, since it was pointed out in research before that the interaction of the character with the perceiver is crucial (Blascovich, 2002; Bailenson et al., 2005). It could also be possible that the screen based experiment was equally immersive than the virtual reality environemnt (we used a large TV screen for observation). Virtual reality could have a positive effect on people's motivation to explore the environment, since it is a relatively new platform which many people have not experienced before. This was apparent especially in the Personality experiment, where we put more importance in the appearance of the environment as well as the functionality of the experiment. How exactly that would affect the data in our experiment is not certain but it could have influenced the behavioural measure where people would be more motivated to go closer to the character to observe details.

Overall, we are confident that this chapter presented a more complex and in–depth understanding of the effects the character's appearance has on the perception of the socially relevant traits. We also successfully implemented a character of higher level of realism in the virtual reality environment, which should encourage other studies to begin exploring realism in interactive immersive environments as well.

Chapter 5

Conclusions and Future Direction

The purpose of this thesis was to present two main approaches of manipulating the appearance of virtual characters: creating a mismatch in visual components (Chapter 4) and changing the render style (Chapter 5), to explore differences in perceptual response towards these characters. In this Chapter, our main findings are presented and discussed in separate paragraphs. Later, we describe interesting outcomes of our work which were not part of our main interest but could have important implications for future research, which we present in the last section.

5.1 Contributions

The contributions of our work provide important explanations and guidelines for the design of appealing virtual characters for various graphics applications. Many different, sometimes even contradicting, findings exist in the literature around this topic. Our study takes a novel approach in first providing a theoretical discussion about social perception of humans and then making predictions about how this could apply to virtual characters. This approach resulted in many beneficial outcomes, which we summarise below:

- **Research contributions**

  Our main finding is that the appearance of a virtual character plays an important role in how its behaviour is perceived and interpreted by the observer. Appearance can change the way perceivers empathise with the character, how they judge its personality and behave around it. Perhaps the most interesting outcome is that a visual style of a character is not unappealing on its own. It interacts with what the character is expressing and can result in engaging combinations which guide the perceiver's response. A visual style, which is rated appealing might not be likable any more if portraying behavior which does not "match" the appearance, and an unappealing style could actually become engaging following the same pattern.

Important outcomes were also related to visual realism. Apart from our result that realism consists of components which have differential importance on the overall realism (shape, material, motion), the perceived realism might also trigger the schema used in processing of social behaviour of real humans. However, the full effect of visual realism will be apparent only when the character is expressing socially relevant behaviour. We were able to detect this by including personality of the character into our research. Also, the perception of realistic characters in an immersive platform (virtual reality) could be particularly crucial for developing social cognition, similar to when interacting with real humans.

In addition, we showed that social cognition is involved in many aspects of character perception. In gender perception, stereotypes are involved in the gender recognition process. The attribution bias is affected by the render style. By analysing shape and material effects on appeal, we found that perceivers dislike realistic material on virtual characters due to an excess in skin details, which are unattractive in real humans since blemishes are a sign of unhealthy skin.

- **Methodology Contributions**

  We found that indirect measures of proximity and bias (gender, personality and attribution) are promising ways of studying the appeal of virtual characters, which can result in a more well-rounded understanding of the effect that a particular virtual character has on the observer. We also focused on the use of immersive environments to study behavioral effects of the interaction with virtual characters in addition to measuring the subjective reports with Likert scales. And lastly, by conducting the study described in Section 4.2 on a very large sample of population (1106 participants) from various different backgrounds, we added a significant amount of relevance and reliability to our conclusions.

  The results of proximity in our studies did not indicate only the level of comfort to approach the character, but also the level of interest to observe him. The important novelty of our research was therefore the use of the measure of proximity as something other than co-presence.

- **Material Contributions**

  For the purposes of our research, we created some material which could be useful for other researchers and studies. For example, we added to the library of natural motions for virtual characters. We captured facial and/or body motion from professional actors expressing social behaviour – emotional walks, emotional conversation and motions, exhibiting personality traits. We also added some motions expressing neutral conditions, i.e. exercising motion, idle standing motion, neutral walk and neutral conversation. We also created a library of static virtual faces (with expressions) with different levels of stylisation and mismatching (or matching) material and shape combinations, which can be used for other studies interested in the perception of virtual faces.

By relying on the psychology literature evaluating the theories on human personality, specifically the Big Five factors, we were able to create scenarios which include an observable set of personality traits, exhibited through actor's motion and voice. The situations, along with the non–verbal behaviour and language specifics, are presented in the Appendix and can be used by other researchers, interested in creating personality scenarios for their work.

An important practical contribution was also the design of an engaging and interactive screen–based and virtual reality environment, created to be a game and an experiment trial at the same time. Most participants who took part in the Science Gallery exhibition experiment described in Section 4.2, reported high levels of motivation to finish the experiment and an overall enjoying experience. While most of their reaction was related to the first exposure to virtual reality, we are confident that the great effort which was put into creating the appropriate design (e.g. speed of motion through the environment, sound effects, high frame rate, proper lighting and materials, etc.) contributed to their positive response.

- **Guidelines for Character Design**

  One of the important features to be considered when designing a virtual character, is the prominent role of social perception. For example, stereotypes can bias the perception of gender away from the biological sex, when the motion cues are not rich with gender cues already. While this is a specific case, it could have implications on other aspects of character design, when particular information about the character are obscured. In addition, extreme mismatches in shape and material realism should be avoided, unless the aim is to create an unappealing and eerie character. When the designer wishes to influence the perceived realism to a greater extent, the easiest way is to change the realism of shape of the character. And also, since appearance interacts with the personality the character is exhibiting, care should be taken in choosing the appropriate visual style for the character, since it can make a difference in how the information he is portraying will be perceived by the audience.

In summary, our research shows that the visual components and their effect on the perception of virtual characters reveal a very complex nature of interactions between visual elements. Therefore, when designing an experiment, additional definitions of factors should be considered, such as realism of sub-components (shape or material, etc.), or what scenario the character is portraying and on which platform. Also, the social context in which the character is integrated, should also be considered since it can change the perception of information that the character is expressing. Our research is the first to study perception of personality and agency of a character in an immersive environment and on an extensive sample of participants. By using state of the art technology and software, we were able to create a highly realistic character, which is also rarely examined in immersive platforms. We also implemented indirect and behavioral measures in ways which were not used before in the related research.

Apart from our main contributions, we found some interesting outcomes, which are important for the methodological approach to exploring similar research questions. Firstly, the scales used to study the subjective response of uncanny valley, *appeal* and *eeriness*, were not following a similar pattern in all our studies. When exploring the shape and material mismatch (Section 3.2) and render style in relation to agency (Section 4.1), we found that high level of appeal was paired with lower levels of eeriness and vice versa. This was not observed in our last study of render style and personality (Section 4.2). While eeriness and appeal represent two distinct categories as proposed by Ho and MacDorman (2010), it is unlikely that an eerie character will also be appealing, as shown in our last study. This is an interesting finding and could indicate that a virtual character who exhibits a certain personality could be interesting to observe (appealing and engaging) while being eerie at the same time. For example, an engaging zombie character could be interesting to watch, similarly to the motivation for watching horror movies and particularly disturbing scenes in movies. However, it is possible this response is associated with an artificial representation, which cannot be expected to exist in real life, therefore the threat is removed. A study exploring this effect of zombie vs. soldier characters in an immersive environment, found a result which could confirm this assumption (Bruneau et al., 2015). In this study, walking through a crowd of characters was affected more when they were plausible in real life (soldiers), then when they were not (zombies). Both types of characters could be perceived as threatening, but only soldiers could be encountered in real life. Of course, there are other possible explanations for this result as well (specifics of motion or appearance of the crowd, etc.)

Secondly, our virtual reality experiments have shown that our most realistic render style of the character was an appealing and even preferred design option for this platform. Perhaps an immersive environment is closer to real life, therefore perceivers also expect real events, people and realistic behaviour. It is also possible, however, that due to the novelty and excitement of virtual reality, people's reports are inflated and therefore some responses might not be observed in the future any more, when people become accustomed to this novelty. For example, the interest to explore eerie characters, which affected the measure of proximity in our experiments the most, could be due to the fascination of exploring details of the character, especially because of the added spacial component which the virtual reality provides as opposed to screen–based environments. Participants in our virtual environment, particularly in the last experiment, reported to have a very good experience overall and were thrilled to have participated in the experiment.

And lastly, an interesting but unexpected side–effect of our variation in experiment designs showed that when participants were seeing only one render style and one example of personality, responses were different to the ones collected from participants which rated multiple styles and personalities in one session. This could simply be due to differences in statistical analysis itself (between-group design as opposed to repeated measures statistics) but it could also signal some effects which repeated exposure to a character in different styles have. Perhaps when the observer sees the character for the first time, the response to it is less predictable since there is nothing to base the judgment on. While we tried to avoid this by presenting another render style in the environment prior to the experiment in our last study (Section 4.2), it is still difficult to assess which design would give more reliable (and generalisable) results. We argue that the between-group design provides more reliable results, however this conclusion is open to discussion. None the less, we conclude that the experiment design is a crucial consideration when conducting research in the area of perception of virtual characters.

5.2 Future Work

There are certain limitations to the interpretation of our results, which prompt further investigation. Also, the conducted research gave us new insights into particular topics which could direct our future work.

One of the main drawbacks of our research is that we were not able to achieve a high level of co-presence with the characters in virtual reality. Our attempt to study social cognition was therefore not completely achieved. Even though we found indications of social cognition (stereotypes, bias when assessing personality traits), our behavioural and indirect measures were not affected by social response of the participants. Therefore, our conclusions about realism indicating the triggering of social schema for real people, could not be investigated. In addition to raising the levels of co-presence, our results would need to be compared to real–

life situations to evaluate what the actual effect of realistic render style on social cognition is. This is also an important part of our future work.

To achieve this, we propose a couple of solutions. We have conducted our experiments in virtual reality with some limitations, such as movements through the virtual environment being performed by using the gamepad joystick as opposed to actually walking through the environment. This is a limitation which should be overcome in future research by allowing the participant to move through virtual space with his body. For example, by using the Vive® headset, product of ©HTC[1], the position of the participant in the real environment can be accurately tracked and translated in the virtual space. By doing so, we would achieve a higher level of immersion in the virtual environment and our behavioural data would be more reliable as well. In addition to making the motion through the environment more natural, greater care should be given to creating a responsive character, who would react to the presence of the perceiver and therefore induce a higher level of co-presence in him. Both aspects, behavioural realism of the character and perceiver's motion through the environment, are frequently used in other experiment designs in the literature to increase immersion and co-presence (Nowak & Biocca, 2003; Bailenson et al., 2003; Garau et al., 2003).

In addition to limitations when interpreting attribution bias, we did not discuss the effect of this bias on stylised virtual characters. For example, we assumed that the realistic style is associated with slower perceptual processes since it would trigger the social schema for real humans but not quite reaching it. However, stylised characters could also trigger slower processing mechanisms, which could be related to how familiar people are with a particular style. Familiarity has been known to affect social attribution (Klauer et al., 2012) and is an important component of the uncanny valley (Ho & MacDorman, 2010), therefore it is possibly related to a faster processing pathway for stylised characters as well. It would be appropriate to test this possibility in future research as well.

We would also like to point out that our theoretical approach, explaining the processing of perceptual information, heavily relies on contemporary research in neuroscience and social psychology, which support the outcomes of previous research done on the perception of virtual characters. We do not discuss alternative approaches which could have different outcomes for our work. For example, we explain the attribution process from the viewpoint of categorical perception (Sherman et al., 1998; Macrae et al., 1999) which is further based on the theory that this is the most efficient way for the brain to conserve cognitive energy (Friston, 2010). We do not discuss alternative explanations to brain functioning which would have different outcomes for our research. In this context, our approach to studying the social cognitive elements while engaging with virtual characters is open to discussion.

[1]https://www.vive.com/eu/, Retrieved 20 October 2016

In our future work, we would like to explore more methods of retrieving information from participants. We would like to include other subjective responses, e.g. when evaluating the responses of participants to expressions of emotion, scales such as believability and confidence ratings could be used (Cunningham et al., 2003). Implicit measures, such as priming, could be considered as well. Priming is a phenomenon in which the speed or accuracy of a decision to a target is increased when a previous stimulus or prime is presented. Priming effects can be used to investigate the effects of prior context upon the recognition of stimuli and have been widely used to study word and object recognition but also more complex stimuli, such as facial identity (Lander & Bruce, 2004). These methods would only allow analysis of fast impressions of virtual characters, therefore it would not be useful for dynamic stimuli or longer exposures. Particularly useful for measuring responses in longer and engaging scenarios are physiological measures, such as changes in skin conductance, heart rate and respiration (Meehan et al., 2002). These measures should be added in the future as well.

We are also interested to explore the motion in more detail, since we confirmed the motion has a very important role in perception of virtual characters. Our motion capture system allows us to study natural human motion, however stylisations of this motion would enable us to study realism of motion in combination with realism of appearance as well. Stylisation could be performed with the help of animators, which typically follow a set of rules while animating the cartoon characters. We could also introduce artifacts in the natural motion in order to make it less realistic. This way we could explore mismatches in both appearance and motion dimensions and how they affect the perception of virtual characters. It would be particularly interesting to focus on facial animation, since it can produce exceptionally disturbing results when visual elements are distorted (Hodgins et al., 2010; Tinwell et al., 2013).

We believe the work presented in this thesis is a valid contribution to the knowledge on the perception of virtual characters. This does not include only the field of graphics, but also psychology and other social sciences which deal with human perception.

We are experiencing a great leap in computer graphics, where virtual characters and environments are generated to appear very similar to real life while the technology which powers this experience is becoming consumer available (e.g. Oculus for displaying virtual reality). Humans seem to have an innate drive for replicating (or modifying) reality and this motivation is unlikely to cease. The advances in the field of computer science, in particular computer graphics, are a way of providing a very powerful platform for this drive to reach a whole new level of realisation. It is exciting to anticipate how our way of living will be affected by this and what psychological consequences of such an evolution will be.

Chapter 6

Appendix

Personality (Big Five Factors)	**Descriptions of Personality** (high scores on Big Five)	**Situation Description**
EXTRAVERSION	Captures the attention of others, is expressive and dominant, humorous, fashionable, positive, likeable, sociable, talkative, enthusiastic, playful, has high self-esteem, exhibits high energy level, shows involvement with others, says or does interesting things; Body language: Is expressive in face, voice or gestures, is physically animated, energetic; Language: Is informal in conversation, looser and less specific in descriptions, uses stylistic expressions such as "catch up", more likely to use plural "we".	**RECEIVING A NEW TV AS A BIRTHDAY PRESENT FROM FRIENDS:** EXPRESSING GRATITUDE, POSITIVE EMOTIONS EXPRESSING HOW GOOD AND IMPORTANT HIS FRIENDS ARE HE KNOWS A LOT ABOUT THE GIFT
AGREEABLENESS	Friendly and non-aggressive, warm in interpersonal behaviour, trustworthy, straightforward, altruistic, compliant, modest, tender-minded, spend less time at home, good natured, sympathetic, forgiving; Language: Uses first-person singular pronouns (I, me, my) (to convey personal rapport);	**TALKING ABOUT A WORK ASSIGNMENT HE DID WITH HIS COLLEAGUE:** EXPRESSING THE GOOD RELATIONS WITH HIS COLLEAGUE BEING SUPPORTIVE EXPRESSING POSITIVE EMOTIONS
CONSCIENTIOUSNESS	Behaving responsibly, carefully, and with self-discipline, spends a lot of time in class or work, reliable, well-organized, competent, dutiful, achievement striving, deliberate, value organization and order; Language: Use first-person singular pronouns (I, me, my) (to convey personal rapport);	**EXPLAINING HOW HE ACHIEVED SUCCESS IN COLLEGE:** WELL-ORGANISED, SELF-DISCIPLINED LONG TIME SPENT IN COLLEGE SHOWS A HINT OF BEING ABOVE OTHER STUDENTS (BETTER THAN THEM)
OPENNESS	Curious, imaginative, and having a wide range of interests, spends more time in public places like restaurants, bars, and coffee shops, are creative, complex, understanding; Value sentience, nurturance, innovation, tolerance, change, fantasy, aesthetics, feelings, actions, ideas, values; Language: Use first-person singular pronouns (I, me, my) to convey personal rapport;	**TALKING ABOUT HIS INTERESTS, SPARE TIME:** VERY OPEN TO DIFFERENT THINGS, NEW EXPERIENCES HAS A LOT OF HOBBIES, SOMETIMES THEY ARE RISKY PUTTING IMPORTANCE ON DOING A LOT OF DIFFERENT THINGS
EMOTIONAL STABILITY	Relaxed, secure, hardy, autonomous, is a risk taker, calm, not easily upset, and low in self-consciousness; spends less time arguing, goes to public places other than restaurants, bars, and coffee shops; calmly reacts in stressful situations – treats stressors as a challenge;	**EXPLAINING WHAT HAPPENED AT THE MEETING WITH HIS BOSS:** TURNING A BAD SITUATION AROUND BEING PRODUCTIVE, POSITIVE, CONSTRUCTIVE DOES NOT LET HIMSELF BE STRESSED OUT ABOUT IT

Table 6.1: Situation descriptions for the Big Five personality factors. The responses to these situations were free for actor's interpretation.

Personality (Big Five Factors)	Descriptions of Personality (low scores on Big Five)	Situation Description
LOW EXTRAVERSION	Introverted, reserved, inhibited, quiet (talking to oneself), expresses criticism, seems detached from the interaction, talks at rather than with partner, exhibits an awkward interpersonal style, behaves in a fearful or timid manner, is reserved and un-expressive; Language: Greater use of quantifiers, introverts use more first-person singular self-references ("I"), negations, more tentative and timid in language ("trying-"," going- " (to)).	**DESCRIBING A PARTY HE WENT TO:** NOT GETTING ALONG, PEOPLE NOT NOTICING HIM BORING CONVERSATIONS, NOT KNOWING ANYTHING ABOUT THE TOPIC WANTING TO GO HOME AND BE ALONE
LOW AGREEABLENESS	Opposite of Agreeable; Critical, rude, harsh, callous, self-defendant, quarrelsome; uses aggression to get what he wants.	**EXPRESSING OPINION ABOUT THE RESTAURANT CHOICE:** NOT HAPPY WITH THE CHOICE OF RESTAURANT THE SERVICE IS POOR, THE FOOD IS NOT GOOD BEING ARGUMENTATIVE, QUARRELSOME
LOW CONSCIENTIOUSNESS	Opposite of Conscientiousness; Disorganized, undependable, negligent, impulsive, playful;	**EXPLAINING WHY HE DID NOT SHOW UP AT THE CONCERT:** EXPLAINING HIMSELF TO FRIENDS SHOWING HIGHLY UNORGANIZED LIFE-STYLE, NEGLIGENCE, UNINTEREST
LOW OPENNESS	Opposite of Openness; Conventional, uncreative, spends time in the same public places, narrow minded, simple.	**TALKING ABOUT HOBBIES, FREE TIME:** NOT A LOT OF HOBBIES, CONVENTIONAL CONTENT WITH HOW THINGS ARE
LOW EMOTIONAL STABILITY	Socializing, particularly in same-sex groups, and spending time in bars; values conformity, interpersonal affect; strives for social recognition, harm-avoidance; anxious, angry, depressed, self-conscious, impulsive, vulnerable; gives up when faced with obstacles; Language: Using words like "with, and, include " (indicating inclusion) first person pronouns, such as I, me, we (indicating preoccupation with self);	**CONFRONTING A FRIEND WHO CRASHED HIS CAR:** WORRYING ABOUT WHAT OTHER PEOPLE WILL SAY, CATASTROPHIZING BEING STRESSED OUT, NOT HANDLING STRESS WELL

Table 6.2: Situation descriptions for the polar opposites of the Big Five.

Ten Item Personality Inventory (TIPI) for Virtual Characters

Modified from Gosling et al. (2003)

Here are a number of personality traits that may or may not apply to the virtual character. Please write a number next to each statement to indicate the extent to which you agree or disagree with that statement (1 - disagree strongly, 7 - Agree strongly). You should rate the extent to which the pair of traits applies to the character, even if one of the traits applies more strongly than the other.

This character is Extraverted, enthusiastic:

Strongly Disagree — Strongly Agree

1 2 3 4 5 6 7

This character is Critical, quarrelsome:

Strongly Disagree — Strongly Agree

1 2 3 4 5 6 7

This character is Dependable, self-disciplined.:

Strongly Disagree — Strongly Agree

1 2 3 4 5 6 7

This character is Anxious, easily upset:

Strongly Disagree — Strongly Agree

1 2 3 4 5 6 7

This character is Open to new experiences, complex:

Strongly Disagree — Strongly Agree

1 2 3 4 5 6 7

Figure 6.1: TIPI questionnaire which was used to measure the perceived personality traits of the character (*continues on next page*).

<u>This character is Reserved, quiet:</u>

Strongly Disagree — Strongly Agree

1 2 3 4 5 6 7

<u>This character is Sympathetic, warm:</u>

Strongly Disagree — Strongly Agree

1 2 3 4 5 6 7

<u>This character is Disorganized, careless:</u>

Strongly Disagree — Strongly Agree

1 2 3 4 5 6 7

<u>This character is Calm, emotionally stable:</u>

Strongly Disagree — Strongly Agree

1 2 3 4 5 6 7

<u>This character is Conventional, uncreative</u>:

Strongly Disagree — Strongly Agree

1 2 3 4 5 6 7

Note: a version for participants was used also. In this alternative version, all statements changed the words at the beginning, e.g. *"The character is Conventional, uncreative:"* to "I see myself as *Conventional, uncreative:"* .

Figure 6.2: TIPI questionnaire which was used to measure the perceived personality traits of the character. We also used the original version of the TIPI to measure the participants' personality.

Bibliography

Albrecht, S. & Carbon, C.-C. (2014). The fluency amplification model: fluent stimuli show more intense but not evidently more positive evaluations. *Acta Psychologica*, *148*, 195–203.

Alexander, O., Rogers, M., Lambeth, W., Chiang, M., & Debevec, P. (2009). The digital Emily project: photoreal facial modeling and animation. In *ACM SIGGRAPH 2009 Courses*, (pp.~12). ACM.

Alley, T. R. & Cunningham, M. R. (1991). Averaged faces are attractive, but very attractive faces are not average. *Psychological Science*, *2*(2), 123–125.

Allport, G. W. (1979). *The Nature of Human Prejudice*. Basic books.

Ambady, N., Hallahan, M., & Rosenthal, R. (1995). On judging and being judged accurately in zero-acquaintance situations. *Journal of Personality and Social Psychology*, *69*(3), 518.

Animatrix. (2003). [motion picture]. Arias, M., Lamm, S., Wachowski, A., Wachowski, L. (Producers), Morimoto, K., Watanabe, S., Maeda, M., Chung, P., Jones, A., Kawajiri, Y., Koike, T. (Directors). USA: Warner Home Video.

Argyle, M. (1988). Bodily communication. *Methuen & Co. Ltd, New York*.

Atkinson, A. P., Dittrich, W. H., Gemmell, A. J., Young, A. W., et al. (2004). Emotion perception from dynamic and static body expressions in point-light and full-light displays. *Perception*, *33*(6), 717–746.

Aviezer, H., Trope, Y., & Todorov, A. (2012). Body cues, not facial expressions, discriminate between intense positive and negative emotions. *Science*, *338*(6111), 1225–1229.

Back, M. D., Schmukle, S. C., & Egloff, B. (2011). A closer look at first sight: Social relations lens model analysis of personality and interpersonal attraction at zero acquaintance. *European Journal of Personality*, *25*(3), 225–238.

Bailenson, J. N., Aharoni, E., Beall, A. C., Guadagno, R. E., Dimov, A., & Blascovich, J. (2004). Comparing behavioral and self-report measures of embodied agents' social presence in immersive virtual environments. In *7th Annual International Workshop on Presence, Valencia, Spain*.

Bailenson, J. N., Blascovich, J., Beall, A. C., & Loomis, J. M. (2001). Equilibrium theory revisited: Mutual gaze and personal space in virtual environments. *Presence*, *10*(6), 583–598.

Bailenson, J. N., Blascovich, J., Beall, A. C., & Loomis, J. M. (2003). Interpersonal distance in immersive virtual environments. *Personality and Social Psychology Bulletin*, *29*(7), 819–833.

Bailenson, J. N., Swinth, K. R., Hoyt, C. L., Persky, S., Dimov, A., & Blascovich, J. (2005). The independent and interactive effects of embodied-agent appearance and behavior on self-report, cognitive, and behavioral markers of copresence in immersive virtual environments. *Presence*, *14*(4), 379–393.

Bar, M. (2009). The proactive brain: memory for predictions. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *364*(1521), 1235–1243.

Baran, I. & Popović, J. (2007). Automatic rigging and animation of 3d characters. In *ACM Transactions on Graphics (TOG)*, volume 26, (pp.$\tilde{7}$2). ACM.

Barsalou, L. W. (2009). Simulation, situated conceptualization, and prediction. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *364*(1521), 1281–1289.

Bartneck, C., Kulić, D., Croft, E., & Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International journal of social robotics*, *1*(1), 71–81.

Bartz, D., Cunningham, D., Fischer, J., & Wallraven, C. (2008). The role of perception for computer graphics. *Eurographics state-of-the-art-reports*, 65–86.

Bassili, J. N. (1978). Facial motion in the perception of faces and of emotional expression. *Journal of Experimental Psychology - Human Perception and Performance*, *4*(3), 373–379.

Battocchi, A., Pianesi, F., & Goren-Bar, D. (2005). A first evaluation study of a database of kinetic facial expressions (DaFEx). In *Proceedings of the 7th International Conference on Multimodal Interfaces*, ICMI '05, (pp. 214–221).

Beane, A. (2012). *3D Animation Essentials* (1st ed.). Alameda, CA, USA: SYBEX Inc.

Beeler, T., Bickel, B., Beardsley, P., Sumner, B., & Gross, M. (2010). High-quality single-shot capture of facial geometry. *ACM Transactions on Graphics*, *29*(4).

Ben-David, B. M., van Lieshout, P. H. H. M., & Leszcz, T. (2011). A resource of validated affective and neutral sentences to assess identification of emotion in spoken language after a brain injury. *Brain Injury*, *25*(2), 206–220.

Benson, P. & Perrett, D. (1992). Face to face with the perfect image. *New Scientist*, *133*(1809).

Beowulf. (2007). [motion picture]. Zemeckis, R., Bing, S., Rapke, J., Starkey, S. (Producers), Zemeckis, R. (Director). USA: Paramount Pictures.

Bernstein, D. (2010). *Essentials of Psychology.* Cengage Learning.

Blackwell, H. R. (1952). Studies of psychophysical methods for measuring visual thresholds. *JOSA*, *42*(9), 606–614.

Blascovich, J. (2002). Social influence within immersive virtual environments. In *The social life of avatars* (pp. 127–145). Springer.

Borkenau, P. & Liebler, A. (1992). Trait inferences: Sources of validity at zero acquaintance. *Journal of Personality and Social Psychology*, *62*(4), 645.

Bouchard, S., Bernier, F., Boivin, E., Dumoulin, S., Laforest, M., Guitard, T., Robillard, G., Monthuy-Blanc, J., & Renaud, P. (2013). Empathy toward virtual humans depicting a known or unknown person expressing pain. *Cyberpsychology, Behavior, and Social Networking*, *16*(1), 61–71.

Brewer, M. B. (1988). *A dual process model of impression formation.* Lawrence Erlbaum Associates, Inc.

Brody, L. R. & Hall, J. A. (2000). Gender, emotion, and expression. In Lewis, M. & Haviland-Jones, J. (Eds.), *Handbook of Emotions*, (pp. 338–349). Guilford.

Bruneau, J., Olivier, A.-H., & Pettré, J. (2015). Going through, going around: a study on individual avoidance of groups. *IEEE transactions on visualization and computer graphics*, *21*(4), 520–528.

Burleigh, T. J., Schoenherr, J. R., & Lacroix, G. L. (2013). Does the uncanny valley exist? an empirical test of the relationship between eeriness and the human likeness of digitally created faces. *Computers in Human Behavior*, *29*(3).

Calder, A. J. (1996). Facial emotion recognition after bilateral amygdala damage: Differentially severe impairment of fear. *Cognitive Neuropsychology*, *13*(5).

Carter, E., Mahler, M., & Hodgins, J. (2013). Unpleasantness of animated characters increases viewer attention to faces. In *Proceedings of the ACM Symposium in Applied Perception*, (pp. 35–40).

Carter, E. J., Hyde, J., Williams, D. L., & Hodgins, J. K. (2016). Investigating the influence of avatar facial characteristics on the social behaviors of children with autism. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, (pp. 140–151). ACM.

Cast Away. (2000). [motion picture]. Hanks, T., Rapke, J., Starkey, S., Zemeckis, R. (Producers), Zemeckis, R. (Director). USA: 20th Century Fox.

Catmull, E. E. (1974). *A Subdivision Algorithm for Computer Display of Curved Surfaces.* PhD thesis. AAI7504786.

Chaminade, T., Hodgins, J., & Kawato, M. (2007). Anthropomorphism influences perception of computer-animated characters' actions. *Social Cognitive and Affective Neuroscience*, *2*(3), 206–216.

Cheetham, M. & Jancke, L. (2013). Perceptual and category processing of the uncanny valley hypothesis' dimension of human likeness: some methodological issues. *JoVE (Journal of Visualized Experiments)*, (76), e4375–e4375.

Cheetham, M., Suter, P., & Jancke, L. (2014). Perceptual discrimination difficulty and familiarity in the uncanny valley: more like a "happy valley". *Frontiers in psychology*, *5*, 1219.

Clavel, C., Plessier, J., Martin, J.-C., Ach, L., & Morel, B. (2009). Combining facial and postural expressions of emotions in a virtual character. In *Proc. of the 9th International Conference on Intelligent Virtual Agents*, IVA '09, (pp. 287–300).

Cloudy with a Chance of Meatballs. (2009). [motion picture]. Marsden, P. (Producer), Lord, P., Miller, C. (Directors). USA: Sony Pictures Animation.

Costa, P. T. & McCrae, R. R. (1992). Professional manual: revised neo personality inventory (NEO-PI-R) and neo five-factor inventory (NEO-FFI). *Odessa, FL: Psychological Assessment Resources.*

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to loftus and masson's method. *Tutorials in Quantitative Methods for Psychology*, *1*(1), 42–45.

Crane, E. & Gross, M. (2007). Motion capture and emotion: Affect detection in whole body movement. In *Affective computing and intelligent interaction* (pp. 95–101). Springer.

Cross, E. S., Ramsey, R., Liepelt, R., Prinz, W., & Hamilton, A. F. d. C. (2015). The shaping of social perception by stimulus and knowledge cues to human animacy. *371*(1686).

Cunningham, D. & Wallraven, C. (2011). *Experimental Design: From User Studies to Psychophysics*. AK Peters/CRC Press.

Cunningham, D. W., Breidt, M., Kleiner, M., Wallraven, C., & Bulthoff, H. (2003). How believable are real faces? towards a perceptual basis for conversational animation. In *Computer Animation and Social Agents, 2003. 16th International Conference on*, (pp. 23–29). IEEE.

Cunningham, D. W., Kleiner, M., Bülthoff, H. H., & Wallraven, C. (2004). The components of conversational facial expressions. In *Proceedings of the 1st Symposium on Applied perception in graphics and visualization*, (pp. 143–150). ACM.

Cupchik, G. C. (1995). Emotion in aesthetics: Reactive and reflective models. *Poetics*, *23*(1), 177–188.

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of personality and social psychology*, *44*(1), 113.

De Meijer, M. (1989). The contribution of general features of body movement to the attribution of emotions. *Journal of Nonverbal behavior*, *13*(4), 247–268.

Deaux, K. (1993). Commentary: Sorry, wrong number: A reply to gentile's call. *Psychological Science*, *4*(2), 125–126.

Demoulin, S., Cortes, B. P., Viki, T. G., Rodriguez, A. P., Rodriguez, R. T., Paladino, M. P., & Leyens, J.-P. (2009). The role of in-group identification in infra-humanization. *International Journal of Psychology*, *44*(1), 4–11.

Dewaele, J.-M. & Furnham, A. (2000). Personality and speech production: a pilot study of second language learners. *Personality and Individual Differences*, *28*(2), 355–365.

Dill, V., Flach, L. M., Hocevar, R., Lykawka, C., Musse, S. R., & Pinho, M. S. (2012). Evaluation of the uncanny valley in cg characters. In *Intelligent Virtual Agents*.

Draganski, B., Gaser, C., Busch, V., Schuierer, G., Bogdahn, U., & May, A. (2004). Neuroplasticity: changes in grey matter induced by training. *Nature*, *427*(6972), 311–312.

Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion*, *6*(3-4), 169–200.

Ennis, C., Hoyet, L., Egges, A., & McDonnell, R. (2013). Emotion capture: Emotionally expressive characters for games. In *Proceedings of Motion on Games*, (pp. 53–60). ACM.

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: a three-factor theory of anthropomorphism. *Psychological review*, *114*(4), 864.

Evans, G. W. & Howard, R. B. (1973). Personal space. *Psychological bulletin*, *80*(4), 334.

Ferrey, A. E., Burleigh, T. J., & Fenske, M. J. (2015). Stimulus-category competition, inhibition, and affective devaluation: a novel account of the uncanny valley. *Frontiers in psychology*, *6*.

Fetter, W. A. (1982). A progression ofhuman figures simulated by computer graphics. *IEEE Transactions on Computer Graphics and Applications*, 9–13.

Final Fantasy: The Spirits Within. (2001). [motion picture]. Lee, C. and Sakai, A. (Producers), Sakaguchi, H. (Director). USA: Columbia Pictures.

Fink, B., Grammer, K., & Matts, P. J. (2006). Visible skin color distribution plays a role in the perception of age, attractiveness, and health in female faces. *Evolution and Human Behavior*, *27*(6).

Fink, B. & Matts, P. J. (2008). The effects of skin colour distribution and topography cues on the perception of female facial age and health. *Journal of the European Academy of Dermatology and Venereology*, *22*(4).

Fischer, A. H. (1993). Sex differences in emotionality: Fact or stereotype? *Feminism & Psychology*, *3*(3), 303–318.

Fischer, A. H., Rodriguez Mosquera, P. M., Van Vianen, A. E., & Manstead, A. S. (2004). Gender and culture differences in emotion. *Emotion*, *4*(1), 87.

Fischer, J., Bartz, D., & Straber, W. (2005). Stylized augmented reality for improved immersion. In *Virtual Reality, 2005. Proceedings. VR 2005. IEEE*, (pp. 195–202). IEEE.

Fleming, R., Mohler, B. J., Romero, J., Black, M. J., Breidt, M., Piryankova, I., Stefanucci, J., Romero, J., de la Rosa, S., Black, M., et al. (2016). Appealing female avatars from 3d body scans: Perceptual effects of stylization. *ACM Transactions on Applied Perception*, *11*, 1–18.

Franks, C. M. (1963). Ocular movements and spontaneous blink rate as functions of personality. *Perceptual and motor skills*, *16*(1), 178–178.

Frijda, N. H. (1988). The laws of emotion. *American psychologist*, *43*(5), 349.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138.

Garau, M. (2003). *The impact of avatar fidelity on social interaction in virtual environments*. PhD thesis, University College London.

Garau, M., Slater, M., Vinayagamoorthy, V., Brogni, A., Steed, A., & Sasse, M. A. (2003). The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, (pp. 529–536). ACM.

Gawronski, B., Ehrenberg, K., Banse, R., Zukova, J., & Klauer, K. C. (2003). It's in the mind of the beholder: The impact of stereotypic associations on category-based and individuating impression formation. *Journal of Experimental Social Psychology*, *39*(1), 16–30.

Gilbert, D. T. (1998). Ordinary personology. *The Handbook of Social Psychology*, *2*, 89–150.

Gilbert, D. T., Pelham, B. W., & Krull, D. S. (1988). On cognitive busyness: When person perceivers meet persons perceived. *Journal of Personality and Social Psychology*, *54*(5), 733.

Gill, A. & Oberlander, J. (2003). Perception of e-mail personality at zero-acquaintance: Extraversion takes care of itself; neuroticism is a worry. In *Proceedings of the 25th Annual Conference of the Cognitive Science Society*, (pp. 456–461). Citeseer.

Golan, Ofer, B.-C. S. & Hill, J. (2006). The cambridge mindreading (cam) face-voice battery: Testing complex emotion recognition in adults with and without asperger syndrome. *Journal of Autism and Developmental Disorders*, *36*(2), 169–183.

Goldberg, L. R. (1990). An alternative "description of personality": the big-five factor structure. *Journal of personality and social psychology*, *59*(6), 1216.

Gooch, B., Reinhard, E., & Gooch, A. (2004). Human facial illustrations: Creation and psychophysical evaluation. *ACM Transactions on Graphics (TOG)*, *23*(1), 27–44.

Gosling, S. D., Rentfrow, P. J., & Swann Jr, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in personality*, *37*(6), 504–528.

Gouraud, H. (1971). Continuous shading of curved surfaces. *IEEE Transactions on Computers*, *100*(6), 623–629.

Green, R. D., MacDorman, K. F., Ho, C.-C., & Vasudevan, S. (2008). Sensitivity to the proportions of faces that vary in human likeness. *Computers in Human Behavior*, *24*(5).

Greenberg, J., Pyszczynski, T., Solomon, S., Simon, L., & Breus, M. (1994). Role of consciousness and accessibility of death-related thoughts in mortality salience effects. *Journal of Personality and Social Psychology*, *67*(4), 627.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of Personality and Social Psychology*, *74*(6), 1464.

Grelotti, D. J., Klin, A. J., Gauthier, I., Skudlarski, P., Cohen, D. J., Gore, J. C., Volkmar, F. R., & Schultz, R. T. (2005). fMRI activation of the fusiform gyrus and amygdala to cartoon characters but not to faces in a boy with autism. *Neuropsychologia*, *43*(3), 373–385.

Haeske, A. B. (2016). The uncanny valley: Involvement of fast and slow evaluation systems.

Hall, E. T. (1966). The hidden dimension.

Hall, J. A. (1979). Gender, gender roles, and nonverbal communication skills. *Skill in nonverbal communication*, 32–67.

Haller, M., Landerl, F., & Billinghurst, M. (2005). A loose and sketchy approach in a mediated reality environment. In *Proceedings of the 3rd international conference on Computer graphics and interactive techniques in Australasia and South East Asia*, (pp. 371–379). ACM.

Han, S., Jiang, Y., Humphreys, G. W., Zhou, T., & Cai, P. (2005). Distinct neural substrates for the perception of real and virtual visual worlds. *NeuroImage*, *24*(3), 928–935.

Hard Woman. (1985). [video file]. Jagger, M. [Screen Name]. Retrieved from https://www.youtube.com/watch?v=IcejzV-KGPI.

Harris, L. T. & Fiske, S. T. (2006). Dehumanizing the lowest of the low neuroimaging responses to extreme out-groups. *Psychological science*, *17*(10), 847–853.

Haslam, N. (2006). Dehumanization: An integrative review. *Personality and social psychology review*, *10*(3), 252–264.

Hayduk, L. A. (1983). Personal space: Where we now stand. *Psychological bulletin*, *94*(2), 293.

Herbelin, B., Riquier, F., Vexo, F., & Thalmann, D. (2002). Virtual reality in cognitive behavioral therapy: a preliminary study on social anxiety disorder. In *VSMM2002*, number VRLAB-CONF-2007-036.

Hess, U., Adams, R. B., & Kleck, R. E. (2004). Facial appearance, gender, and emotion expression. *Emotion*, *4*(4), 378–388.

Hill, H., Jinno, Y., & Johnston, A. (2003). Comparing solid-body with point-light animations. *Perception*, *32*(5), 561–566.

Hill, H. & Johnston, A. (2001). Categorizing sex and identity from the biological motion of faces. *Current Biology*, *11*(11), 880–885.

Ho, C.-C. & MacDorman, K. F. (2010). Revisiting the uncanny valley theory: Developing and validating an alternative to the godspeed indices. *Computers in Human Behavior*, *26*(6), 1508–1518.

Hobson, R. P., Ouston, J., & Lee, A. (1988). What's in a face? the case of autism. *British Journal of Psychology*, *79*(4), 441–453.

Hodgins, J., Jörg, S., O'Sullivan, C., Park, S. I., & Mahler, M. (2010). The saliency of anomalies in animated human characters. *ACM Transactions on Applied Perception (TAP)*, *7*(4).

Hodgins, J. K. (1996). Three-dimensional human running. In *Robotics and Automation, 1996. Proceedings., 1996 IEEE International Conference on*, volume 4, (pp. 3271–3276). IEEE.

Hugill, N., Fink, B., & Neave, N. (2010). The role of human body movements in mate selection.

Hyde, J., Carter, E., Kiesler, S., & Hodgins, J. (2013). Perceptual effects of damped and exaggerated facial motion in animated characters. In *Proceedings of the 10th IEEE International Conference on Automatic Face and Gesture Recognition.*

Jack, D. R., Garrod, M. O., & Schyns, P. (2014). Dynamic facial expressions of emotion transmit an evolving hierarchy of signals over time. *Current Biology*, *24*(2).

Jensen, H. W., Marschner, S. R., Levoy, M., & Hanrahan, P. (2001). A practical model for subsurface light transport. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, (pp. 511–518). ACM.

Jimenez, J. (2012). Separable subsurface scattering. In *ACM SIGGRAPH 2012 Computer Animation Festival*, SIGGRAPH '12, (pp. 82–82). ACM.

Jimenez, J., Echevarria, J. I., Oat, C., & Gutierrez, D. (2011). Practical and realistic facial wrinkles animation. *GPU Pro*, *2*, 15–27.

Jimenez, J. & Gutierrez, D. (2010). *GPU Pro: Advanced Rendering Techniques*, chapter Screen-Space Subsurface Scattering, (pp. 335–351). AK Peters Ltd.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics*, *14*(2), 201–211.

John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative big five trait taxonomy. *Handbook of personality: Theory and research*, *3*, 114–158.

Johnson, K. L., McKay, L. S., & Pollick, F. E. (2011). He throws like a girl (but only when he's sad): Emotion affects sex-decoding of biological motion displays. *Cognition*, *119*(2), 265–280.

Johnson, K. L. & Tassinary, L. G. (2005). Perceiving sex directly and indirectly meaning in motion and morphology. *Psychological Science*, *16*(11), 890–897.

Jörg, S., Hodgins, J., & O'Sullivan, C. (2010). The perception of finger motions. In *Proceedings of the 7th Symposium on Applied Perception in Graphics and Visualization*, (pp. 129–133).

Joshi, P., Tien, W. C., Desbrun, M., & Pighin, F. (2005). Learning controls for blend shape based realistic facial animation. In *ACM SIGGRAPH 2005 Courses*, (pp.$\tilde{8}$). ACM.

Jurassic Park. (1993). [motion picture]. Kennedy, K., Molen, G.R. (Producers), Spielberg, S. (Director). USA: Universal Pictures.

Kätsyri, J., Förger, K., Mäkäräinen, M., & Takala, T. (2015). A review of empirical evidence on different uncanny valley hypotheses: support for perceptual mismatch as one road to the valley of eeriness. *Frontiers in Psychology, 6*.

Klauer, K. C., Voss, A., & Stahl, C. (2012). *Cognitive methods in social psychology.* Guilford Press.

Klehm, O., Rousselle, F., Papas, M., Bradley, D., Hery, C., Bickel, B., Jarosz, W., & Beeler, T. (2015). Recent advances in facial appearance capture. *Computer Graphics Forum (Proceedings of Eurographics)*, *34*(2).

Kobayashi, H., Ichikawa, Y., Senda, M., & Shiiba, T. (2003). Realization of realistic and rich facial expressions by face robot. In *Intelligent Robots and Systems, 2003.(IROS 2003). Proceedings. 2003 IEEE/RSJ International Conference on*, volume 2, (pp. 1123–1128). IEEE.

Kokkinara, E. & McDonnell, R. (2015). Animation realism affects perceived character appeal of a self-virtual face. *Proceedings of the ACM conference on Motion in Games*, *In Press*.

Kozlowski, L. T. & Cutting, J. E. (1977). Recognizing the sex of a walker from a dynamic point-light display. *Perception & Psychophysics*, *21*(6), 575–580.

Kramer, A. & Buck, L. A. (1997). Encountering people labeled “schizophrenic”. *Journal of Humanistic Psychology*, *37*(4), 12–29.

Krikorian, D. H., Lee, J.-S., Chock, T. M., & Harms, C. (2000). Isn’t that spatial?: Distance and communication in a 2-d virtual environment. *Journal of Computer-Mediated Communication*, *5*(4), 0–0.

Kringelbach, M. L. (2005). The human orbitofrontal cortex: linking reward to hedonic experience. *Nature Reviews Neuroscience*, *6*(9), 691–702.

Lander, K. & Bruce, V. (2004). Repetition priming from moving faces. *Memory & cognition*, *32*(4), 640–647.

Lee, J. M., Ku, J. H., Jang, D. P., Kim, D. H., Choi, Y. H., Kim, I. Y., & Kim, S. I. (2002). Virtual reality system for treatment of the fear of public speaking using image-based rendering and moving pictures. *Cyberpsychology & behavior*, *5*(3), 191–195.

Lewis, M., Haviland-Jones, J., & Barrett, L. (2008). *Handbook of Emotions, Third Edition*. Guilford Publications.

Li, H., Weise, T., & Pauly, M. (2010). Example-based facial rigging. In *ACM Transactions on Graphics (TOG)*, volume 29, (pp.$\tilde{3}$2). ACM.

Little, A. C. & Hancock, P. J. (2002). The role of masculinity and distinctiveness in judgments of human male facial attractiveness. *British Journal of Psychology*, *93*(4).

Loomis, J., Silva, J. D., Fujita, N., & Fukusima, S. (1992). Visual space perception and visually directed action. *Journal of Experimental Psychology Human Perception and Performance*, *18*(4), 906–921.

MacDorman, K. F. (2005). Mortality salience and the uncanny valley. In *5th IEEE-RAS International Conference on Humanoid Robots, 2005.*, (pp. 399–405). IEEE.

MacDorman, K. F., Vasudevan, S. K., & Ho, C.-C. (2009). Does japan really have robot mania? comparing attitudes by implicit and explicit measures. *AI & society*, *23*(4), 485–510.

Macrae, C. N. & Bodenhausen, G. V. (2000). Social cognition: Thinking categorically about others. *Annual review of psychology*, *51*(1), 93–120.

Macrae, C. N., Bodenhausen, G. V., Schloerscheidt, A. M., & Milne, A. B. (1999). Tales of the unexpected: executive function and person perception. *Journal of personality and social psychology*, *76*(2), 200.

Magnenat-Thalmann, N. & Thalmann, D. (2005). *Handbook of virtual humans*. John Wiley & Sons.

Mar, R. A., Kelley, W. M., Heatherton, T. F., & Macrae, C. N. (2007). Detecting agency from the biological motion of veridical vs animated agents. *Social cognitive and affective neuroscience*, *2*(3), 199–205.

Marschner, S. R., Jensen, H. W., Cammarano, M., Worley, S., & Hanrahan, P. (2003). Light scattering from human hair fibers. In *ACM Transactions on Graphics (TOG)*, volume 22, (pp. 780–791). ACM.

Martin, T., Woflgang, E., Thibieroz, N., Yang, J., & Lacroix, J. (2014). Tressfx: Advanced real-time hair rendering. *GPU Pro*, *5*, 193–209.

Matsumoto, K., Shibata, S., Seiji, S., Mori, C., & Shioe, K. (2010). Factors influencing the processing of visual information from non-verbal communications. *Psychiatry and clinical neurosciences*, *64*(3), 299–308.

McBride, G., King, M., & James, J. (1965). Social proximity effects on galvanic skin responses in adult humans. *The Journal of psychology*, *61*(1), 153–157.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychological review*, *102*(3), 419.

McCloud, S. (1993). Understanding comics: The invisible art. *Northampton, Mass.*

McCroskey, J. C. & McCain, T. A. (1974). The measurement of interpersonal attraction.

McDonnell, R., Breidt, M., & Buelthoff, H. (2012). Render me real Investigating the effect of render style on the perception of animated virtual humans. *ACM Transactions on Graphics*, *31*(4), 91:1–91:11.

McDonnell, R., Jörg, S., Hodgins, J. K., Newell, F., & O'sullivan, C. (2009). Evaluating the effect of motion and body shape on the perceived sex of virtual characters. *ACM Transactions on Applied Perception (TAP)*, *5*(4), 20.

McDonnell, R. & O'Sullivan, C. (2010). Movements and voices affect perceived sex of virtual conversers. In *Proceedings of the 7th Symposium on Applied Perception in Graphics and Visualization*, (pp. 125–128). ACM.

Meehan, M., Insko, B., Whitton, M., & Brooks Jr, F. P. (2002). Physiological measures of presence in stressful virtual environments. In *ACM Transactions on Graphics (TOG)*, volume 21, (pp. 645–652). ACM.

Mehl, M. R., Gosling, S. D., & Pennebaker, J. W. (2006). Personality in its natural habitat: manifestations and implicit folk theories of personality in daily life. *Journal of personality and social psychology*, *90*(5), 862.

Mehrabian, A. (1980). Basic dimensions for a general psychological theory implications for personality, social, environmental, and developmental studies.

Minato, T., MacDorman, K. F., Shimada, M., Itakura, S., Lee, K., & Ishiguro, H. (2004). Evaluating humanlikeness by comparing responses elicited by an android and a person. In *Proceedings of the 2nd International Workshop on Man-Machine Symbiotic Systems*, (pp. 373–383).

Misselhorn, C. (2009). Empathy with inanimate objects and the uncanny valley. *Minds and Machines*, *19*(3), 345–359.

Mitchell, J. P., Banaji, M. R., & Macrae, C. N. (2005). General and specific contributions of the medial prefrontal cortex to knowledge about mental states. *Neuroimage*, *28*(4), 757–762.

Monsters, Inc. (2001). [motion picture]. Anderson, D. K. (Producer), Docter, P. (Director). USA: Pixar Animation Studios.

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to cousineau (2005). *reason*, *4*(2), 61–64.

Mori, M. (1970). The uncanny valley. *Energy*, *7*(4), 33–35.

Morrison, E. R., Gralewski, L., Campbell, N., & Penton-Voak, I. S. (2007). Facial movement varies by sex and is related to attractiveness. *Evolution and Human Behavior*, *28*(3), 186–192.

Musek, J. (2007). A general factor of personality: Evidence for the big one in the five-factor model. *Journal of Research in Personality*, *41*(6), 1213–1233.

Naumann, L. P., Vazire, S., Rentfrow, P. J., & Gosling, S. D. (2009). Personality judgments based on physical appearance. *Personality and Social Psychology Bulletin*, *35*(12), 1661–1671.

Neff, M., Wang, Y., Abbott, R., & Walker, M. (2010). Evaluating the effect of gesture and language on personality perception in conversational agents. In *Intelligent Virtual Agents*, (pp. 222–235). Springer.

Nehab, D., Rusinkiewicz, S., Davis, J., & Ramamoorthi, R. (2005). Efficiently combining positions and normals for precise 3d geometry. *ACM Trans. Graph.*, *24*(3), 536–543.

Nelder, J. A. & Baker, R. J. (1972). Generalized linear models. *Encyclopedia of Statistical Sciences*.

North, M. (1975). *Personality assessment through movement*. Plays, Incorporated.

Nowak, K. & Biocca, F. (2003). The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments. *Presence*, *12*(5), 481–494.

Osofsky, M. J., Bandura, A., & Zimbardo, P. G. (2005). The role of moral disengagement in the execution process. *Law and Human Behavior*, *29*(4), 371.

Parke, F. I. (1974). A parametric model for human faces. Technical report, DTIC Document.

Patel, H. (2015). *The persuasiveness of humanlike computer interfaces varies more through narrative characterization than through the uncanny valley*. PhD thesis.

Paunonen, S. V. & Ashton, M. C. (2001). Big five factors and facets and the prediction of behavior. *Journal of personality and social psychology*, *81*(3), 524.

Paunonen, S. V. & Hong, R. Y. (2013). The many faces of assumed similarity in perceptions of personality. *Journal of Research in Personality*, *47*(6), 800–815.

Pertaub, D., Slater, M., & Barker, C. (2001). An experiment on fear of public speaking in virtual reality. *Studies in Health Technology and Informatics*, 372–378.

Piwek, L., McKay, L. S., & Pollick, F. E. (2014). Empirical evaluation of the uncanny valley hypothesis fails to confirm the predicted effect of motion. *Cognition*, *130*(3), 271–277.

Plant, E. A., Hyde, J. S., Keltner, D., & Devine, P. G. (2000). The gender stereotyping of emotions. *Psychology of Women Quarterly*, *24*(1), 81–92.

Plutchik, R. (2001). The nature of emotions human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American Scientist*, *89*(4), 344–350.

Pocoyo. (2004). [television series]. Matas, C., Cubria, P. (Producers), Carsi, G.G., Cantolla, D., Rodriguez, A. (Directors). Spain UK: ITV, Granada Kids and Granada International.

Pollick, F. E., Paterson, H. M., Bruderlin, A., & Sanford, A. J. (2001). Perceiving affect from arm movement. *Cognition*, *82*(2), B51–B61.

Quattrone, G. A. (1982). Overattribution and unit formation: When behavior engulfs the person. *Journal of personality and social psychology*, *42*(4), 593.

Rameson, L. T. & Lieberman, M. D. (2009). Empathy: A social cognitive neuroscience approach. *Social and Personality Psychology Compass*, *3*(1), 94–110.

Ramey, C. H. (2005). The uncanny valley of similarities concerning abortion, baldness, heaps of sand, and humanlike robots. In *Proceedings of views of the uncanny valley workshop: IEEE-RAS international conference on humanoid robots*, (pp. 8–13).

Rao, R. P. & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87.

Rauthmann, J. F., Seubert, C. T., Sachse, P., & Furtner, M. R. (2012). Eyes as windows to the soul: gazing behavior is related to personality. *Journal of Research in Personality*, *46*(2), 147–156.

Ready, R. E., Clark, L. A., Watson, D., & Westerhouse, K. (2000). Self-and peer-reported personality: Agreement, trait ratability, and the "self-based heuristic". *Journal of Research in Personality*, *34*(2), 208–224.

Ring, L., Utami, D., & Bickmore, T. (2014). The right agent for the job? In *International Conference on Intelligent Virtual Agents*, (pp. 374–384). Springer.

Rosegrant, T. J. & McCroskey, J. C. (1975). The effects of race and sex on proxemic behavior in an interview setting. *Southern Journal of Communication*, *40*(4), 408–418.

Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. *Advances in Experimental Social Psychology*, *10*, 173–220.

Sagi, Y., Tavor, I., Hofstetter, S., Tzur-Moryosef, S., Blumenfeld-Katzir, T., & Assaf, Y. (2012). Learning in the fast lane: new insights into neuroplasticity. *Neuron*, *73*(6), 1195–1203.

Saygin, A., Chaminade, T., Ishiguro, H., Driver, J., & Frith, C. (2012). The thing that should not be: Predictive coding and the uncanny valley in perceiving human and humanoid robot actions. *Social Cognitive Affective Neuroscience*, *7*(4).

Schacter, D., Gilbert, D. T., & Wegner, D. M. (2011). *Psychology (2nd Edition)*. New York: Worth.

Schneider, E., Wang, Y., & Yang, S. (2007). Exploring the uncanny valley with japanese video game characters. In *Situated Play: Proceedings of the Digital Games Research Association Conference*.

Schyns, P. G., Petro, L. S., & Smith, M. L. (2009). Transmission of facial expressions of emotion co-evolved with their efficient decoding in the brain: behavioral and brain evidence. *PLoS One*, *4*(5), e5625.

Serfass, D. G. & Sherman, R. A. (2013). Personality and perceptions of situations from the thematic apperception test. *Journal of Research in Personality*, *47*(6), 708–718.

Seyama, J. & Nagayama, R. S. (2007). The uncanny valley: Effect of realism on the impression of artificial human faces. *Presence: Teleoperators and Virtual Environments*, *16*(4), 337–351.

Sherman, J. W., Lee, A. Y., Bessenoff, G. R., & Frost, L. A. (1998). Stereotype efficiency reconsidered: Encoding flexibility under cognitive load. *Journal of Personality and Social Psychology*, *75*(3), 589.

Singer, T. (2006). The neuronal basis and ontogeny of empathy and mind reading: review of literature and implications for future research. *Neuroscience & Biobehavioral Reviews*, *30*(6), 855–863.

Smith, E. R. & DeCoster, J. (2000). Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review*, *4*(2), 108–131.

Sommer, R. (2002). Personal space in a digital age. *Handbook of Environmental Psychology*, 647–660.

Sousa, T., Kasyan, N., & Schulz, N. (2012). Cryengine 3: Three years of work in review. *GPU PRO 3: Advanced Rendering Techniques*, 133.

Tangled. (2010). [motion picture]. Conli, R., Lasseter, J., Keane, G. (Producers), Greno, N., Howard, B. (Directors). USA: Walt Disney Studios.

Terminator 2: Judgment Day. (1991). [motion picture]. Cameron, D. (Producer), Cameron, D. (Director). USA: TriStar Pictures.

The Adventures of Tintin. (2011). [motion picture]. Spielberg, S., Jackson, P., Kennedy, K. (Producers), Spielberg, S. (Director). USA: Sony Pictures Animation.

The Lawnmower Man. (1992). [motion picture]. Everett, G. (Producer), Leonard, B. (Director). USA: New Line Cinema.

The Lord of the Rings: Fellowship of the Ring. (2001). [motion picture]. Osborne, B.M., Jackson, P., Walsh, F., Sanders, T. (Producers), Jackson, P. (Director). USA, New Zealand: New Line Cinema.

The Polar Express. (2004). [motion picture]. Starkey, S., Zemeckis, R., Goetzman, G., Teitler, W. (Producers), Zemeckis, R. (Director). USA: Warner Bros. Pictures.

Thompson, J. C., Trafton, J. G., & McKnight, P. (2011). The perception of humanness from the movements of synthetic agents. *Perception*, *40*(6), 695–704.

Tinwell, A., Grimshaw, M., Nabi, D. A., & Williams, A. (2011). Facial expression of emotion and perception of the uncanny valley in virtual characters. *Computers in Human Behavior*, *27*(2), 741–749.

Tinwell, A., Grimshaw, M., & Williams, A. (2011). The uncanny wall. *International Journal of Arts and Technology*, *4*(3), 326–341.

Tinwell, A., Nabi, D. A., & Charlton, J. P. (2013). Perception of psychopathy and the uncanny valley in virtual characters. *Computers in Human Behavior*, *29*(4), 1617–1625.

Toy Story. (1995). [motion picture]. Guggenheim, R., Arnold, B. (Producers), Lasseter, J. (Director). USA: Pixar Animation Studios.

Trope, Y. (1986). Identification and inferential processes in dispositional attribution. *Psychological review*, *93*(3), 239.

Unger, R. K. (1979). Toward a redefinition of sex and gender. *American Psychologist*, *34*(11), 1085–94.

Volante, M., Babu, S. V., Chaturvedi, H., Newsome, N., Ebrahimi, E., Roy, T., Daily, S. B., & Fasolino, T. (2016). Effects of virtual human appearance fidelity on emotion contagion in affective inter-personal simulations. *IEEE transactions on visualization and computer graphics*, *22*(4), 1326–1335.

Volkova, E. P., Mohler, B. J., Dodds, T. J., Tesch, J., & Bülthoff, H. H. (2014). Emotion categorization of body expressions in narrative scenarios. *Frontiers in psychology*, *5*.

Wallraven, C., Breidt, M., Cunningham, D. W., & Bülthoff, H. H. (2008). Evaluating the perceptual realism of animated facial expressions. *ACM Transactions on Applied Perception*, *4*(4).

Wallraven, C., Bülthoff, H. H., Cunningham, D. W., Fischer, J., & Bartz, D. (2007). Evaluation of real-world and computer-generated stylized facial expressions. *ACM Transactions on Applied Perception*, *4*(3).

Ward, K., Bertails, F., Kim, T.-Y., Marschner, S. R., Cani, M.-P., & Lin, M. C. (2007). A survey on hair modeling: Styling, simulation, and rendering. *IEEE Transactions on Visualization and Computer Graphics*, *13*(2), 213–234.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the panas scales. *Journal of personality and social psychology*, *54*(6), 1063.

Weise, T., Bouaziz, S., Li, H., & Pauly, M. (2011). Realtime performance-based facial animation. *ACM Transactions on Graphics*, *30*(4).

Willis Jr, F. N. (1966). Initial speaking distance as a function of the speakersâ€™ relationship. *Psychonomic Science*, *5*(6), 221–222.

Winnemöller, H., Olsen, S. C., & Gooch, B. (2006). Real-time video abstraction. In *ACM Transactions On Graphics (TOG)*, volume 25, (pp. 1221–1226). ACM.

Wisessing, P., Dingliana, J., & McDonnell, R. (2016). Perception of lighting and shading for animated virtual characters. In *Proceedings of the ACM Symposium on Applied Perception*, (pp. 25–29). ACM.

Yamada, Y., Kawabe, T., & Ihaya, K. (2013). Categorization difficulty is associated with negative evaluation in the â€œuncanny valleyâ€ phenomenon. *Japanese Psychological Research*, *55*(1), 20–32.

Yang, P.-F., Laszlo, J., & Singh, K. (2004). Layered dynamic control for interactive character swimming. In *Proceedings of the 2004 ACM SIGGRAPH/Eurographics symposium on Computer animation*, (pp. 39–47). Eurographics Association.

Zell, E. & Botsch, M. (2013). Elastiface: Matching and blending textured faces. In *Proceedings of the Symposium on Non-Photorealistic Animation and Rendering*.

Zhu, J.-Y., Agarwala, A., Efros, A. A., Shechtman, E., & Wang, J. (2014). Mirror mirror: Crowdsourcing better portraits. *ACM Transactions on Graphics*, *33*(6).