Department of

Human Sciences for Education “Riccardo Massa”

Ph.D. program in Education in Contemporary Society, Cycle XXXIII

DIGITAL HUMANITY.

Do Users' Gaming Habits Affect the Perceived Human-Likeness of Virtual Agents in a Simulated Human Interaction?

Surname FERRARI Name AMBRA

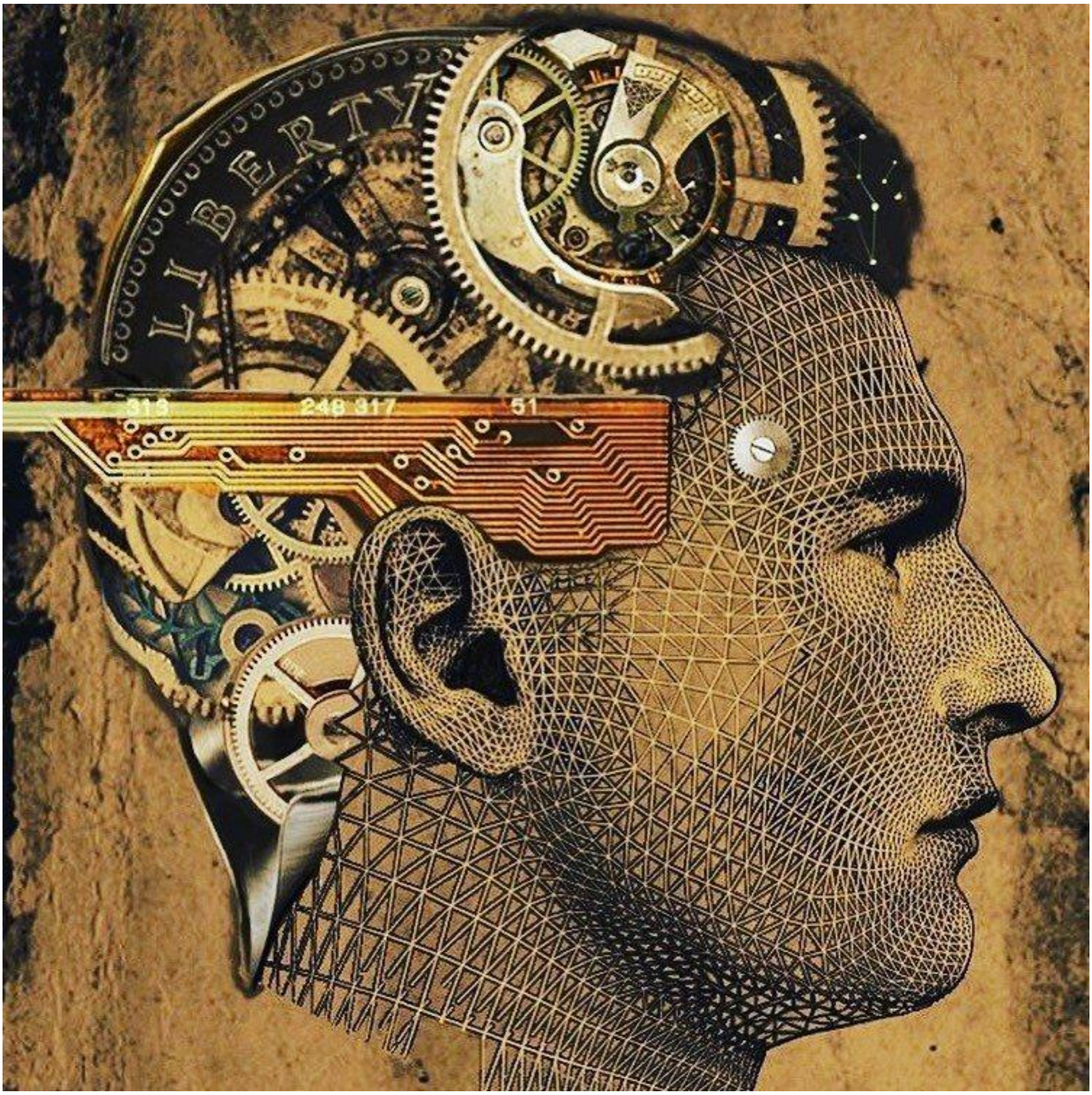
Registration number 737607

Supervisor: FABRIZIA MANTOVANI

Coordinator: FRANCESCA ANTONACCI

ACADEMIC YEAR 2019/2020

Cover Art: "Connor" by Facial Animation Supervisor @mce_rk800
Prologue Art: Detroit Become Human artwork from Quantic Dream Games (2018)


LIBERTY
248 317
51

[this page intentionally left blank]

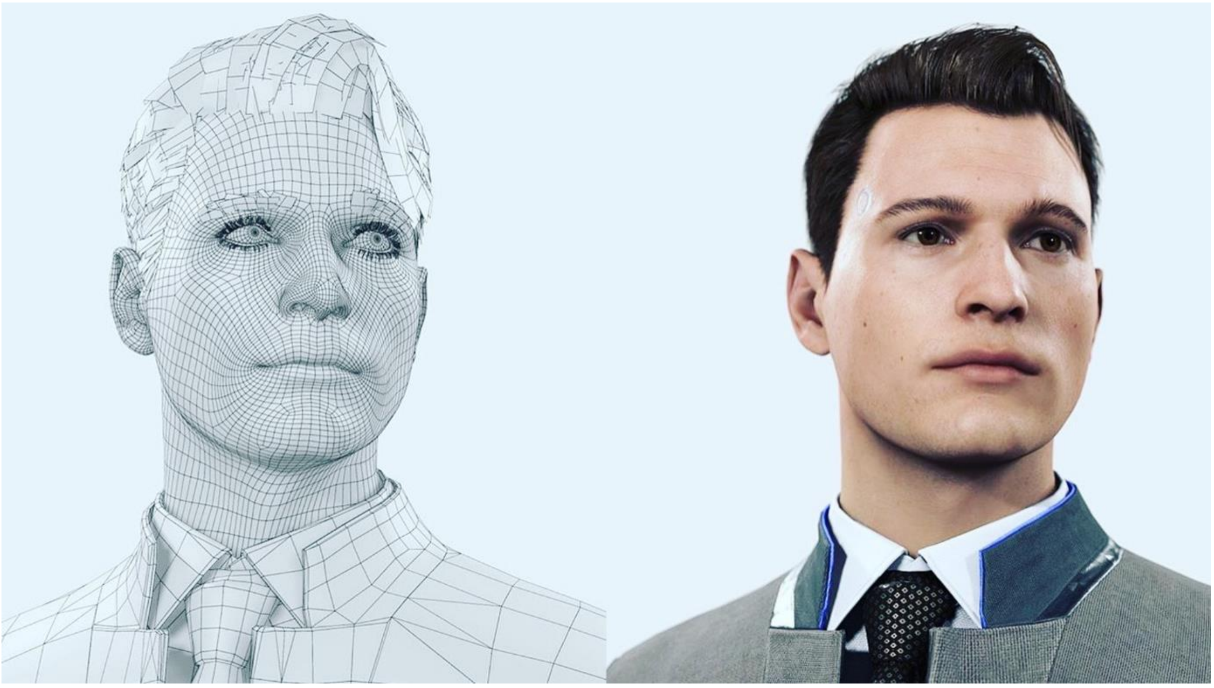
TABLE OF CONTENTS

SECTION 1 - Theoretical Framework
Virtual Agents

SECTION 2 - Theoretical Framework
Negotiation Training in the XXI Century

SECTION 3 - Empirical Works

[this page intentionally left blank]



"You cannot kill me. I am not *alive*, Lieutenant."

Connor,

Detroit: Become Human (Quantic Dream, 2018)

[this page intentionally left blank]

Prologue

What does it mean to be human?

Such a question has bugged the mind of Mankind for millennia, prancing through philosophy, religion, biology, and physics. We may think to narrow down such a problem to the definition of indispensable human qualities regarding physical appearance and intellectual talents, such as intelligence, responsivity, or the ability to interact and socialize with other humans. Nevertheless, it is anything but simple to pinpoint the very essence of such qualities and there are many people affected by psychological or physical disorders who have impediments when it comes to such attributes or abilities. However, it is abhorrent even to think that we should regard them as "not human."

During the last decades, technological advancements have been pushing the already unresolved question of humanity even further. Research laboratories worldwide have been experimenting on the possibility of creating increasingly realistic artificial human bodies, either electric or hydraulic powered. If paired with Artificial Intelligence (AI), defined as a simulation of human intelligence through machines, technology may lead to the creation of androids, human-level intelligent robots with a fully human appearance.

The 2018 blockbuster video game by Quantic Dream, *Detroit: Become Human*, has tackled the issue: should androids have rights? Should they be paid for their work? Do they deserve humans' respect? The video game also attempts to answer such questions, suggesting that Humanity ultimately lies in the freedom of choice, including choosing anger and violence.

Despite what science fiction and media may lead us to believe, androids are still far from reality. According to Neuralink CEO Elon Musk, we will have to wait at least until the 2050s for human-like intelligent robots to roam the Earth. Nonetheless, new challenges to humanity's concept will eventually be born. There is an increasingly urgent need to reason with technology and humanity, as we are already live in close contact with sub-human artificial intelligence.

Such a kind of AI differs from general intelligence because it is hyper-specialized and, therefore, can complete a restricted number of tasks. Deep Blue, the IBM computer, won a chess match against Garry Kasparov in 1997 and Google Duplex, the 2018 virtual assistant on our smartphones, is capable of calling restaurants and making reservations for us. Hyper-specialized sub-human artificial intelligence has been a consistent part of our lives for decades.

Although intelligent androids with physical bodies still represent a challenge for engineers, digital sub-human intelligent beings such as non-playable characters in video games are technical realities. Though non-physical, virtual characters are embodied representations of human beings who do not exist in our world, yet they exist in – and are consistent with – the space of digital worlds, i.e., the space of possible worlds (Monnin, 2010; Schröter & Thon, 2018). The issue of Digital Humanity is already part of our everyday life, as players have started forming affective relationships with virtual characters, defined as "parasocial relationships" (Coulson, Barnett, Ferguson, & Gould, 2012).

At a certain point in *Detroit: Become Human*, the human Lieutenant Anderson asks Connor, a robotic detective, a crucial question: "Why did they make you look so goofy and give you that weird voice?". Connor responds: "CyberLife androids are designed to work harmoniously with humans. Both my appearance and voice were specifically designed to facilitate my integration". Gaining inspiration from what the fictional company CyberLife has achieved, the present dissertation's objective is, therefore, to begin to scratch the surface of what it takes for a virtual being to be recognized as Human.

Though such interrogation may seem strictly theoretical, it does have several practical applications, especially considering its urgent nature. In the context of the present work, it is worth considering that the use of digitally Simulated Human Interactions with Virtual Agents is continually rising in companies seeking to spice up effective communication training. In this context, Virtual Agents' creation with realistic human qualities is essential for the training process's success and generalizability. The issue is further complicated because the Entertainment video game industry posits higher and higher human-likeness standards through its characters, creating an increasingly wide gap between them and small independent developers such as research laboratories.

Introduction

I. Virtual Worlds and Virtual Agents

During the last half-century, computer-based simulated environments have become increasingly common in many fields, from cinematography to advertisement, entertainment, and education. Among the sea of different virtual environments, a discriminant characteristic lies in the degree of interactivity between the user and the virtual world. While several virtual environments cannot be interacted with (e.g., digital renderings of worlds in movies), in some other cases, users can interact with the digital world by looking in different directions and manipulating in-world objects.

This interactivity can be obtained through pressing a button on a controller or through motions with a varying degree of sensorimotor correspondence between the actions performed in the real world and their translation in the virtual world. For instance, Sony's PlayStation4 Dualshock controller's motion sensor allows for simple movements with a vague resemblance to real gestures (e.g., washing dishes or cradling newborns). Motion-controlled consoles, such as Nintendo Switch, allow realistic tracking of real-world gestures (e.g., milking a cow, throwing a bowling ball). The chance to realistically interact with virtual worlds is ever increasing. For what is more, the last decades have seen the introduction to the mass market of Virtual Reality, "a computer-generated reality, which allows a learner or group of learners to experience various auditory and visual stimuli (…) through the use of specialized ear and eyewear" (INACSL Standards Committee, 2016). Users can interact with objects in such immersive virtual spaces using a variety of different devices. From controllers allowing users to move each finger independently (e.g., Oculus Touch) up to experimental Virtual Reality Systems tracking the user's tendons movements through sensors placed on gloves (Fig.I, following page), the chance to interact with virtual worlds realistically is ever increasing. Realistic interactions can sustain embodiment, which is experienced when an illusion of non-mediation ensues in a virtual environment (Jung, 2009), and the player experiences an almost physical connection with their avatars, i.e., their digital representation in the virtual world (van Looy 2012; Ratan, 2015).

A further step into such an illusion is immersion, defined as the feeling of becoming part of the virtual experience (De Castell, 2007) to the point that it "takes over (…) our whole perceptual apparatus" (Murray, 1997). When feeling immersed in a virtual environment, the user can feel present in it, allowing real emotions to be activated (Parsons & Rizzo, 2008; Price, Mehta, Tone & Anderson, 2011).

Figure I. Realistic sensory-motor interactive virtual worlds in virtual reality: Haptic gloves let the user feel virtual objects and perform actions with high fidelity to reality (HaptX, 2017).

Most importantly, for the sake of this dissertation, users can also interact with a variety of Embodied Virtual Agents, digital visual representations of computer interfaces represented by human or animal bodies (Cassell, 2000). Among the many different classifications found in the relevant literature (see Table I, following page), Virtual Agents acquire the adjective "Conversational" when capable of simulating a face-to-face conversation with users. Such Virtual Agents have abilities such as recognizing and responding to input, turn-taking, and contributing to various degrees to the conversation topic with new information and utterances, in written or oral form (Cassell, 2000). Contrary to Intelligent Virtual Assistants, which may not have a digital body and are only capable of responding to a single query submitted by the user (McTear, Callejas & Griol, 2016), Conversational Virtual Agents can simulate conversations at various degrees of similarity with the ones between human dyads, including their emotional, behavioral, and cognitive elements (Gratch, Wang, Gerten, Fast, & Duffy, 2007).

ENTITY	DEFINITION
Avatar	"A digital representation of a human user that facilitates interaction with other users, entities, or the environment" (Nowak & Fox, 2018)
Embodied Virtual Agents	"Computer interfaces represented by way of human or animal bodies (…) which human or animal bodies appear lifelike or believable in their actions and their reactions to human users." (Cassell, 2000)
Intelligent Virtual Assistants	"Interfaces that enable users to interact with smart devices using spoken language in a natural way and provide a singular response to a query similar to speaking to a person." (McTear, Callejas & Griol, 2016)
Embodied Conversational Agents	"Virtual embodied representations of humans that communicate multimodally with the user through voice, facial expression, gaze, gesture, and body movement." (Hartmann, Mancini, & Pelachaud, 2005)

Table I. Definitions of different kinds of Virtual Entities

Conversational Agents can also be categorized into several different classes based on the intelligence supporting forms of communication increasingly independent from human intervention. For instance, some Virtual Agents' intelligence is based on the wizarding method. Such a method requires the presence of a real person, often hidden from the user, who is in charge of choosing which actions the Virtual Agent should perform during an interaction, among a series of possible options (e.g., Gratch, DeVault, Lucas, & Marsella, 2015). This method takes its name from L. Frank Baum's 1900 children's book *The Wonderful Wizard of Oz*, as a person is hidden behind a computer just as, in the world of Oz, a person hides behind the (in)famous curtain. The second kind of Virtual Agents' intelligence is "scripted", i.e., based on the Agent's programming code based on computer-generated choices among the Agent's pool of behaviors. Being independent of human intervention once programmed, the benefits of this kind of Agent are more comprehensive than Wizarding-based Agents, as they can be made available at any time and place (Ruijten, 2015). There is yet another category, which is Autonomous Agents. An Autonomous Agent represents a highly realistic representation of a person whose behavior is controlled by Artificial Intelligence and possesses technology-based senses (Blascovich & Bailenson, 2011; Straßmann & Krämer, 2018). Artificial intelligence (AI) is a simulation of human intelligence through machines, especially computer systems. As AI processes include learning (the acquisition of information and rules for using such information), reasoning (using rules to reach conclusions), and self-correction (Shanahan, 2015), these Agents are not only independent from human intervention once a learning algorithm is provided, but also potentially able to improve themselves without further human intervention.

Independently from their degree of Intelligence, some Virtual Agents can, therefore, be capable of simulating a conversation with their human users to various degrees of human-likeness. The consequent Simulated Human Interactions can be used as training methods in various settings, as described in the following paragraphs.

II. More Than Conversational Partners: Virtual Agents as Trainers

The design and creation of realistic digitally simulated Virtual Agents have become a prerequisite for the success of many applications, from non-interactive contents (such as VFX in the movie industry) to human-machine interactions such as video games and Serious Games (Ruhland et al., 2015). Virtual Agents have been increasingly used in a variety of areas, as well as in a variety of roles. For instance, they can act as mentors and deliverers of notions in the context of e-learning University programs or on-the-job safety seminars, allowing organizations to save money and experts to save time. Similarly, they can be used as drivers of positive behavioral change, persuading and encouraging users to take on new positive habits, such as staying motivated for a daily physical exercise or learning how to act more ecologically in everyday life (e.g., Ruijten, de Kort, & Kosnar, 2012).

Moreover, Virtual Agents have been the subject of recent developments in health-related assistance, especially for people in need of support such as elderly or cognitively impaired people. Human-like 2D Agents and robots have been introduced in retirement homes and private houses to remind patients and guests at which time to take their medications (e.g., Yaghoubzadeh, Kramer, Pitsch, & Kopp, 2013), to monitor them and alert a nurse in case of sudden movements, thanks to motion sensors detecting falling, and even to play games and chat with them (e.g., European Commission, 2018). Even though these categories of people often cannot do without other people's help, Virtual Agents can still assist them in some daily tasks, sustaining a more self-determined life, when possible (Straßmann & Krämer, 2018).

Virtual Agents can also act as conversational partners with young and non-cognitively impaired people, either online (e.g., Cleverscript's "Eviebot", www.eviebot.com) or in the context of specific and organized virtual experiences for the enhancement and training of effective communication strategies. This particular set of skills refers to the ability of people to effectively communicate through verbal and non-verbal components (Gibson & Sodeman, 2014) and to successfully work and interact with people of diverse ethnicities, social statuses, and cultural backgrounds, building and maintaining a healthy and positive relationship with them (Kivunja, 2015). Especially over the last decades, the Western world is becoming more and more connected and prone to the movement of great masses of people (Mezzaroba & Silveira, 2018; Serzhanova, 2011). The consequent necessity of addressing the blossoming social demands of a world set in the XXI century has driven attention towards effective communication and Life Skills in general, defined as positive and versatile behaviors enabling people to successfully adapt to an ever-changing world (World Health Organization, 1994; Jain & Anjuman, 2013). Nonetheless, there seems to be a gap between what contemporary formal education equips young adults with, which often still focuses on conveying information and techniques (Black & Wiliam, 2005), and "those skills that they need to build a future for themselves" (Wyn, 2009). Such a consideration poses a challenge to high school and university education alike, even if to different extents (Tynjala, Slotte, Nieminen, Lonka, & Olkinuora, 2006; Wyn, 2009). This translates into the fact that, while young people display abilities such as skilled technology usage, flexibility, and multitasking skills (Özçelik, 2015; Robinson & Stubberud, 2012; Martin, 2005), young generations seem to lack abilities in terms of face-to-face interactions - while being comfortable and effective interacting via technology (Hershatter & Epstein, 2010; Robinson & Stubberud, 2012). There is, therefore, a growing need for efficient training methods capable of addressing Life Skills, as effective communication strategies result to be relevant for the individual's general well-being (Vincent, 2005) but also as a requirement in the XXI-century workplace (Bedwell, Salas, & Fiore, 2010; National Research Council, 2012).

Despite a general tendency of schools and universities not to provide students with appropriate Life Skills training, there have been several exceptions to this trend since the 1980s, involving frontal lectures of different duration intended to improve students' self-confidence in communication skills (e.g., Taylor, 2008; Coleman & Lim, 2001). Frontal lectures are described as one of the most commonly used methods to teach with, even though there is a risk for the trainees not to be able to actively participate in the lecture, increasing the danger of attention decrease during the course (Castagna, 2014).

Role-playing is another Life Skills training method. With the objective of training communication skills, human relations cases are generally proposed to groups for discussion after being introduced by experts in charge of conducting a debriefing after the end of the session (Bales et al., 1951). With the advantage of involving expert people's know-how, all these traditional techniques can nonetheless present different disadvantages. First of all, experts are not always readily available, and the monetary cost of paying them should not be underestimated (Lewin & Gollan, 2018). Moreover, companies need to make sure expert evaluations are consistent with each other, without being negatively affected by the expert's fatigue due to the same exercise's repetition on different occasions. Finally, traditional methods are time-consuming for both experts and trainees (Kron et al., 2017).

According to some authors, technological innovations represent an effective need-supporting method to respond to many of the disadvantages mentioned above. Rather than traditional training techniques, technological innovations such as e-learning, serious games, videos, or chat rooms, represent compelling interactive learning opportunities (Ozcelik, 2015). Many technology-based teaching methods not only provide consistent, personalized feedback (Van Keuren, 2006), but they also include an entertainment aspect in the learning process (Chelliah & Clarke, 2011; Weigel, James, & Gardner, 2009). Such entertainment aspect appears to be particularly relevant, as a substantial core of literature states that it seems to increase embeddedness into a team (Tews, Michel, Xu, & Drost, 2015), task performance, creative performance, and engagement, especially for the Millennial generation (Fluegge-Woolf, 2014). Technologically enhanced methods can be useful training tools for younger generations, intrinsically motivating to a technologically-savvy generation to effectively use electronic devices to access and process information (Gibson & Sodeman, 2014). Many technology-based teaching methods provide continuous feedback instead of traditional post-training-only assessments (Farrell & Hurt, 2014; Van Keuren, 2006). This represents a further advantage, as younger generations tend to be "continuous learners", expecting high-quality feedback to continuously evaluate their progress, leading with clear directions (Martin, 2009; Brack & Kelly, 2012).

Among technologically enhanced Life Skills training methods, we can find e-learning methods, videos, and interactions in which users can interact with each other (e.g., Greco & Murgia, 2007) or with an embodied Virtual Agent (e.g., Ross, Pollman, Perry, Welty, & Jones, 2001) through a chat system based on written communication. By sustaining interactivity without an expert's need, these methods are deemed more effective than traditional training systems. Nonetheless, they do not represent the gold-standard for technologically enhanced Life Skills training, as they lack information regarding the interlocutor's non-verbal communication components, such as facial expressions, body language, gestures, voice volume, or intonation. These, instead, represent essential information and clues regarding a variety of internal states of both the talker and the listener (e.g., Hömke, Holler, & Levinson, 2017, 2018; Ruhland et al., 2014; Wood et al., 2015).

It is in this context that Simulated Human Interactions come into play. Using Simulated Human Interactions as training seems to improve the users' competencies more efficiently and flexibly if compared to traditional teaching methods (e.g., Azadegan, Riedel, & Hauge, 2012), thanks to the combination of playful features and simulation aspects (Anolli & Mantovani, 2011).

Regarding the latter, a complete and accurate interactive simulation involving Virtual Agents' ability to communicate with the user through facial expressions, speech, and gestures sustains learning of complex scripts. A general human disposition lies in anticipating future phenomena and, consequently, adjusting strategies to tackle them (Anolli & Mantovani, 2011) effectively. This kind of learning can be defined as global because of its involving theory and practice and perception, cognition, and motor aspects at the same time, differently from traditional teaching methods (Raybourn, 2007). Moreover, simulations offer trainees interactive and realistic situations to act in while avoiding the risks of operating in a real-life context with tangible consequences, which might be costly for the institution providing the training in case of grave mistakes (Cano & Sáenz, 1999; Datar, Garvin, & Cullen, 2010). Thus, Simulated Human Interactions provide experiential insights thanks to trial-and-error and learning-by-doing methodologies, in the safety of a realistic yet harmless context (Laborde Torres, 2016).

As for the former component, game-like features have been long associated with highly motivated learning (Gee, 2003), and in particular with the so-called "state of flow". Flow has been defined as the complete engagement in a particular activity, "an ordered state of mind that is highly enjoyable" (Csikszentmihalyi, 1990). While particular personality traits seem to be related to an increased chance of entering a state of flow (i.e., people presenting a high level of the "Conscientiousness" trait as measured by the Big Five Questionnaire; Ilies et al., 2017), this state can be achieved whenever an individual engages with an activity complying with a specific set of characteristics. It provides clear goals and immediate feedback for the person's actions, is effortless and intrinsically rewarding, proposes balanced challenges to the individual's skills, and therefore it is perceived as easy to control (Csíkszentmihályi, 1990). Therefore, the person is wholly concentrated on the task to the point that they lose track of time and self-consciousness about one's inadequacies in other aspects of life.

The strict relationship between flow experiences and learning outcomes has been robustly demonstrated by literature during the last decades (e.g., Choi & Baek, 2011; Csikszentmihalyi, 1990; Hoffman & Novak, 2009; Webster, Trevino, & Ryan, 1993) in teenagers and young adults (Rathunde & Csikszentmihalyi, 2005), but also for higher education (Ghani, 1995). Moreover, the relaxation induced by the state of flow and the "safe environment" provided by simulation aspects enables trainees to learn things more easily and effectively (Almeida et al., 2011; Popescu et al., 2013), encouraging a positive attitude from trainees towards the simulation (Zgodavova, Kisela, & Sutoova, 2016). Flow, therefore, represents an essential feature for virtually any successful learning process (Almeida et al., 2011; Popescu, Romero, & Usart, 2013): thanks to the intense concentration and enjoyment that arise from such a state (Ghani & Deshpande, 1994), it fosters the trainee's efforts without causing bore or resentment, enabling them to learn things more easily and effectively (Almeida et al., 2011; Popescu et al., 2013).

III. Main Research Question and Objectives of The Present Dissertation

Replicating the level of human-likeness in the Virtual Agent's appearance and behavior required to "convince users' perceptual systems that a virtual human is the real thing" (Ruhland et al., 2015) is a contemporary challenge. In the specific context of game-based training and interactive simulations, plausible Virtual Agents have already been shown to contribute to higher immersion levels, fun, and learning among users. On the contrary, flawed Virtual Agents may negatively affect the users' perceived immersion and cause disturbing emotional reactions in users, which in turn can disrupt the simulative nature of an educational Simulated Human Interaction and impede the users' performance (Von Bergen, 2010; MacDorman & Ishiguro, 2006).

During the last decades, video games have been increasingly representing a consistent entertainment source for a growing number of young people, to the point that 56% of people who habitually play videogames are under the age of 35 (Entertainment Software Association, 2017). According to such a report, players consider the quality of graphics the most crucial aspect to consider before purchasing a video game, even above the quality of storytelling. An example of the significant consideration graphics quality is taken into lies in turmoil following the release of the expected blockbuster video game *Mass Effect: Andromeda* (BioWare, 2017). The reason for such an uprising of video game consumers (accompanied by calls for refunds) was the characters' eyes and precisely the ratio between the size of the sclera and the iris. Though negligible to the eyes of uninitiated professionals, such detail was responsible for disruptive feelings in the video game players: at the sight of the main characters, either hilarity or terror ensued in unsuspecting players. They defined the eyes as "cartoonishly large" or as "piercing through your soul". Such an unpleasant situation was worsened by the fact that *Mass Effect: Andromeda* can be considered as a story-based video game; therefore, close-ups of the weird faces were quite frequent, and the design flaw quickly became blatant and impossible to ignore (Fig II, following page).

Such a consideration has led the entertainment video game industry to a "graphics race" towards photorealism and simulation-driven design, for virtual worlds to become increasingly involving (Batt, 2015). Consequently, the industry has encouraged the acceleration of technological progress related to visual and auditory data acquisition from real-life situations, using sophisticated systems and real actors as reference.

Fig. II – The ratio between Addison's sclera and iris size before and after the Mass Effect: Andromeda patch released by BioWare.

Contemporarily, the use of digitally Simulated Human Interactions for educational purposes is continually rising in popularity because of their flexibility and motivational value. However, there are several obstacles to developing Simulated Human Interactions for educational purposes by small independent research labs, such as processing power and the economic cost of using modern techniques and specialized workers.

The first objective of the present dissertation is, therefore, to explore the urgency of the situation. In particular, the aim is to investigate whether the Entertainment Industry is truly posing higher and higher standards. In other words, if habitual video game players have harsher reactions to Simulated Human Interactions for educational purposes than non-habitual video game players. The last decades have seen the transformation of video games from a niche hobby to mainstream glory, to the point that there are an estimated 2.47 billion video game players globally (Statista, 2019). Therefore, an increasingly larger population may find Simulated Human Interactions for educational purposes inadequate, causing perceptual disparities that may contribute to unsettling emotional reactions, overruling the benefits of using such a training tool.

On the bright side, there is a vast amount of psychological knowledge about Agent Human-likeness and the Uncanny Valley that might make up for research labs' technological and economic disadvantages. The second objective of the present dissertation is to explore which realistic human qualities are essential for a Virtual Agent to be considered human-like and, therefore, explore whether it is possible to identify several "easy wins" on the short development run. In other words, this research seeks to find key elements that do not require expensive interventions in terms of money and time but can dramatically increase the perception of the product's quality, similar to what happened with the implementation of new eyes in Mass Effect: Andromeda.

This dissertation is articulated into three Sections and seven Chapters.

The first section revolves around Virtual Agents, providing the Theoretical Framework of the present dissertation. Chapter 1 is dedicated to a discussion over literature concerning Virtual Agents' perceived realism, the matter of the Uncanny Valley, and how the original theory evolved over the years. Specific focus is dedicated to intervening factors in uncanniness perception, such as individual differences in users, movement, and sounds in Virtual Agents. Chapter 2 provides a brief insight into the contemporary Entertainment video game industry. After describing contemporary market trends and game consumption habits in Europe, space is dedicated to contemporary technology and techniques used to design and create realistic Virtual Agents in commercial video games. Finally, resources and challenges encountered by research labs developing Simulated Human Interactions for educational purposes relying on limited time and budget are explored.

The second section provides a Theoretical Framework for Negotiation and Conflict Management Skills. Chapter 3 defines Negotiation as the primary target Life skill of the present dissertation, stating the importance of being an effective negotiator in the context of both personal and professional life. Chapter 4 provides several examples of training for Negotiation Skills based on Simulated Human Interactions. It then presents the dissertation's target Simulated Human Interaction, ENACT, its structure, functions, and the available Virtual Agents.

The final section is dedicated to empirical works. Chapter 5 describes the first experimental study. The study's objective has been to explore whether gameplay habits affect users' perception of an Agent's human-likeness and whether the introduction of random eye blinks positively contributes to the perceived human-likeness of the Agent. Chapter 6 is dedicated to the second experimental study of the present work. The objective has been to explore whether the Agent's gibberish voice positively contributes to its perceived human-likeness and whether gameplay habits affect users' perception. Finally, theoretical and practical take-aways of the present dissertation and Future Directions are debated in Chapter 7.

DIGITAL HUMANITY.
Do Users' Gaming Habits Affect the Perceived Human-Likeness of Virtual Agents in a Simulated Human Interaction?

SECTION 1

Theoretical Framework

Virtual Agents

[this page intentionally left blank]

Chapter 1
Digital Humanity

Virtual Agents, digital visual representations of computer interfaces represented by human bodies (Cassell, 2000), have been increasingly used as game-based training and interactive content for educational purposes, simulating a face-to-face conversation with users. Using Simulated Human Interactions as training improves the users' communication competencies more efficiently and flexibly compared to traditional teaching methods (e.g., Azadegan, Riedel, & Hauge, 2012). This is partially thanks to playful features and because Virtual Agents are capable of simulating complex and multimodal communicative behaviors to various degrees of human-likeness (Cassell, 2000), supporting real-world generalizability of what is learned in the digital world (Anolli et al., 2011).

Plausible Virtual Agents have been shown to contribute to higher levels of immersion, fun, and learning. However, replicating a satisfactory level of human-likeness in a Virtual Agent remains a challenge to this day (Ruhland et al., 2015). To respond to such a challenge, the entertainment video game industry has encouraged the acceleration of technological progress related to visual and auditory data acquisition from real-life situations. However, similar sophisticated systems and specialized workers are often unavailable to small independent research labs, increasing the pressure on an often already fatigued resource system. Far from being a mere aesthetic issue, flawed Virtual Agents are likely to negatively affect the users' perceived immersion and cause disturbing emotional reactions in users, disrupting the simulative nature of the Simulated Human Interaction and impeding the users' performance and learning process (Von Bergen, 2010; MacDorman & Ishiguro, 2006).

Therefore, this Chapter will trace preliminary knowledge regarding the topic of Realism of Virtual Agents, starting from the complex and multilayered operationalization of the term itself. The Uncanny Valley phenomenon will then be introduced, along with its original definition and the debate over the source of such disruptive feelings, describing evolutionary, cognitive, or philosophical stances regarding its origins. Finally, the Chapter will describe the successive redefinitions of the concept and intervening factors identified over the years, such as Virtual Agents' movements and vocal utterances, as well as individual differences in uncanniness perception.

1.1 Operationalizing Realism

According to Literature, the degree of realism of a Virtual Agent can heavily affect the degree of perceived appeal and friendliness of the Agent and predict the extent to which users will build relationships with it (Novick & Gris, 2014; Ring, Utami, & Bickmore, 2014; Terada, Jing, & Yamada, 2015). Generally speaking, a more realistic Virtual Agent leads, in fact, to more positive Simulated Interactions compared to Agents presented with a lower degree of realism (Yee, Bailenson, & Rickertsen, 2007). This is particularly true for older people, as they are more likely not to be familiar with technology and might need a higher degree of naturality as a motivation to engage with Virtual Agents (Straßmann & Krämer, 2017).

A challenge in this regard lies in the operationalization of the term "Realism" in the context of Virtual Agents. Users seem to evaluate the Virtual Agent's face attractiveness using the same principles they use with human interlocutors in real life (Sobieraj & Krämer 2014). Nonetheless, Virtual Agents' likeness contains much more factors than that of a human, since its visual appearance and behavior are intrinsically and deliberately based on technological resources and design aspects (Straßmann & Krämer, 2017). The concept of realism in the context of Virtual Agents is, therefore, complex and multilayered, and many dimensions have to be taken into consideration (Wang, Lilienfeld, & Rochat, 2015; Bartneck, Kanda, Ishiguro, & Hagita, 2009).

An operationalization of the concept of "Realism" might be the term "Anthropomorphism", here defined as the users' perception of the degree of credibility the Agent would have in an offline world, to the point that they would socially respond to the Agent at a spontaneous level (Nowak & Fox 2018; Nass & Moon, 2000; Roubroeks, Ham, & Midden, 2011). In this dissertation's context, anthropomorphism is, therefore, conceptualized as the result of an interaction with a Virtual Agent instead of a predisposition of the user to attribute human-like qualities to Artificial Agents (e.g., Lemaignan, Fink, Dillenbourg, & Braboszcz, 2014).

To support perceived anthropomorphism, the Virtual Agent should display typically human characteristics and uniquely human ones (Ruijten, Bouten, Rouschop, Ham, & Midden, 2014). The former set of characteristics refers to features and skills which are generally associated with human beings but are also displayed by other living beings, such as the ability to dynamically respond to the environment (Haslam, Loughnan, Kashima, & Bain, 2008). On the contrary, uniquely human characteristics include both cognitive - such as moral decision-making skills – and non-cognitive skills - such as fine manipulation skills. Uniquely human characteristics are exclusive to humans and are more difficultly attributed to Virtual Agents than typically human ones (Ruijten, Bouten, Rouschop, Ham, & Midden, 2014).

1.2 The Uncanny Valley

Although several studies agree that more realistic Virtual Agents lead to a more positive evaluation of the quality of the Interaction compared to Agents presented with a lower degree of realism (e.g., Yee, Bailenson, & Rickertsen, 2007), this might be a simplification. For instance, results from comparisons among a spectrum of different degrees of realism, ranging from line drawing to highly realistic, indicate that users perceived both Agents with a low realism and those with a high realism as equally appealing (McDonnell, Breidt & Bülthoff, 2012). According to some literature, such results are linked to a specific phenomenon, usually referred to as the Uncanny Valley.

The concept of "uncanny" was first referred to during the first years of the Twentieth Century. German psychiatrist Jentsch first used the term in connection to the frightening feeling a viewer might experience when observing dolls, automata, or wax figures resembling humans, unable to solidly distinguish if the objects are to be considered "animate or inanimate" (Jentsch, 1906). Subsequently, Jentsch also used the term to describe the feeling experienced when witnessing an ongoing epileptic seizure. In such a situation, the interlocutor is aware that the affected person no longer has control over their own body and, therefore, they find themselves confused and frightened, unable to predict how the affected person will behave in the following minutes (Jentsch, 1906).

Nonetheless, Mori's theorization is known for having systematized the Uncanny Valley theory basics, here defined as the abrupt decline in humans' perceived familiarity of a stimulus when it nears perfection (Mori, 1970; Mori, MacDorman, & Kageki, 2012). The Uncanny Valley refers to the observer's unpleasant impression of a robotic or virtual being with an almost, but not entirely, realistic human form (Seyama & Nagayama, 2007). In other words, virtual beings whose movements, behavior, and appearance are almost but not wholly regarded as human are likely to cause a sense of repulsion or rejection in the user (Hodgins, Jörg, O'Sullivan, Park, & Mahler, 2010; Nowak & Fox 2018), to the point where the character's attractiveness and human-like appearance result to be negatively correlated (Scheneider, Wang, & Yang, 2007; Fig.1).

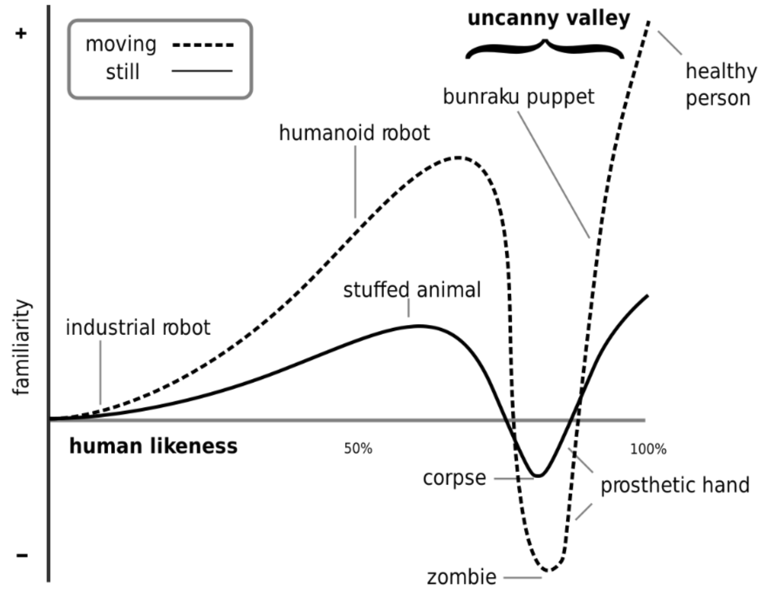


Figure 1. Mori's theorization of the Uncanny Valley (Mori, 1970).

This implies that increasing human-likeness is associated with more familiarity only until it exceeds a certain threshold. At that point, there is a sharp drop in the perceived familiarity of the Agent before becoming better again (Kätsyri, Förger, Mäkäräinen, & Takala, 2015). In theory, the state of the highest perceived familiarity - associated with more positive experiences - is achieved when in contact with a completely believable Agent, where the differences with real humans are impossible to detect. However, a relatively high level can also be reached with moderate realism, just before the valley (Seymour, Riemer, & Kay, 2017). Therefore, the relationship between the Agent's human-likeness degree and the user's perceived familiarity (accompanied by positive feelings) is very non-linear in fashion.

Exploring the cognitive foundations of the Uncanny Valley, starting from the very basics of "how people perceive each other" (Wang, Lilienfeld, & Rochat, 2015) results to be of importance for the sake of better understanding the nature of the uncanny feeling, as well as to provide more precise definitions and reliable measurements of human-likeness beyond superficial physical similarities. Several theories can be found in the literature regarding the possible cause of the Uncanny Valley phenomenon.

First of all, fMRI data have demonstrated the existence of an anatomical specialization for the recognition and elaboration of human faces, the fusiform face area (FFA). Such data sustain the hypothesis that the Uncanny Valley might be caused by specialized perceptual processing (Andrews, Schluppeck, Homfray, Matthews, & Blakemore, 2002; Meng, Cherian, Singal, & Sinha, 2012; MacDorman & Chattopadhyay, 2016). Interestingly, macaque monkeys have shown more prolonged eye fixation on unrealistic synthetic faces and real faces if compared to realistic synthetic faces, suggesting a degree of preference for the former two and, possibly, a form of uncanny valley effect in primates (Steckenfinger & Ghazanfar, 2009). The unpleasant phenomenon might, therefore, be caused by an evolutionary-based, automatic, stimulus-driven threat-avoidance mechanism, triggering a "fight or flight" response (Ohman, 1993) towards possible pathogens or poor health (MacDorman, Green, Ho, & Koch, 2009; Dill, Flach, Hocevar, Lykawka, Musse, & Pinho, 2012). In the optic of mate selection and personal survival, the primal pathogen avoidance mechanism leans towards an abundance of false negatives, fuelling a sense of fear and disgust in the viewer while watching an imperfect human, and therefore causing withdrawal from possible illness, lesions, and deformities (Burleigh, Schoenherr, & Lacroix, 2013; MacDorman & Chattopadhyay, 2016).

Alternatively, the sense of eeriness may arise from the fact that the user may regard a highly realistic but not perfect Virtual Agent as a member of the "non-human" and "human" category simultaneously (Cheetham, Suter, & Jäncke, 2011). If an Agent looks almost but not entirely human, it will elicit a person's detailed normative expectations of a human, simultaneously violating several of its norms, especially if the stimuli are close to the category boundary (Feldman, Griffiths, & Morgan, 2009). The hypothesis connected to categorization ambiguity is supported by the fact that the slowest response times in a human/non-human identification task coincide with characters on the left and right threshold of the boundary (e.g., Pisoni and Tash, 1974; de Gelder, Teunisse, & Benson, 1997), in this case, Virtual Agents with inconsistently artificial and human-like features (e.g., MacDorman, Green, Ho, & Koch, 2009; Pollick, 2010). This occurrence may cause a so-called "cognitive dissonance", i.e., two conflicting multisensory perception cues held simultaneously, resulting in an increased risk of perceiving uneasiness and disruptive feelings (Festinger, 1962; Burleigh, Schoenherr, & Lacroix, 2013; Moore, 2012). Biologically, two main pathways link information from the sensory organs (thalamus) to the emotional system (amygdala): a direct pathway, receiving rough information on visual stimuli, and an indirect pathway, which includes a perception process as it involves a recognition system (cortex). Therefore, it is hypothesized that the feeling of eeriness is derived from contradictory information provided by the low-level path, indicating that the Agent is human, and the high-level path, which does not (Shimada, Minato, Itakura, & Ishiguro, 2018). When exposed to Virtual Agents, users' fMRI scans revealed a non-linear activation of the dorsomedial prefrontal cortex, a region linked to mental state attributions, and in the fusiform gyrus, the function of which is to discriminate between living and non-living entities (Rosenthal-von der Pütten, Krämer, Maderwald, Brand, & Grabenhorst, 2019).

Finally, a "Terror Management Theory" has been hypothesized by various authors (e.g., MacDorman, Green, Ho, & Koch, 2009; Dill, Flach, Hocevar, Lykawka, Musse, & Pinho, 2012). The "terror" which the human perceives while in contact with a human-like entity that is not *alive* may arise from the subconscious salience of humans' mortality, a subliminal reminder of their death (Solomon, Greenberg, & Pyszczynski, 1998). A second hypothesis traces the origin of such terror to the perception of a human-like entity as a threat, either to human safety (Kang, 2009) or to human uniqueness (MacDorman, 2006; Dill, Flach, Hocevar, Lykawka, Musse, & Pinho, 2012; Funk, 2018). In the former case, the entity is perceived as dominant because it is perceived as "beyond human control" (Kang, 2009). In the latter case, the uneasiness felt by users might derive from the disruption brought to our knowledge of "the essence of humanity" (Kang, 2009), potentially leading to a technological singularity, i.e., an existential crisis related to the moment in which the user feels "human affairs as we understand them to come to an end" when in contact with an alien yet seemingly intelligent being (Shanahan, 2015). Indeed, it has to be underlined that, whichever underpinning reasons sustain the Uncanny Valley, they should not be regarded as a reason to attribute the users' adverse reactions to uncanny Agents to the users themselves. Instead, unpleasant feelings should always be attributed to the Virtual Agent's shortcomings, continually aiming to advance the Agents' design rules (Bartneck, Kanda, Ishiguro, & Hagita, 2009).

Since the Seventies, several authors have been revisiting and revising Mori's original theorization. First of all, there has been a debate regarding the translation of Mori's original term to label the vertical axis. According to several authors, although the term "shinwa-kan" cannot be directly translated into English, it can be roughly regarded as "the sense of being mutually friendly" or "the sense of having a similar mind" (Bartneck, Kanda, Ishiguro, & Hagita, 2009). Based on such a version, more precise translations rather than "familiarity" have been proposed, i.e., "affinity" or "likeability" (Kätsyri, Förger, Mäkäräinen, & Takala, 2015;

Bartneck, Kanda, Ishiguro, & Hagita, 2009). Such novel translations seem to be appropriate as, according to a study, the perceived familiarity and likeability of a Virtual Agent resulted in a highly correlated correlation, suggesting that the two concepts were closely related. Nevertheless, "likeability" is deemed more informative regarding Virtual Agents. That is because, although people indeed tend to prefer familiar options compared to unknown ones when given a choice, "this does not mean that they will like all of the options they know" (Bartneck, Kanda, Ishiguro, & Hagita, 2009).

Secondly, it has been proposed that the feelings perceived by users exposed to an uncanny Agent might not be unavoidably negative. For instance, people might perceive an Agent as strange or surreal without necessarily being frightened by it. This seems to be especially true in artistic contexts, in which people might experience a sense of wonder when observing a being that only somewhat resembles a human (Hanson, Olney, Pereira, & Zielke, 2005).

Finally, it has also been suggested that the Uncanny Valley is a theoretical model that is too simple to be informative, considering the complexity and dimensionality of an Agent's human-likeness (Wang, Lilienfeld, & Rochat, 2015; Bartneck, Kanda, Ishiguro, & Hagita, 2009). Therefore, the following paragraphs will describe previous literature regarding several intervening factors in the perception of uncanniness. In particular, several studies have underlined the presence of individual differences in the degree of predisposition to anthropomorphize an Agent (e.g., Epley, Waytz, & Cacioppo, 2007; Kätsyri, Förger, Mäkäräinen, & Takala, 2015). Moreover, static elements of a Virtual Agent, such as skin grain quality and muscles (e.g., MacDorman, Green, Ho, & Koch, 2009), as well as dynamic elements (e.g., body and face movements, Saygin, Chaminade, Ishiguro, Driver, & Frith, 2012) have been shown to affect the perceived human-likeness of an Agent, and they will also be debated in the following pages. Lastly, the presence of an "Audio Uncanny Valley" has been suggested by some authors, and a paragraph will be dedicated to a brief review of the central literature.
Further details about intervening factors will be treated in the following paragraphs.

1.3 Intervening Factors in the Perception of the Uncanny Valley

1.3.1 Individual Differences

As previously stated, exploring the cognitive foundations of the Uncanny Valley, starting from the very basics of "how people perceive each other" (Wang, Lilienfeld, & Rochat, 2015) results to be a rich source of information for the sake of better understanding the nature of the uncanny feeling. In this context, the exploration of individual differences in feelings of uncanniness gains particular importance (Kanai & Rees, 2011).

Results from several previous studies have given evidence that feeling of anthropomorphism might not be universal and personality traits or psychological states can cause individual differences in the degree of predisposition to anthropomorphize an Agent (e.g., Epley, Waytz, & Cacioppo, 2007; Kätsyri, Förger, Mäkäräinen, & Takala, 2015). For instance, the perceived anthropomorphism as measured by the Individual Differences in Anthropomorphism Questionnaire (IDAQ) seems to decrease if the neural correlates of mental state attributions (amygdala) are in any way damaged, or in subjects on the autistic spectrum (Waytz, Cacioppo, & Epley, 2010). IDAQ results also seem to accurately predict the intensity of the trust put into the Agent and of the concern towards the Agent, resulting in feelings of morality revolving around the ancient query "can it suffer?" (Bentham, 1843; Shanahan, 2015).

Moreover, people with a lower inclination towards effortful cognitive activities tend to be more likely to attribute human-like characteristics to objects than people with a higher inclination (Epley, Waytz, & Cacioppo, 2007). This may be linked to the fact that when objects (including animals and fruits) are hard to categorize, they are perceived negatively, as they decrease processing fluency and raise a "cognitive dissonance", i.e., two conflicting multisensory perception cues that are held simultaneously (Yamada, Sasaki, Kunieda, & Wada, 2014).

This categorization-difficulty hypothesis assumes a two-stage process. People who have a stronger tendency to avoid novel experiences also seem to have more difficulties in attributing human-like characteristics to Virtual Agents and, on the contrary, to find eeriness in hard-to-categorize stimuli (Yamada, Sasaki, Kunieda, & Wada, 2014). This might be because uncategorized new stimuli are more likely to be seen as a threat by people opposing novel experiences, in avoidance would serve as a defensive mechanism. Notably, a correlation has been found between traits that support the avoidance of novel foods and the degree to which hard-to-categorize objects are perceived negatively (Yamada, Kawabe, & Ihaya, 2012).

Finally, another contributing factor lies in the user's degree of social connectedness: a higher degree of anthropomorphism is perceived by users who are chronically lonely as opposed to those who are habitually connected with others (e.g., Epley, Akalis, Waytz, & Cacioppo, 2008; Waytz, Cacioppo, & Epley, 2010). Relatedly, introverts and users with lower emotional stability preferred the mechanical looking appearance to a greater degree than other participants, suggesting a stronger preference towards Mori's Uncanny curve (Sasaki, Ihaya, & Yamada, 2017).

In conclusion, the Uncanny Valley remains a complex phenomenon linked to the simultaneous integration of a variety of factors in neurophysiological (Cheetham et al., 2011; Saygin et al., 2012), cognitive (Seyama & Nagayama, 2007), and social approaches (MacDorman & Ishiguro, 2006; MacDorman et al., 2009b).

1.3.2 The Uncanny Valley and Static Virtual Agents

Literature regarding social categorization and social judgments in human-to-human relationships has suggested that positive impressions of the interlocutor depend on the interlocutor's visual characteristics (e.g., Clark & Rutter, 1985). For instance, in the context of work interviews, a study demonstrated how interviewers would determine whether a candidate will be hired in the first 2 minutes of their face-to-face interaction (Berg & Piner, 1990). The main takeaway is that people tend to perform social categorizations and judgments very quickly and based on what they see, although they might be unaware of the visual elements that influence them (Bartneck, Kanda, Ishiguro, & Hagita, 2009). Similarly, a Virtual Agent's perceived likeability seems to depend on specific aspects of its appearance directly.

For instance, in the context of reactions to static faces, human-like digital faces' proportions such as the amount of space between the Agent's eyes or a digital face's height seem to be particularly relevant for users' positive reactions to Virtual Agents (MacDorman, Green, Ho, & Koch, 2009). On the contrary, adverse reactions can be caused by inconsistent levels of realism in the Virtual Agent, such as grossly digital-looking eyes on an otherwise human-like face (Kätsyri, Förger, Mäkäräinen, & Takala, 2015; MacDorman et al., 2009). Moreover, the eyes' size is particularly relevant, as significantly enlarged eyes on a human-like face cause negative feelings in its observers (Seyama & Nagayama, 2007).

An imperfect form anthropomorphism (a human-like appearance, Gong & Nass, 2007) has been previously intentionally used in horror video games by pairing an almost human appearance to inhuman details to increase the perceived anxiety in the player when encountering an antagonist (Tinwell, Grimshaw, & Williams, 2010).

Famous examples are Pyramid Head (*Silent Hill 2*, Konami, 2001) or Shade / Light Woman (*The Evil Within DLC - The Assignment* and *The Consequence*, Bethesda, 2015), both fusions between human beings and non-biological elements, respectively, a metallic tetrahedron and a lamp (Fig. 2).

Figure 2. Example of human and artificial components in the design of unsettling video game characters. Shade / Light Woman (The Evil Within DLC - The Assignment and The Consequence, Bethesda, 2015).

The degree of realism of the skin's texture also results to be relevant. Human skin is not a mere surface: on the contrary, it is sustained by muscles, tendons, and bones, which apply tension and cause wrinkles when moved. Nonetheless, a higher level of graphical detail leads to a higher sensitivity of the user to said details, causing the Agent to fall into the Uncanny Valley (Tinwell, Grimshaw, Nabi, & Williams, 2011). Such an aspect is well demonstrated by the fact that the recommended degree of photorealism for skin texture results to be 75% (MacDorman, Green, Ho, & Koch, 2009).

Regarding muscles, the presence of the sternocleidomastoid results to be of importance to avoid the Uncanny Valley. Its thickness and uniqueness to mammals serve as a primary landmark defining the degree of rotation and orientation of the head. When creating movie characters that are supposed to look attractive or familiar to viewers, special effects artists make sure to include the sternocleidomastoid in the design (Pueringer, 2019b, Fig. 3a, Fig. 3b).

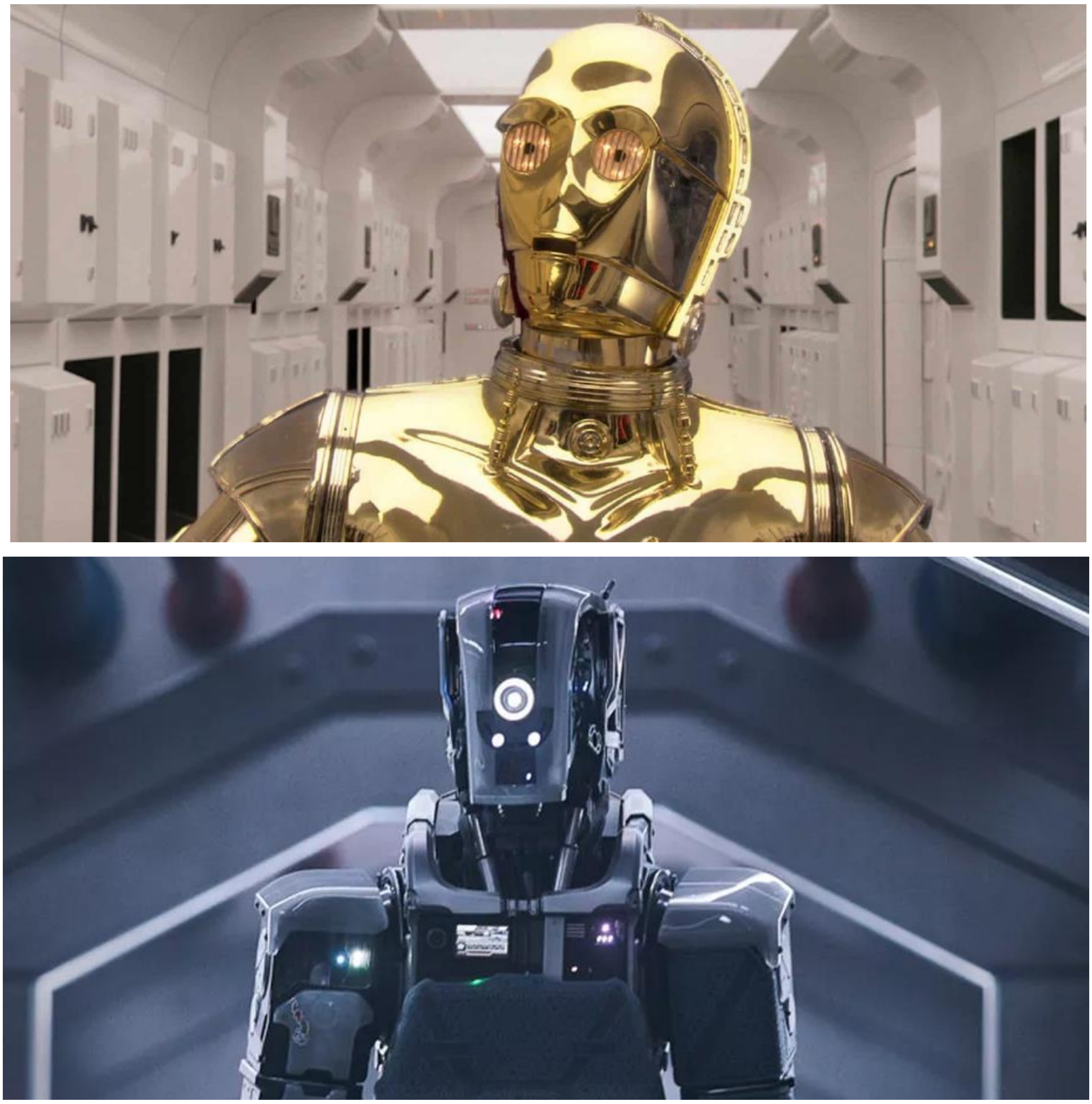


Figure 3 a and b. Examples of the inclusion of sternocleidomastoids in the design of robotic digital characters in popular movies, C3PO in the Star Wars Series (Lucasfilm, 1977), above, and Mother in the homonymous Netflix movie (Netflix, 2020), below.

1.3.3 The Uncanny Valley and Moving Virtual Agents

Although static appearance is rather crucial for the sake of perceived human-likeness, many applications move beyond it, also incorporating the Agent's movements (Damiano, Dumouchel, & Lehmann, 2014).

From a general perspective, motion is a relatively strong focusing point in human perception, as humans are able to subconsciously anticipate the movement's speed and direction in other people (Sakata, Takubo, Inoue, Nonaka, Mae, & Arai, 2004; Akiwa, Suga, Ogata, & Sugano, 2004). Such an ability results in a heightened sensitivity to discrepancies in an Agent's movements. The more an Agent tries to resemble a human being, the more it raises the expectation of behaving like a human in terms of body language and facial expressions (Lehmann, Roncone, Pattacini, & Metta, 2016; Johnson, Veltri, & Hornik, 2008), and such an assumption is graphically represented by the increased steepness of the Valley in Mori's theorization. For instance, a humanoid Agent might be regarded as human-like when static but, if its eyes move anomalously, the uncertainty regarding its human-like status might increase, causing it to be perceived as uncanny (Moore, 2012). It is worth to underline that, as several studies have suggested that characters with complex movements can be considered more human-like, more familiar, and less eerie (Piwek, McKay, & Pollick, 2014; McDonnell, Breidt, and Bülthoff, 2012; Thompson, Trafton, & McKnight, 2011), adding movement to an Agent does not inevitably lead to an increased feeling of uncanniness. Instead, there is an increased *risk* of uncanniness if users notice inaccuracies of movements.

From a neurological point of view, different areas are dedicated to the perception of static and dynamic facial features, respectively, the Fusiform Face Area (FFA) in the ventral temporal lobes (Gobbini and Haxby, 2007; Vuilleumier and Pourtois, 2007), and the inferior occipital gyrus and the superior temporal sulcus (STS) (Haxby et al., 2000; Adolphs, 2002; Ishai, 2008). Nonetheless, the link between appearance and motion may be inextricable, as the Virtual Agent's appearance dictates its expected movement, leading to a "difficulty of isolating the effect of behavior from that of appearance" (Minato, Shirmada, Ishiguro, & Itakura, 2004). Consequently, Mori's original two-dimensional theorization might not provide a sufficiently informative encapsulation of the complex and multisensory relationship between likeability and perceived human-likeness and the consequent intricate patterns of users' response (Seymour, Riemer, & Kay, 2017; (Piwek, McKay, & Pollick, 2014; Moore, 2012). Several examples of the interaction between appearance and movement can be found in recent literature. For instance, in one study, participants were asked to categorize Virtual Agents' movements with different appearances (point-lights, ellipses, robot, alien, clown, jogger) as either biological or artificial-looking. Results showed how characters with the most uncomplicated appearance (i.e., point-lights) were categorized as moving more naturally than more complex characters (Chaminade, Hodgins, & Kawato, 2007). Impressive results were also obtained from fMRI data: a higher brain activation was recorded when users watched a mechanically moving Virtual Agent if compared to the activity recorded when watching a biologically moving Virtual Agent or a mechanically moving humanoid robot. Therefore, it is hypothesized that the Uncanny Valley might be caused by a violation of perceptual expectations linking appearance to movement (Saygin, Chaminade, Ishiguro, Driver, & Frith, 2012).

The perception of uncanniness in moving Agents is furtherly affected by the user's attention. Several studies have suggested how facial anomalies are more disturbing than even significant body motion errors because humans are highly drawn towards emotional cues given by faces and, therefore, facial expressions are attentively scrutinized by the users (Hodgins, Jörg, O'Sullivan, Park, & Mahler, 2010; Lehmann, Roncone, Pattacini, & Metta, 2016; Donath, 2001). When experiencing asymmetric facial expressions, users experience emotions of lower intensity towards the Agent, except sympathy. Therefore, it is hypothesized that users might be sympathetic towards what may resemble a disease at a superficial perception level (Burleigh, Schoenherr, & Lacroix, 2013). For instance, asymmetries in facial movements may resemble Bell's palsy symptoms, paralysis of facial muscles resulting from a dysfunction of cranial nerve VII (Hodgins, Jörg, O'Sullivan, Park, & Mahler, 2010).

In human-to-human communication, young people tend to explore the top half of the interlocutor's face to search for emotional clues, especially when identifying anger, sadness, fear, and surprise (Calder, Young, Keane, & Dean, 2000; Sullivan, Campbell, Hutton, & Ruffman, 2017; Birmingham, Svärd, Kanan, & Fischer 2018). Along the same lines, downcast, upraised, and eyes looking sideways suggest states of mind in illustrated faces (Faigin, 2012). Specifically, the upper area of a digital human face (i.e., eyes, eyebrows, and forehead) provides a high number of emotional clues (Tinwell, Grimshaw, Nabi, & Williams, 2011) through elements such as pupil size, blink rate, and saccades, as well as eye movement parameters derived from Facial Action Units, basic facial movements the combinations of which univocally identify basic emotions (Ekman, 1992; Li & Mao, 2012).

Humans can accurately recognize facial expressions of five out of six basic emotions in digital faces, except for disgust, which is easily mistaken for anger (Dyck, Winbeck, Leiberg, Chen, Gur, & Mathiak, 2008). However, inadequate movements in the upper face area affect such accuracy, but they also cause uneasiness in the users. When asked to assess and recognize the emotions displayed by a virtual human face where the upper section's animations were too subtle or absent, users would mistake surprise for fear and consider uncanny many virtual expressions. The uncanniness of happiness and anger was less evident, and it is hypothesized that the lower portion of the face contributes to conveying such emotions (Tinwell, Grimshaw, Nabi, & Williams, 2011). On the contrary, fear and sadness were deemed significantly more uncanny, especially in the lack of animations condition (Tinwell, Grimshaw, Nabi, & Williams, 2011; Dill, Flach, Hocevar, Lykawka, Musse, & Pinho, 2012).

Except for older adults, who tend to look at mouths more often than eyes when searching for emotional evidence in a conversation (Sullivan, Campbell, Hutton, & Ruffman, 2017), the lower area of a face is generally explored for clues regarding articulation. In human-to-human communication, when engaging in conversations, people tend to accompany listening with lip-reading techniques to interpret what is being said (Macaluso et al., 2004; Mattys et al., 2000). If the information typically gained from observation of mouth movements is somewhat inadequate, the listener's comprehension can be impaired, leading to interpretative conflicts (Tinwell, Grimshaw, & Nabi, 2015; Szerszen, K. A. (2011). Virtual Agents' mouth articulation patterns are generally much more limited in number if compared to humans. Therefore, lip-reading Virtual Agents can lead to misrecognition of sounds that muscularly resemble the artificial articulation, causing audio/visual interpretative conflicts (McGurk and MacDonald 1976). Conflicts are likely to occur when the asynchronization between the Agent's lip movements and produced sound exceeds a range of ±400 ms. According to a study, the magnitude of a Virtual Agent's perceived uncanniness was directly proportional to the magnitude of asynchrony between lip movement and audible sound (Tinwell, Grimshaw, & Nabi, 2015). It is also worth to notice that the sensitivity to asynchrony seems to be higher if the Agent's voice is perceived before the onset of the mouth's movements: participants were able to perceive an asynchrony when auditory preceded visual input by 50 ms, but visual inputs needed to precede auditory ones by 220 ms for them to notice an asynchrony (Grant, Wassenhove, & Poeppel 2004).

1.3.4 The Uncanny Valley and Talking Virtual Agents

The concept of an audio Uncanny Valley has been introduced in the last few years, as "the closer simulated sound comes to natural sound, the more human ears pick up on the subtle differences that mark it as strange" (Ramsey, 2014).

Voice human-likeness can be operationalized into two constructs, i.e., semantic content and voice qualities. Regarding the former, results suggest that different personalities are assigned to an Agent's voice based on the wording used (Nass & Lee, 2001). For instance, a Virtual Agent capable of "small talk" and making suggestions (as opposed to orders) is perceived as friendlier and having a better personality (Garcia & Lopez, 2019). As for the latter construct, paraverbal components of the Agent's voice, such as the pitch, volume, pace, tone, and timbre, have also been shown to affect users' perception of Virtual Agents. For instance, tone shallowness and lack of emotional expression negatively affected the level of perceived friendliness in children users (Druga, Williams, Breazeal, & Resnick, 2017). Nevertheless, results over the years have been mixed.

Studies have indicated that an Agent communicating through a simple text-to-speech voice was perceived as more trustworthy than a text-only Virtual Agent (Qiu & Bensabat, 2005). According to several others, a computer-synthesized voice resulted in an inferior evaluation of the Agent as a conversational partner compared to a human voice (e.g., Mayer, 2014), even lower than the evaluation of Agents relying on purely text-based communication (Gong & Nass, 2007). While increasing the human likeness of a computer's voice did not matter in terms of perceived credibility according to an experimental study (Stern, Mullennix, &Yaroslavsky, 2006), another more recent study showed that a Virtual Agent speaking through a modern text-to-speech voice engine was rated as credible as an Agent with a human voice, outperforming an Agent speaking through an older text-to-speech voice engine (Craig & Schroeder, 2017). Finally, users showed aversion towards an Agent's voice so realistic they could not decipher whether they were talking to a real person or an Agent (Garcia & Lopez, 2019). Such results seem to suggest that an Uncanny Valley may be present in the perception of human-likeness in talking Virtual Agents, as increasing human-likeness is associated with more positive reactions (e.g., an Agent with a human voice or a modern text-to-speech voice). However, there is a sharp drop in the positivity of reactions when the Agent's voice combines elements of machine-like distortion with human naturalness (e.g., older synthetic voice).

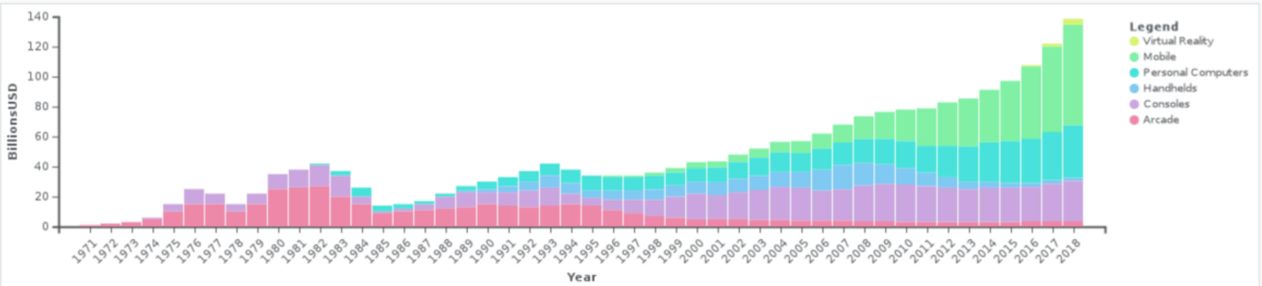
Besides paralinguistic components, adverse reactions seem to be common if there is a mismatch between the Virtual Agent's appearance and voice. For instance, the combination of a synthetic face with a synthetic voice was preferred by users compared to a visibly synthetic face and a recorded human voice, which had an adverse effect (Gong & Nass, 2000). Similarly, pairing a synthetic child-like voice with a male adult appearance ensured that a social robot fell into the Uncanny Valley (Zlotowski et al., 2015). On the contrary, users perceived Virtual Agents as more credible and attractive if their face respected human proportions and had a human voice, or a text-to-speech voice added to a face with exaggerated features (Stein & Ohler, 2018). Results seem to confirm that the level of human-likeness of a Virtual Agent's appearance and voice should match in order not to elicit feelings of eeriness and support perceived human-likeness and interpersonal warmth during the interaction (Mitchell, Szerszen Lu, Schermerhorn, Scheutz, & MacDorman, 2011).

Chapter 2

The Contemporary Entertainment Video Games Industry

During the last decades, video games have been increasingly representing a consistent entertainment source for a growing number of people. Globally, the number of video game players has risen to an astounding 2.47 billion (Statista, 2019). In Europe, 51% of the population plays video games, which equals roughly 250 million players (ISFE, 2019). As a consequence, the traditional and popular identikit of a young male video game player has changed towards a more complex profile: the average player in Europe is now 31 years old (Entertainment Software Association, 2017), and nearly half of the players are female (45%; ISFE, 2019).

As a consequence, transforming from an unusual hobby to a mainstream habit, revenue in the Video Game Industry has been steadily increasing for decades (Graph 1). In particular, the Italian share of Europe's spending on video games increased gradually, jumping from seven percent of Europe's overall spending in 2015 to about nine percent in 2019 (Statista, 2019).



Graph 1. Global revenues of the video game industry from 1971 to 2018, not adjusted for inflation (Nakamura, 2019).

The European video game industry has reached a worth of €20bn in 2019 (ISFE, 2019), driving technological innovation and progress related to video game design and production, as described in the following paragraph.

2.1 Contemporary Game Habits: Photorealism in Video Games

Over the last three years, best-selling video games have been increasingly including realistic Virtual Agents as Avatars and Non-Playable Characters, with just a few exceptions (Table 1).

2018	2019	2020
1. Red Dead Redemption 2 (Rockstar Games)	1. Call of Duty: Modern Warfare 2019 (Activision Blizzard)	1. Call of Duty: Modern Warfare (Activision Blizzard)
2. Call of Duty: Black Ops 4 (Activision Blizzard)	2. NBA 2K20 (2K Games)	2. Animal Crossing: New Horizons (Nintendo)**
3. NBA 2K19 (2K Games)	3. Madden NFL 20 (Electronic Arts)	3. Madden NFL 21 (Electronic Arts)
4. Madden NFL 19 (Electronic Arts)	4. Borderlands 3 (Gearbox)	4. The Last of Us: Part II (Sony)
5. Super Smash Bros. Ultimate (Nintendo)**	5. Mortal Kombat 11 (NetherRealm)	5. Ghost of Tsushima (Sony)

*Table 1. Top-5 highest-grossing video games 2018-2020 (retrieved from venturebeat.com). ** indicates video games with cartoonish graphics.*

Players have started considering the quality of graphics to be the most crucial aspect to consider before purchasing a video game, even above the quality of storytelling (Entertainment Software Association, 2017). Such a consideration has led the entertainment video game industry to a "graphics race" towards photorealism and simulation-driven design, for virtual worlds to become increasingly involving (Batt, 2015). Consequently, the industry has also encouraged the acceleration of technological progress related to data acquisition from real-life situations. In Entertainment products, face synthesis is studied in detail not only from the aesthetic point of view but also from an anatomical point of view.

Although the same work pipelines are shared among different approaches (Table 2, following page; Foley, Van Dam, Feiner, & Hughes, 1995; Riemer & Kay, 2019), graphical accuracy can be achieved through several methods, including hand-animation, manual approaches using real footage of people as a reference, physics-based simulations, or computer-aided scanning.

PIPELINE PHASES	PHASE DESCRIPTION
Scanning or modelling	"A face is created either from computer-aided scanning such as photogrammetry, or artist interpretation using computer modelling tools."
Expressions or poses	"This stage defines the range of motion. Often these key poses relate to the theory of Facial Action Coding System, which break down the face's expressions into Action Units (Ekman, 1992)."
Correspondence	"Correspondence is achieved between expressions so that the model may move between key expressions seamlessly."
Rigging	"The rigging stage allows controls for moving the face to be presented for either manual or data manipulation of the face."
Texturing	"The fifth stage of texturing adds realism with skin and hair detail, and the correct responses to light. The face is now complete."
Animation and Rendering	"The last two stages animate the face and render a final output at the appropriate frame rate and resolution with appropriate lighting."

Table 2: The seven stages of face synthesis pipeline, as described in Riemer & Kay, 2019.

An example of the possibilities unlocked by the use of physics-based simulation models in the Entertainment Industry is the accurate reproduction of the modality with which light is reflected from the characters' clothes, skin, and eyes (Dalibard, Magnenat-Talmann, & Thalmann, 2012). For instance, such a technique realistically ensures wrinkling and reflective qualities to human skin and pores, especially when facial muscles contract (Pueringer, 2019a; Fig.4).

Figure 4. Virtual Skin Microstructure Deformation when closing one eye (ICT Vision and Graphics Lab).

However, human characters’ animation is increasingly based on natural human motion through whole-body sensor-based techniques borrowed from cinematographic VFX (Dalibard, Magnenat-Talmann, & Thalmann, 2012; Klehm et al. 2015). Among these, Motion Capture is defined as the process of digitally recording actors' movements to use them as a reference for the animation of Virtual Agents. Movements are captured through spatial markers, rubber balls attached to a full-body suit worn by the actor (Fig.5, following page). Markers can either be passive, coated with a retroreflective material, or active, which emit light autonomously in a unique identification pattern. The actor’s movements are sampled many times per second by a system of 2 to 48 cameras (Fig. 6, following page). Successively, retrieved data is mapped to a rough digital 3D model of the Virtual Agent.

Figure 5. Actor Troy Baker and actress Ashley Johnson respectively playing the roles of Joel and Ellie in "The Last of Us" (Naughty Dog, 2013). Clearly visible on their head and torso, the white dots acting as space markers for the digital rendering of their body movements.

Figure 6. Death Stranding's (Kojima Productions, 2019) camera setup, as revealed during a panel Business Insider attended featuring Kojima and Reedus at the Tribeca Film Festival in May 2019.

When motion capture techniques include facial expressions, it is often referred to as performance capture (Hart, 2012). Virtual Agents in entertainment video games can compellingly convey emotions thanks to professional actors' employment displaying said emotions. In this case, the actors’ facial movements are recorded through markers placed at specific points on an actor's face during facial optical motion capture (see Fig.7). Thanks to new technologies, complex movements and physical interactions of muscles and tendons, such as weight and forces, can be easily recreated in a physically accurate manner. Moreover, the amount of data produced is tremendous compared to traditional animation techniques, therefore constituting a far more efficient technique. Such a technological advancement does not merely have an aesthetic value. On the contrary, accuracy in digital facial expressions often becomes an integral part of the gameplay, as the player is often encouraged to sense which characters are to be trusted (Domsch, 2017).

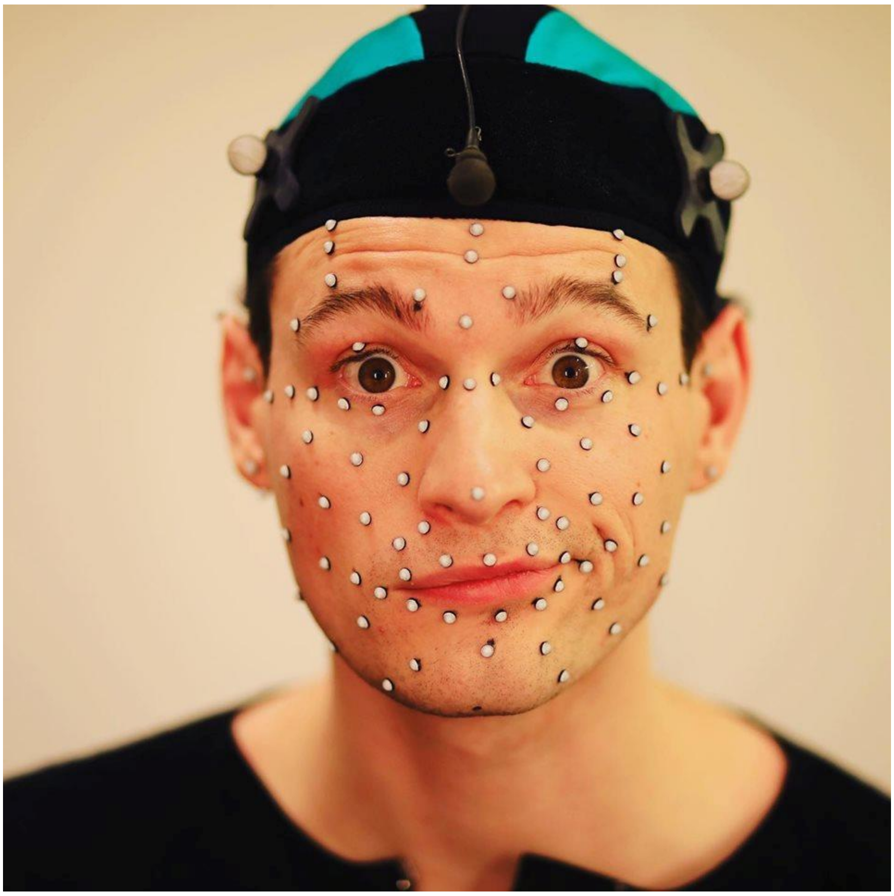


Figure 7. Actor Bryan Dechart playing the role of Connor in “Detroit: Become Human” (Quantic Dream, 2018). Clearly visible on his face, the white dots acting as space markers for the digital rendering of his facial expressions.

2.2 Contemporary Game Habits: Voice Acting in Video Games

Games have also continuously evolved towards new categories, to the point that there is no standard accepted taxonomy of genres. One of the most adopted is Herz’s system (Herz, 1997), defining ten different genres of commercial video games: action games, driving-racing games, puzzle games, strategy games, simulation games, exergames, horror games, commercial brain-training programs, arcade games, adventure games. However, some video games are centered around mechanics across genres, defined as “methods invoked by agents for interacting with the game world” (Sicart, 2008), and some others on storytelling. Story-Driven games belong to this latter category, in which the focus is highly narrative.

In particular, the current video game industry is most interested in developing highly interactive experiences, not only in the traditional sense, in which the player controls the character movements in the virtual environment but in a much more complex way. The player is, in fact, often able to affect the direction of the game’s narration by selecting the course of actions that the character should take. In such genre of video games, the player is therefore required to make decisions on their avatar’s behalf, often in the form of “quick-time events”: a limited amount of time is granted for the decision in order to encourage a fair degree of authenticity to the response (Tavinor, 2017).

Such video games are defined as the “Butterfly Effect” games, a term borrowed by a mathematical theory according to which a minuscule localized change in a complex system can have a large effect elsewhere in the same system (Gawade, 2018). A rather famous example of a Butterfly Effect game is Until Dawn, a survival horror game (Supermassive Games, Sony Computer Entertainment, 2015). In such a game, the player can take decisions which affect the chance of survival of various characters, but also the gameplay: for instance, if the player states that they are afraid of cockroaches, the game will scatter such animals in the virtual environment, increasing the level of uneasiness of the player. During recent years there has been a clear tendency towards “tailoring the story to how the subject plays” (Telltale, 2012), allowing the player to choose where to direct the course of events, from what to eat during breakfast to which life to save. *Until Dawn* (SuperMassive Games), various Telltale series (*The Walking Dead*, *Game of Thrones*), and *Life is Strange* (Square Enix) are all famous examples of successful games based on these mechanics, which allows the player to be the director of their adventure.

Therefore, Dialogues represent an essential aspect of story-driven games, as they provide information that substantially sustains the payers' sense-making and decision-making. Consequently, a further relevant aspect of a Virtual Agent's human-likeness can be found in vocal communication. Earlier video games mainly relied on soundtracks and sound effects while delegating in-game dialogues to text-only sentences. Some retro-style video games, or long-running video game series, still use this set-up to this day (Domsch, 2017; Fig. 8).

Figure 8. Dialogues in text-only form in "Pokémon Red" (above, Nintendo, 1996) and in "Pokémon Sword"" (below, Nintendo, 2019).

Nonetheless, as the entertainment industry increasingly relies on actors' performances to convey visually credible characters, fully human-voiced Virtual Agents become the industry standard in video games, adding layers of multimodal information to a dialogue (Domsch, 2017). In most cases, the same actor working as the motion capture reference also records the Virtual Agent's voice, which in the final product is often accompanied by subtitles in different languages. Only in rare cases, Virtual Agents' voices are dubbed in different languages, especially in countries where cinematography has already paved the way.

In minor productions or in video games that provide a remarkably complex experience by allowing the player to affect the very direction of the game's narration, Virtual Agents can be only partially voiced. In these cases, actors record a limited number of sentences and vocal non-verbal exclamations that are deemed essential for the plot or to understand the Virtual Agent's personality. An example of this approach to voicing can be found in the *Fire Emblem* series (Intelligent Systems, Nintendo).

2.3 The Research Industry: Resources and Challenges

The present dissertation's first objective is to investigate whether the Entertainment Industry is truly posing higher and higher standards for Virtual Agent human-likeness, and, therefore, if habitual video game players have harsher reactions to Simulated Human Interactions for educational purposes if compared with non-habitual video game players. As the last decades have seen the transformation of video games from a niche hobby to mainstream glory, to the point that there is an estimated 2.47 billion video game players globally (Statista, 2019), an increasingly larger population may find Simulated Human Interactions for educational purposes inadequate, causing perceptual disparities that may contribute to unsettling emotional reactions, overruling the benefits of using such a training tool.

Contemporary research has been dedicated to the creation of multiple approaches to animate underlying muscle mechanics and innervation patterns of Virtual Agents' faces, such as polynomial curve fits (Lee, Badler, & Badler, 2002) and data-driven approaches (Deng, Lewis, & Neumann, 2005; Le, Ma, & Deng, 2012). Despite such virtuous examples, advanced technology and statistical models allowing high-fidelity face synthesis are often not available to smaller research laboratories developing Simulated Human Interactions for educational purposes. There are also several challenges connected to using sophisticated techniques such as Motion Capture, especially in research teams working on the development of Simulated Human Interactions for educational purposes. One of the main obstacles to photorealism is processing power, as the process of compiling movements of the digital models can take hours per single frame and require high-end graphics cards. Secondly, the economic cost of using modern techniques might be unsustainable for a relevant number of research teams, as a Motion Capture suit alone can cost up to 80.000 $, and it requires the involvement of specialized workers. As for a Virtual Agent's natural-sounding speech emphasis, voice synthesizing relies on a limited amount of data, i.e., a limited set of available phrases, and recording human sentences is time-consuming and, to a certain extent, expensive if professional actors are hired to do so (Wu, Ning, Zang, Jia, Meng, Meng, & Cai, 2015).

In present times, there is an urgent need for researchers to communicate with the industry, as the aforementioned technological advancements are very likely to soon become widely available on accessible computers (Seymour, Riemer, & Kay, 2017), posing higher and higher standards to younger generations. On the bright side, communication with film and video game industry professionals such as Oscar-Winning Chief Technology Officers and VFX Supervisors is becoming increasingly widespread. Interestingly, experts contacted for the sake of a Delphi Study did not consider the Rendering phase as the critical element for achieving realistic Virtual Agents, but rather the Animation phase (Riemer & Kay, 2019). Therefore, practical knowledge seems to confirm what has already been suggested by previous research regarding the Uncanny Valley and moving Virtual Agents. Reaching a significantly high level of static realism might, therefore, raise the users' expectations, as the more an Agent tries to resemble a human being, the more it raises the expectation of behaving like a human in terms of body language and facial expressions (Lehmann, Roncone, Pattacini, & Metta, 2016; Johnson, Veltri, & Hornik, 2008). Such an assumption is graphically represented by the increased steepness of the Valley in Mori's theorization (Mori, 1970). A further critical theme pointed out by the experts during the Delphi study was that reproducing a known individual, either a famous actor or someone who is personally known by the user, was much harder than an unknown person (Riemer & Kay, 2019). Along the same lines, previous research has suggested that, as humans can subconsciously anticipate the movement's speed and direction in other people (Sakata, Takubo, Inoue, Nonaka, Mae, & Arai, 2004; Akiwa, Suga, Ogata, & Sugano, 2004), the Uncanny Valley might be caused by a violation of perceptual expectations linking appearance to movement (Saygin, Chaminade, Ishiguro, Driver, & Frith, 2012). For what is more, most of the interviewed VFX experts stated that many familiar aspects of human faces are hard to point out and articulate, but when missing one of these, the face feels “wrong” (Riemer & Kay, 2019). This is just another way to say what Freud was writing one hundred years before: “One of the most successful devices for easily creating uncanny effects is to leave the reader in uncertainty (…) and to do it in such a way that his attention is not focused directly upon his uncertainty” (Freud, 1919).

Besides practical know-how, there is also a vast amount of knowledge about Virtual Agents from various sources across a range of disciplines (e.g., anatomy, psychology, neuroscience, and social sciences) that might make up for the technological and economic disadvantages of small research labs. The challenge resides in integrating knowledge from such a massive number of disciplines into comprehensive models capable of synthesizing credible appearance and behavior for Virtual Agents, ensuring their effectiveness in eliciting positive reactions among users of Simulated Human Interactions (Ruhland et al., 2014). The present dissertation explores such different sources in the search for possible “easy wins” in the short development run. In other words, this research has the secondary objective of finding key elements that do not require expensive interventions in terms of money and time but can dramatically increase the perception of the product's quality.

[this page intentionally left blank]

DIGITAL HUMANITY.
Do Users' Gaming Habits Affect the Perceived Human-Likeness of Virtual Agents in a Simulated Human Interaction?

SECTION 2

Theoretical Framework
Negotiation Training in the XXI Century

[this page intentionally left blank]

Chapter 3
Traditional Negotiation Training

3.1 Negotiation and Conflict Resolution Strategies: Tools for Intergenerational Comprehension

The necessity of addressing the blossoming social demands of a world set in the XXI century has driven attention towards effective communication and Life Skills in general, defined as positive and versatile behaviors enabling people to successfully adapt to an ever-changing world (World Health Organization, 1994; Jain & Anjuman, 2013). During recent years, the Western world has witnessed a reorganization of the workforce because of the slow but steady generational turnover between older and younger generations (Istat, 2014). In particular, Millennials, i.e., people born from the late 1980s to the early 2000s (McCrindle, 2009), have now officially entered the job market, bringing their peculiar values, point of views, strengths, and weak points into the workplace.

According to literature, work values, defined as “standards that are held in high esteem by an individual and are related to all aspects of one's personal and work life” (Locke, Sirota, & Wolfson, 1976), considerably vary according to the generation to which the worker belongs (e.g., Gursoy, Chi, & Karadag, 2013; Kaifi, Nafei, Khanfar, & Kaifi, 2012; Twenge, 2010). Above all, social values on the job seem to be considered as rather crucial by Millennials, as opposed to previous generations such as Baby Boomers, born in the years following the Second World War, marked by an increase in the birth rate, and GenXers or members of Generation X, born from the early 1960s to the early 1980s (McCrindle, 2009). For instance, Millennials show a high expectation of recognition, longing for their employers' approval and rewards (Martin, 2005; Jennings, 2000). Millennials expect to be led with clear directions and provided with immediate feedback, indicating a need for support and more in-depth care about creative expression than about playing leadership roles (Martin, 2005). Differently from representatives of Generation X, who mainly preferred individualism over collectivism (Jurkiewicz, 2000), Millennials work best in teams, showing advanced collaboration skills (Martin, 2005; Wilson & Gerber, 2008), and valuing relationship boundaries on the job (Wrzesniewski, LoBuglio, Dutton, & Berg, 2013).

GenXers and Baby Boomers may see the vital need for social interaction and immediate feedback as weaknesses of the millennial generation (Gilbert, 2011), and the lack of previous work experiences could lead older coworkers to abstain from offering recognition (Fenzel, 2013), which instead results to be a rather significant value to the younger generation (Gorman, Nelson, & Glass, 2004). Therefore, the millennial generation has been described as entitled and challenging to work with (Deal Altman, & Rogelberg., 2010). Similarly, as Millennials expect to be able to openly communicate with supervisors without being intimidated because of their inferior status (Chou, 2012), to be taken seriously (Espinoza, 2012), they also expect to be treated as individuals and to be genuinely invested on by managers (Murphy, 2012; Raines, 2012). Moreover, while some studies state that the expectations about the workplace of previous generations tended to shift after transitioning from education to work (Jin & Rounds, 2012; Krahn & Galambos, 2014), no significant differences in displayed work values between pre-career and working Millennials have been found (Kuron et al., 2015). This implies that Millennials are, once more, different from previous generations because their work values tend to maintain their salience over time and fail to adjust to the pace and structures of the traditional marketplace (Kuron et al., 2015).

The characteristics mentioned above that members of the millennial generation share are inevitably and increasingly permeating the workplace and affecting on-the-job behavior (Chou, 2012; Elam, Stratton, & Gibson, 2007; Peebles, 2008). Literature suggests that a failure to consider the generation's values that will increasingly represent the workforce may result in conflicts, misunderstandings (Krug, 1998), and miscommunication (Jurkiewicz, 2000). This is especially true as, in the contemporary marketplace, interpersonal relationships are considered as vital sources of information to build a stable network and to fully integrate newcomers into the team and the company's mission (Buha, 2014; Casalnuovo, Vasilescu, Devanbu, & Filkov, 2015), as well as to learn social norms and expectations in the workgroup (Myers & Sadaghiani, 2010).

Among various communication skills, effective Negotiation strategies sustain effective team-work and cross-generational communication skills, as they can be defined as the ability to discern, respect, and grasp the perspectives and the alternative norms provided by different ethnicities, statuses, and cultural backgrounds (Dede, 2010). Negotiation skills, therefore, represent a crucial requirement in the XXI century workplace (Bedwell & Fiore, 2010; Pellegrino & Hilton, 2012). In this latter context, in fact, besides being relevant for successful communication between workers and customers (Mast & West, 2016), negotiation also seems to be crucial in the relationship among employees, from a perspective of sharing meanings (i.e., group interpretation) and co-construction of problem-solving strategies (Dede, 2010). On the contrary, without a proper Negotiation strategy, staff's productivity and morale are likely to decrease, resulting in higher costs, not only for the companies themselves (e.g., productivity losses, high turnover, and absenteeism rates), but also for the individuals (e.g., psychological stress, lower quality of life; Huyler, Pierre, Ding, & Norelus, 2015; Fernet, Austin, & Vallerand, 2012). Therefore, both being profitable in the long term and providing the opportunity for employees to thrive intertwine into a single objective for modern companies, rather than being antithetical aims (Deci, Olafsen, & Ryan, 2017).

In conclusion, a successful and effective Negotiation strategy not only sustains effective communication strategies, but it also results to be relevant for the individual's general well-being (Vincent, 2005) and the construction of stable and transparent interpersonal relationships (Buha, 2014; Casalnuovo, Vasilescu, Devanbu, & Filkov, 2015). For such reasons, research dedicated to designing and implementing training dedicated to improving one's Negotiation and Conflict Resolution Strategies is blossoming in contemporary educational settings. Negotiation training has been selected as the primary material of the present dissertation.

3.2 Negotiation Theories

3.2.1 An Integrative Negotiation

Negotiation abilities represent an essential component of Life Skills. Although there is not a single agreed-upon definition of the term "Negotiation", it can be defined as a communication process through which two or more parties who share the same goal but hold different points of view come together to define a solution able to converge their different interests (Raiffa, 1982; Rumiati & Pietroni, 2001; Monoriti & Gabellini, 2018). Thus, Negotiation can allow the resolution of such conflict through strategic interaction modalities.

Such a communication process is, therefore, strictly connected to the concept of conflict. It is also worth to notice that having different interests would not lead the parties to a conflict if the available resources could be equally distributed to all parties. Consequently, conflict arises in the context of scarcity of resources, and it is encouraged by the party's determination to obtain what is desired (Monoriti & Gabellini, 2018). When resources are scarce, the conflict begins when one party perceives another as harming them, causing concern (Thomas, 1992).

As conflict involves parties who hold different points of view, they can emerge between members of different cultures or social identities (Wilmot & Hocker, 2001). A group's social identity is connected to the individual's awareness that they belong to specific values, norms, and beliefs. Social categorization carried out by individuals allows them to identify their group's relative social position compared to others that hold different positions and needs. Such a social comparison process implies a division between "us" and "them", not only involving the individual at a cognitive level but also on an emotional and motivational level (Tajfel, 1981).

Nonetheless, conflicts can also arise among members of the same culture who hold different values when they perceive a violation of their norms and roles. Within a group, communication allows members to share events to create a common perspective on reality. Such shared opinions are, although, partially due to a so-called social influence: individuals tend to adapt to social pressure, especially when they are numerically disadvantaged compared to those who exercise it, when they occupy a lower social position in the group, or when and they have a low level of self-esteem (Asch, 1956).

Traditionally, Negotiation has been regarded as a contentious matter revolving around the concept of conflict. According to such a conception, what is gained by a party is equal to what is lost by the other, involving, therefore, a winner and a loser ("win-lose" logic).

Such a “Distributive” conception of negotiation derives from the Game Theory used for the study and analysis of individuals' decision-making choices on how to distribute a scarce resource (von Neumann & Morgenstern, 1944).

A more recent conception of Negotiation allows seeing the involved parties as allies rather than competitors, shifting from a logic of "win-lose" (distributive negotiation) to a logic of shared interests, i.e., a “win-win” logic (integrative negotiation; Carnevale and Pruitt, 1992). Among the authors sustaining a “win-win” Negotiation logic, Fisher, Ury, & Patton identify four principles that sustain an effective Negotiation strategy (Fisher, Ury & Patton, 1991):

- **The first principle**: It implies the exigence to separate people from matters. Since most negotiations are immersed in complex relationships, the negotiation strategy should be built to safeguard existing networks. This ability underlies the need to separate the emotional sphere from the rational sphere;

- **The second principle**: It relies on the ability to distinguish differing from common interests. To do so, it is necessary to identify the actual interests of the parties, often hidden by psychological and cultural reasons;

- **The third principle**: It concerns the use of creativity and imagination to bring out as many solutions as possible and to be able to identify the best one for both parties;

- **The fourth principle:** It concerns the ability of the negotiators to rely on objective criteria, as opposed to personal interests, to avoid negotiation failure due to excessive persistence on individual positions (Ury, 2007).

3.2.2 Negotiation and Conflict Resolution Styles

Besides the issue of conflict, Negotiation also involves how individuals approach it and act. Individuals tend to choose a model of principles that guides them throughout the entire negotiation process, and such an orientation is manifested in the form of a system of observable behaviors that tends to stay stable over time. The style of conflict management is defined as a general and stable orientation that an individual has towards another and the conflict's object (Kuhn & Poole, 2000).

Among the numerous theories related to Negotiation that have been presented during the years, Rahim's Model states that Negotiation styles depend on the amount of concern the negotiator has for himself or herself and the negotiation partner, and therefore such a definition orbits around the distribution of power in a relationship (Rahim, 1983). Five negotiation styles arise from combining the two dimensions, i.e., concern for oneself and concern for others (Fig. 9).

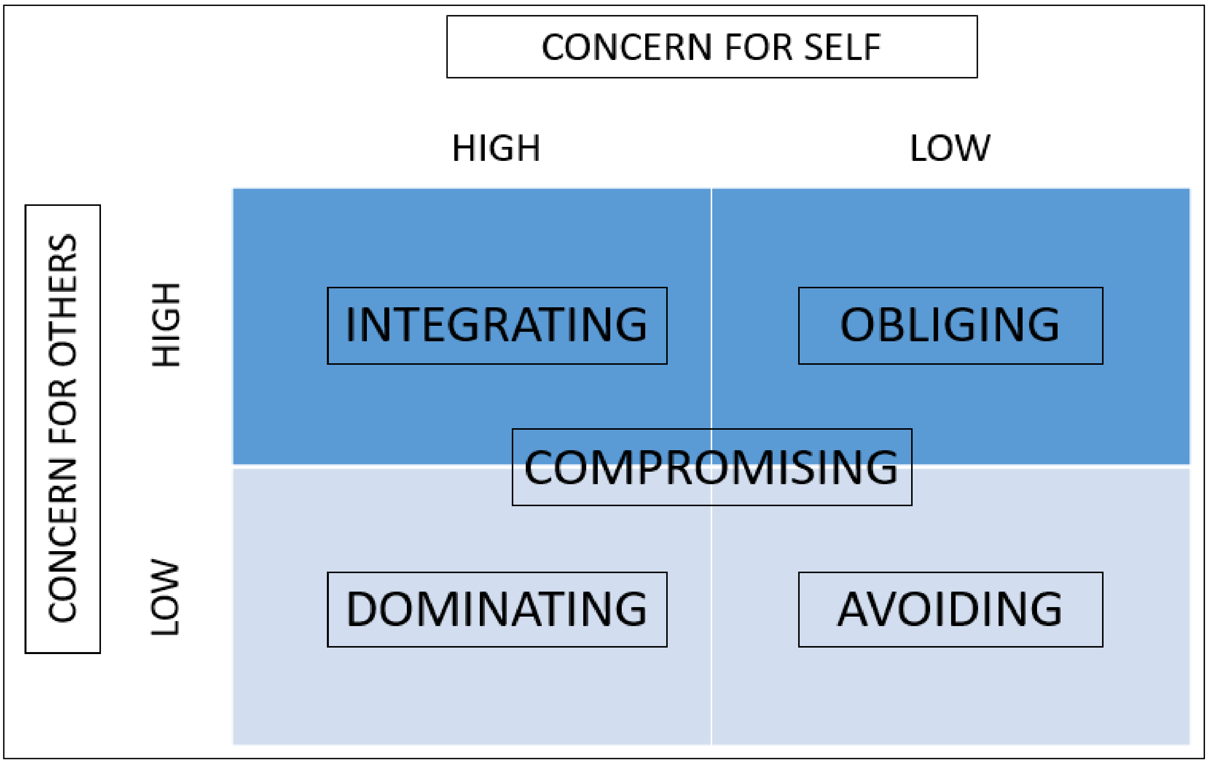


Figure 9. The five styles of conflict management as theorized by Rahim, based on the two dimensions of self-orientation and orientation towards others (Rahim & Bonoma, 1979).

In particular, the five styles are described as follows:

- **The Integrative Style:** It is characterized by a high interest situation for oneself and the other party involved in the conflict. It is a style marked by the collaboration between the two parties (open exchange of information and examination of differences) to reach a durable and optimal solution acceptable for both (Fisher & Ury, 1991). The integrative approach leads the negotiator to understand the interests of the other parties better, thus increasing the chances of reaching an acceptable solution and ensuring a higher value level for all parties (Currall, Friedman, Tidd & Tsai, 2000). Moreover, those who use an integrative style tend to adapt their approach to the approach adopted by the other party, and in this way, the collaboration produces more collaboration (Kelley and Staheiski, 1970). In conclusion, an integrative style is likely to produce an environment with a low level of conflict (Currall, Friedman, Tidd & Tsai, 2000), and it is associated with effective problem-solving, as it can lead to creative solutions (Rahim, Garrett & Buntzman, 1992);

- **The Obliging Style:** It is characterized by a low interest for oneself and high interest for the other party involved in the conflict. It is a style marked by the tendency to minimize the differences and emphasize the points in common to satisfy the other party's interest. The Obliging Style is also associated with an individual's tendency to consider themselves wrong or consider the other party's interests more important than their own (Rahim, Garrett & Buntzman, 1992). Therefore, it is likely to lead to the conflict's solution with the surrender of one part in favor of the other. While this style does not allow a satisfactory solution to be achieved for all parties, it is likely to have a positive impact on others (Yukl & Tracey, 1992; Ferris, Judge, Rowland, & Fitzgibbons, 1994; Wayne, Liden, Graf, & Ferris, 1997) thanks to the reduction of the conflict at the level of relationship between the parties;

- **The Dominating Style:** It is characterized by a high interest rate for oneself and low interest for the other party involved in the conflict. A negotiator using a Dominating style uses a strong competitive tendency to pursue their interest at the expense of the other party (Rahim, Garrett & Buntzman, 1992). The probabilities of reaching a solution are minimal, and it always fails to reach a solution accepted by all parties (Pruitt et al., 1983; Ben-Yoav & Pruitt, 1984a, 1984b). This style tends to be preferred by those who implement rigid behaviors (Pruit and Carnevale, 1993). Some research points out that, in a business environment, this style has significant impacts on behavior and relationships. For instance, if someone has a Dominating style, his collaborators will hardly have the tendency to communicate with them (Richmond, Wagner, & McCroskey, 1983) or to conform to their indications (Rahim & Buntzman, 1990);

- **The Avoiding Style:** It is characterized by a situation of low interest for both oneself and the other party involved in the conflict. It is a style associated with a withdrawal from the situation. Consequently, the individual does not satisfy their interest or the other party (Rahim, Garrett & Buntzman, 1992). It is also characterized by the lack of information exchange between the parties: no party is interested in their own nor the others' demands, and therefore there is no interest in resolving the conflict. Despite the desire to avoid it, the "non-solution" is likely to lead to an ever-increasing conflict (Currall, Friedman, Tidd, & Tsai, 2000);

- **The Compromising Style**: It is characterized by a situation of moderate interest for oneself and the other party involved in the conflict. It is a style associated with sharing, in which both parties give something up to reach a mutually acceptable decision (Rahim, Garrett, & Buntzman, 1992). The literature presents numerous debates on whether to consider it a real style (Van De Vliert, & Hordijk, 1989), as it does not involve active participation in solution-finding, contrary to the Integrative style (Pruitt, 1983).

Some researchers suggest that successful management of conflicts requires the application of specific styles. In particular, it would seem that the Integrative style is the most suitable in most circumstances in terms of transparency and collaboration and the most efficient. As for the other styles, individuals seem to perceive the Dominating style as inappropriate, the Avoiding style as inefficient, and Compromising and Obliging styles as relatively neutral (Gross & Guerrero, 2000). On the contrary, other scholars argue that the most appropriate style to use depends on the current situation, and there is not a single correct answer to all Negotiation problems (Rahim & Bonoma, 1979; Thomas, 1992).

3.3 The Role of Emotional Intelligence in Negotiation

A high level of emotional intelligence, defined as the ability to evaluate, express, and regulate one's emotions, as well as the ability to use them in decision making effectively (Rahim, Psenicka, Polychroniou, & Zhao, 2002), has been linked to the effective use of conflict management strategies. As a consequence, the Emotional Quotient, which includes abilities such as self-awareness, self-regulation, motivation, empathy, and social skills (Goleman, 2001), and the five conflict .management styles indicated by Rahim's theorization seem to have a reciprocal influence, collectively influencing the problem-solving strategy privileged by the subject when confronted with a negotiation (Rahim, Psenicka, Polychroniou, & Zhao, 2002).

Besides focusing internally, an optimal negotiator should also effectively understand the interlocutor's internal states. The Adaptability, Interpersonal Skills, and Stress Management subscales of the Bar-On Emotional Quotient Inventory (Bar-On, 2004) seem to have a significant relationship with conflict management skills, in particular related to the individual's ability to adapt their style to the interlocutor (Hopkins & Yonker, 2013), with positive consequences on their emotional well-being (Benitez, Medina, & Munduate, 2018; Van Niekerk, De Klerk, & Pires-Putter, 2017).

The transition from a self-centered approach to considering all the involved parties' needs is not simple, as individuals tend to take a defensive position toward their values and needs automatically. However, such behavior must be overcome to find a shared winning solution, and empathy seems to positively sustain this change of direction (Monoriti & Gabellini, 2018). Therefore, a key role is played by the individual's social perception, defined as "the ability to understand and react appropriately to the social signals sent out by other people vocally, facially, or through body posture” (Mayer et al., 2016). As facial expressions connected to emotions convey rich social information, studies showed that participants who were more capable of recognizing interlocutors’ emotional facial expressions reached better gains for themselves and their counterparts after the negotiation, while also being perceived as more cooperative and likable (Schlegel, Mehu, van Peer, & Scherer, 2018; Elfenbein, Der Foo, White, Tan, & Aik, 2007).

It is in this context that Virtual Agents capable of realistically portraying emotional facial expressions show their importance. As facial expressions are a rich source of information during a negotiation, plausible Virtual Agents contribute to higher learning levels among users (Cano & Sáenz, 1999; Datar, Garvin, & Cullen, 2010). On the contrary, flawed Virtual Agents may negatively affect the user’s perceived immersion and disrupt a Simulated Human Interaction's simulative nature by impeding generalizability to real-life situations (Von Bergen, 2010; MacDorman & Ishiguro, 2006).

3.4 Traditional Training for Negotiation Skills

Negotiation training has been delivered in a variety of contexts and a variety of declinations. Therefore, the following pages will be dedicated to a brief account of the most-used techniques and methods in which Negotiation training has been delivered during the last decades.

- **Frontal Lectures:** Despite a general tendency of schools and universities not to provide students with specific skills regarding negotiation (Tynjala, Slotte, Nieminen, Lonka, & Olkinuora, 2006; Wyn, 2009)(Black & William, 1998), there have been several exceptions to this trend since the 1980s. For instance, a one-semester long university course delivered in Florida and intended to improve business students’ self-confidence as negotiators positively affected the development of negotiation skills and self-confidence as measured by self-reported assessments (Taylor, 2008). Moreover, it positively impacted negotiation style changes as measured by the Conflict Management Style Scale immediately after the course and eighteen months after the course (Taylor, 2008). Besides, a much shorter training, i.e., a twenty-hour long course about collaborative negotiation delivered throughout three weekends at the Columbia University, resulted in affecting the work environment positively, changing how people think, feel, and act towards their colleagues, according to the Negotiation Evaluation Survey (Coleman & Lim, 2001). Frontal lectures are described as the most commonly used method to teach with, even though precautions should be taken for the trainees to be able to actively participate in the lecture, to avoid attention decrease during the course (Castagna, 2014). Compared to an online course, frontal lectures resulted in more effective in providing students with practical knowledge about negotiation, resulting in higher performance in the skill for identical retained theoretical information (Callister, 2016).

- **Role-Playing Methods:** One of the most used interactive methods of training negotiation skills is role-playing, during which participants play the role of different characters to simulate various scenarios. By providing an active learning experience, role-playing methods allow trainees to experience closely simulated feelings and situations that they might encounter in the future (Kilgour, 2015). For instance, in the National Highway Safety Administration task (Hirokawa, 1983), the objective is to propose a course of action for effectively controlling speeders on Seattle's arterials, while in the "Lost in the Arctic" or the "Desert Island" tasks (Hiltz, Johnson, & Turoff, 1986), the group is asked to find a shared solution to escape from an uncomfortable situation. An alternative to team-building activities followed by an expert-conducted debriefing is experts' involvement as role-players, simulating different interlocutors' personalities and roles, and evaluating the trainee's performance firsthand (Kron et al., 2017). Engaging in face-to-face negotiation tasks seems to be particularly useful if proposed following a one-hour-long lecture regarding bargaining and negotiation, resulting in higher agreement rates and profits for both negotiators, independently from the quality of the assigned goal (i.e., easy or difficult; Northcraft, 1994). More recently, a role-playing exercise regarding salary negotiation was included in a university course for business students (Wesner, 2018). Results showed that such a method encouraged students' interest in the topic and their engagement in the curse while allowing easy applicability to their professional lives (Wesner, 2018).

- **Case Analysis:** The Case Analysis method is based on analyzing specific scenarios, either real or hypothesized, to identify the best general behavior from examining a specific case. Such a method enables trainees to deal with complex topics, as it links them to concrete actions and occurrences, encouraging brainstorming techniques to grasp the cause-effect relationships in diverse situations (Mander, 2014). This is provided trainees are able to draw the appropriate conclusions from the specific case and generalize rules and advice to the appropriate future cases, without overlooking important details that might set the two cases apart (George, 2018). Students who were asked to read several negotiation cases and compare them were three times more likely to use what they learned from the comparison in a face-to-face negotiation simulation if compared to students who were asked to read each case and provide advice to the characters (Thompson, 2000). Similarly, students asked to compare written negotiation cases were more likely to use what they learned in a face-to-face negotiation simulation than students who were simply asked to read each case (Loewenstein, 2003). Such results seem to suggest that case reading, when actively involving the student, is a suitable training option (Loewenstein, 2003).

- **Coaching:** Coaching can be defined as the process in which an expert provides a coachee with training and guidance to reach a specific goal. According to this method, career guidance and social relationships result in equal importance (Chambers, 2011), juxtaposing an emotional approach to the coachee's professional advice (Quaglino, 2004). Therefore, coaches must communicate what to do and how to do it, according to a relationship that is more related to teamwork than to command (Graen & Schiemann, 2013). In such a sense, motivation is regarded as a critical factor of coaching: inspirational motivation provided by coaches continuously reinforce coachees by acknowledging their regular contributions, encourage values, empowerment, and inspiring to facilitate the success of the coachee (Debaro et al., 2015) and to provide a co-construction of knowledge (Hansman, 2016). By collaborating, the coach and coachee can exchange knowledge in an interactive way (Hunzicker, 2011). As Millennials expect to be supported by supervisors (Murphy, 2012; Raines, 2012) and as the role of coach seems to be a fulfilling activity for Babyboomers, provided that the older employee does not perceive the Millennial as a potential job competitor (Hershatter & Epstein, 2010), coaching seems to be a successful and cost-effective training method.

With the advantage of involving expert people's know-how, all these traditional techniques can nonetheless present different disadvantages. First of all, experts are not always readily available, and the monetary cost of paying them should not be underestimated (Lewin & Gollan, 2018). Moreover, companies need to make sure expert evaluations are consistent with each other, without being negatively affected by the expert's fatigue due to the same exercise's repetition on different occasions. Finally, traditional methods are time-consuming for both experts and trainees (Kron et al., 2017).

It is in this context that Simulated Human Interactions come into play. Using Simulated Human Interactions as training improves the users' competencies more efficiently and flexibly than traditional teaching methods (e.g., Azadegan, Riedel, & Hauge, 2012). Several examples of Negotiation Training based on Simulated Human Interactions will be the subject of the next Chapter.

[this page intentionally left blank]

Chapter 4

Training Negotiation through a Simulated Human Interaction - ENACT

According to some authors, technological innovations represent an effective need-supporting method to respond to many of the disadvantages mentioned above. Since competence and a high level of emotional intelligence have been linked to effective conflict management strategies, multimodal simulations of Human Interactions have been used as training tools for Negotiation strategies. Virtual Agents displaying verbal and non-verbal communication components allow the learning of complex scripts through interactive trial-and-error methodologies while avoiding the risk of jeopardizing real-world relationships through clumsy management of conflicts, on the job and in everyday life (Laborde Torres, 2016). The following pages will describe examples from two different kinds of Simulated Human Interactions for Negotiation: Wizarding-based Simulated Human Interactions and Semi-Automated Simulated Human Interactions.

4.1 Wizarding-Based Simulated Human Interactions

Several Agent-based Simulated Human Interactions for negotiation are based on the wizarding method. Such a method requires the presence of a real person, often hidden from the user, who is in charge of choosing which actions the Virtual Agent should perform during an interaction, among a series of possible options (e.g., Gratch, DeVault, Lucas, & Marsella, 2015). Among the Wizarding-based Simulated Human Interactions, the Conflict Resolution Agent (Gratch et al., 2015) allows its users to practice different negotiation scenarios with agents applying different strategies during the negotiation. Both the participants and the agent can communicate with each other through spoken language or by manipulating, looking at, or pointing at the physical objects involved in the negotiation (e.g., an umbrella or a lamp): the agent can speak to the user through the use of text-to-speech software. This is done with the intent to use multiple communication channels to create complex and multimodal scripts involving perception, cognition, and motor aspects at the same time (Raybourn, 2007).

The Conflict Resolution Agent’s upper body is visible and presented in a graphically realistic way (Fig.10). Its behavior is semi-automated, as low-level functions are automatically implemented (e.g., gestures and facial expressions in connection with specific utterances), while researchers manually select high-level behaviors through two separate sections of the interface, related respectively to verbal and non-verbal behaviors. For the sake of rapidity of use, the interface also includes shortcuts to pause fillers (e.g., “hmm”) and gaze behaviors while the researcher searches for the appropriate high-level response.

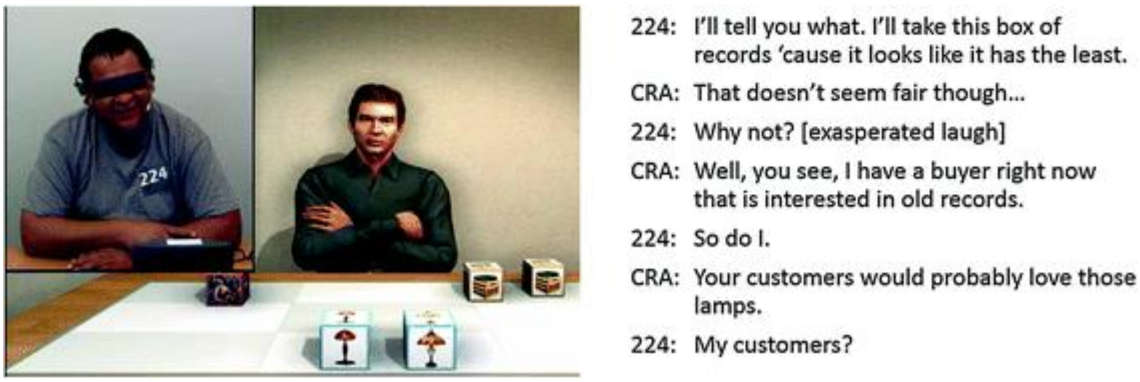


Figure 10. The Conflict Resolution Agent developed by Gratch et al., 2015 (on the right) and a trainee (on the left), and a brief recording of an interaction (image retrieved from Gratch et al., 2015).

Interestingly, results showed that participants who believed to be negotiating with an autonomous virtual agent expended more effort on the proposed multi-issue classic bargaining problems reported a higher level of comfort and punished their negotiation partner less often as opposed to those who were aware of the wizarding method being used (Gratch et al., 2015).

The main advantage of using a Wizarding-based agent lies in the possibility of intelligently and rapidly choosing the behavior to respond to the user among an extremely high number of variants and combinations (e.g., over 10,000 distinct utterances, Gratch et al., 2015). This allows researchers to experiment with different personalities and negotiation policies much more efficiently and in a much more flexible way, without the need to separately code the individual decision-making processes. Among the challenges of choosing a Wizarding-based agent is, instead, the design of an efficient and effective interface for researchers to use. An efficient system has to be implemented for the wizarding researcher to be able to navigate through such a large number of possible behaviors effectively and quickly, approximating the pace of an in-real-life human interaction (e.g., DeVault, Mell, & Gratch, 2015, Fig.11, following page).

Figure 11. Screenshot of a wizarding verbal interface. Items at the upper-left correspond to short, high-frequency utterances. The right half of the interface organizes longer and less frequent utterances by speech acts. Filters at the top allow navigation by speech acts' specific topics (image retrieved from Gratch et al., 2015).

Moreover, along the same lines of traditional methods, Wizarding methods still need an expert's presence, showing benefits and disadvantages similar to the aforementioned traditional methods, to a certain degree.

4.2 Semi-Automated Simulated Human Interactions: ENACT

Several Agent-based Simulated Human Interactions for negotiation are based on computer-generated choices built in the Agent's programming code. Among these, ENACT (Marocco, Pacella, Dell'Aquila, & Di Ferdinando, 2015) is an online Simulated Human Interaction developed in the context of the European project ENACT, "Enhancing Negotiation skills through on-line Assessment of Competencies and interactive mobile Training" (Lifelong Learning Program, Key Activity 3, 2014-15) and funded by the European Agency AECEA (European Executive Agency for Education Audiovisual And Culture). The project is developed by the NAC Laboratory at the Naples "Federico II" University, coordinated by the University of Plymouth, and partnered by Aidvanced Srl, Fondazione Mondo Digitale, Turkish Ministry of Sports, and Fundetec Foundation Madrid.

A systematic investigation of the Life Skills training needs of target groups across Spain, Turkey, and Italy was used to design and develop ENACT. Results from such investigation confirmed the gap mentioned above between recommended educational practice and current practice in schools and training institutions and a gap between guidelines of technology application and the training provided to experts in the use of technology for Life Skills development (Marocco et al., 2015). Results also show that critical elements related to the concept of Negotiation vary across the participants' nationality and occupation. Regarding nationality, Negotiation was considered to be mostly related to effective communication, empathy, and decision making in Spanish subjects, to effective communication, problem-solving, empathy, and creative thinking in Turkish subjects, and finally to critical thinking and active listening in Italian subjects. As for occupation, Negotiation was considered to be mostly related to the ability to successfully manage conflicts according to teachers, skills in managing interpersonal relationships according to students, and problem-solving and effective time management in the business area.

After collecting the data mentioned above, a Simulated Human Interaction was developed on the Unity Engine, a cross-platform engine developed by Unity Technologies to create both three-dimensional and two-dimensional simulations spanning outside video gaming, to be used in automotive, architecture, engineering, and construction simulations.

ENACT aims to improve individuals' communication and negotiation skills through the interaction between a user, in the form of an avatar, and a Virtual Agent. ENACT’s theoretical framework is based on Rahim’s theorization of Negotiation Styles (Rahim, 1983): both the Virtual Agent’s and the user’s behaviors are coded according to the five main Negotiation Styles, and users receive feedback based on their performance, with the added value of allowing users to practice interactions with Agents applying different conflict resolution strategies. By combining Virtual Agents with a solid theoretical psychological framework, ENACT represents a personalized, learner-centered, and innovative tool that can train the users’ skills in conflict resolution and decision-making, offering an alternative to the often expensive and time-consuming training provided by professional consultancy firms.

The choice has fallen on ENACT as the primary material for this dissertation because of previous studies' preliminary results. Data retrieved from 72 participants aged 6-10 and 79 participants aged 11-60 (mean age = 20.6) related to a first version of the Simulation’s content and interface showed overall positive feedback (95.3% of the interviewed participants stated they would want to train with the Simulation again; Marocco et al., 2015). The only negative exceptions could be found in questions regarding the realism of the conversations and the graphics aspects, which scored, respectively, a mean of 3.4 and 3.3 on a scale of 5 points. Another set of data about a second prototype involving four of the five scenarios mentioned above was collected from 39 participants aged 11-28. During this second study, the Simulation’s content and interface were evaluated positively, as 94.7% of the interviewed participants stated they would want to train with the Simulation again (Marocco et al., 2015). However, the question related to how modern the graphics looked still scored less than 3.5.

Considering the young mean age of the participants, the fact that 56% of people who habitually play videogames are under the age of 35 (Entertainment Software Association, 2017) and that players consider the quality of graphics to be the most vital aspect to take into consideration before purchasing a video game, even above the quality of storytelling, results from preliminary studies are particularly interesting to be further explored. In particular, it can be essential to investigate whether the Entertainment Video Game Industry is truly posing higher and higher human-likeness standards through their Virtual Agents, and, therefore, if such results are partially due to the participants’ game habits. As the last decades have seen the transformation of video games from a niche hobby to mainstream glory, to the point that there is an estimated 2.47 billion video game players globally (Statista, 2019), an increasingly larger population may find Simulated Human Interactions for educational purposes inadequate, causing perceptual disparities that may contribute to unsettling emotional reactions, overruling the benefits of using such a training tool.

4.2.1 ENACT's Virtual Agents

In this Simulated Human Interaction, the user is presented with one-on-one scenarios described in written form. Users are then required to interact with the Virtual Agent through a multiple-choice approach. The Simulation uses a first-person view, i.e., a graphic perspective in which the user observes the virtual world from their avatar's point of view. The Virtual Agent and the avatar's upper bodies are visible and presented in a graphically realistic way. ENACT allows the researcher to choose which dyad among nine female Agents and seven male Agents to include in the Simulated Human Interaction, in the form of the user's avatar and their interacting Agent (Fig.12).

Figure 12. The nine female Agents among which both the user's avatar and the interacting Agent are chosen in ENACT (Marocco et al., 2015)

Each Agent is identified by a name, as previous literature has suggested that assigning Agents a typically human name may contribute to the degree of their perceived human-likeness. Moreover, users interacting with an Agent with a human name are more likely to assign higher trustworthiness and credibility to the Agent and have a more positive attitude towards the product (Zhao, 2006; Go & Sundar, 2019). Besides, participants were less concerned about the Virtual Agent influencing their autonomy when identified by a human name (Voorveld & Araujo, 2020).

Besides gender, Agents and avatars also differ for their clothing style (i.e., formal/informal attire), their height, and represent different races (e.g., Caucasian, Arabic, Asian, Latin, Black). The Agent and the avatar communicate through a combination of four different facial expressions (related to, respectively, a neutral, happy, sad, or angry emotion, Fig. 13), 24 automatically implemented gestures and mouth articulations, and ten different body postures and relative gaze directions connected with the five Negotiation Styles theorized by Rahim (Rahim, 1983). Based on face-to-face observational data, each Negotiation Style is portrayed through two alternative body posture configurations (e.g., Fig.14, following page).

Figure 13. Examples of ENACT's Agents and Avatars' facial expressions, from left to right: angry, happy, sad (Marocco et al., 2015).

Figure 14. Examples of ENACT's Agents and avatars body postures, in this case, related to the Dominating Negotiation Style (Rahim, 1983; Marocco et al., 2015).

ENACT's Agents and avatars also communicate vocally through the use of gibberish. Gibberish is a speech modality defined as "vocalizations of meaningless strings of speech sounds, and thus have no semantic meaning" (Yilmazdyildiz, 2015). By conveying non-verbal elements such as the voice's tempo, prosody, and intonation while eliminating the semantic content, gibberish has been used in various cross-cultural and cross-language experimental settings products, including ENACT (Marocco, 2015). Instead, the semantic content is conveyed through written text for both the Agent (in the form of a comics-inspired balloon above their head) and the user's avatar (in the form of text boxes). Written text is presently available in English and Italian. The voice volume is conveyed through different balloon lines, e.g., a jagged balloon implies shouting or yelling.

4.2.2 ENACT's Scenarios

ENACT provides five different scenarios in which the user can practice, and each scenario involves five different variants related to the conditions under which the negotiation process takes place, both regarding the backstory and the involved characters. Versions 1 of the five scenarios are as follows:

- **TV Program:** The user is married to the Virtual Agent, and they live in a house with just one TV. The objective of the negotiation is to decide which TV program to watch during the evening. The user's avatar loves watching Report, a news program, while the spouse loves Masterchef, a cooking show. The spouse is also in a bad mood because of an ominous phone call with their manager.

- **Sport:** The Plymouth football team has decided to celebrate their promotion to the Major League by asking for a redesign of their team logo. The negotiation involves the user and a colleague, both designers of the Company selected by the Plymouth team manager. The objective of the negotiation is to decide who will be the leader of the design team. The user plays a Plymouth team member who is also very skilled at using a design software, while the colleague is a design expert who has been feeling hopeless and sad lately.

- **Motorbike:** The negotiation involves the user and their sibling, who both need to use the family's motorbike. This is unexpected, as they usually have very different habits regarding their "going-out times". The objective of the negotiation is to find a solution. The user plays the role of a sibling who cannot miss a meet-up with their friends, while the Agent is their sibling who needs the motorbike to get to the city center and meet with their friends.

- **Restaurant:** The negotiation involves the user and their friend, both waiting for a mutual friend's suggestion regarding an after-dinner meeting. In the meantime, the objective of the negotiation is to decide where to have dinner. The user plays the role of a person who has a craving for pizza, while the Agent is their friend who loves Chinese food.

- **CD:** The negotiation involves the user and their roommate, who have been waiting in line for 30 minutes to finally purchase the autographed version of their favorite band's new album. Even though they both had a reserved autographed copy, the shop has only one album left on the shelf. The objective of the negotiation is to decide who is going to keep the autographed album. The user plays the role of a person who loves limited editions and band memorabilia and has been waiting for this moment for a month, while the Agent is their roommate who has the habit of listening to a new album on repeat for a week.

Every scenario is introduced by written instructions, explaining the context of the scenario and the matter that needs to be negotiated. Information about the two interlocutors, their relationship, mood, and motivations is also given (e.g., Fig.15).

Figure 15. An example of ENACT's introductory screen, in which the context of the scenario, the matter that needs to be negotiated, the two interlocutors, their relationship, mood, and motivations are explained. In this case, the descripted scenario is "TV Program" (Marocco et al., 2015).

4.2.3 ENACT's Interaction Modality

ENACT's interaction modality is based on a multiple-choice approach. During the Simulated Human Interaction, the Virtual Agent is portrayed on the right side of the screen, and their utterances are displayed in written form in a balloon above their head. Instead, the user's avatar is portrayed on the left side of the screen, in the upper corner.

The avatar is shown as if the user was looking at a mirror and, therefore, the user can watch their avatar's movements and appearance throughout the Simulated Interaction. Below the user's avatar, there are five options through which the user can continue the interaction (Fig.16). Each option is linked to a specific Negotiation Style, as theorized by Rahim (Rahim, 1983). As the user hovers on each of them with the mouse, the avatar offers a preview of the facial expressions, gestures, body posture, and gibberish associated with the utterance.

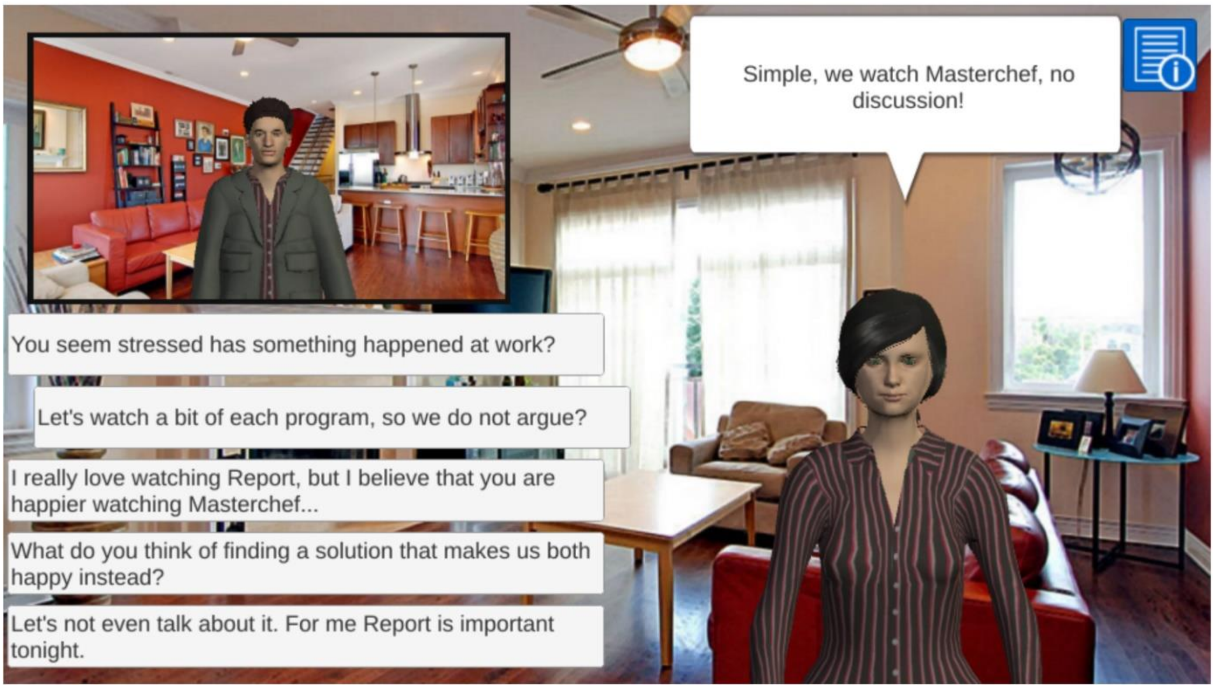


Figure 16. An example of ENACT's multiple-choice interaction modality, in the Italian translation. The avatar is portrayed on the left side of the screen, while the Virtual Agent is portrayed on the right side. The multiple-choice options are located below the avatar, while the Agent's utterances are shown in a balloon above the Agent's head. In this case, the ongoing scenario is "TV Program" (Marocco et al., 2015).

4.2.4 ENACT's Feedback and Debriefing

Immediate feedback is provided to the users in the form of the Agents' facial expressions, conveying emotions such as anger, sadness, or happiness. Moreover, at the end of each session, complete and articulated feedback is also provided. In the first debriefing page, ENACT provides overall feedback, which articulates into two different areas. The first area is related to the negotiation's ending point's direct evaluation, explaining why the conflict resolution was successful/unsuccessful. The second area is instead related to the Negotiation Style the user has selected as a final choice in the negotiation interaction and to the most appropriate Negotiation Style to be used in a situation similar to the one presented in the scenario (Fig.17a). The second debriefing page indicates the negotiation style used in every choice and feedback regarding the Negotiation pathway's consistency in terms of Negotiation Style adopted (Fig.17b, following page). Finally, the third debriefing page adds a reminder of each interaction's textual content to the indication of the negotiation style used in every choice with the Virtual Agent, as well as a theoretical description of the main characteristics of each Negotiation Style used (Fig.17c, following page). The user is free to navigate every choice's description.

Figure 17a. ENACT's first debriefing page. The blue area gives a direct evaluation of the negotiation's ending point. The red area states the user's final Negotiation Style and the most appropriate Negotiation Style to be used in a situation similar to the one presented in the scenario (in this case, Integrating) (Marocco et al., 2015).

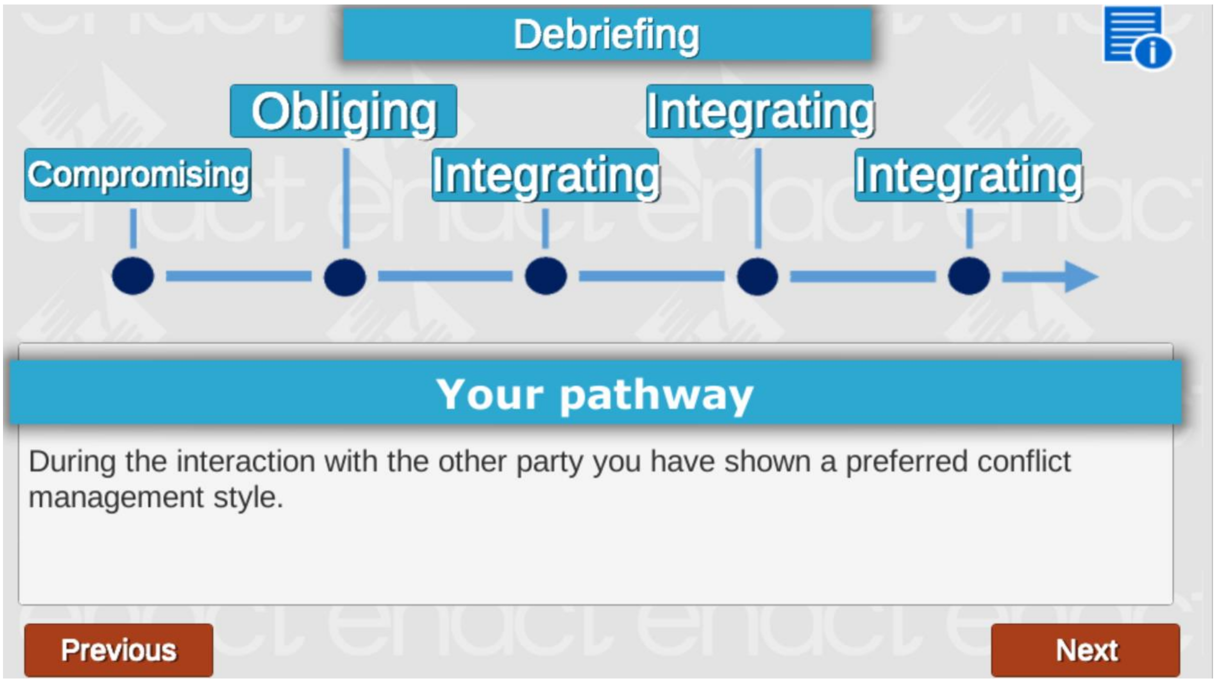


Figure 17b. ENACT's second debriefing page. The top area indicates the negotiation style used in every choice, while the bottom area gives a feedback regarding the consistency of the Negotiation pathway, in terms of Negotiation Style adopted (Marocco et al., 2015).

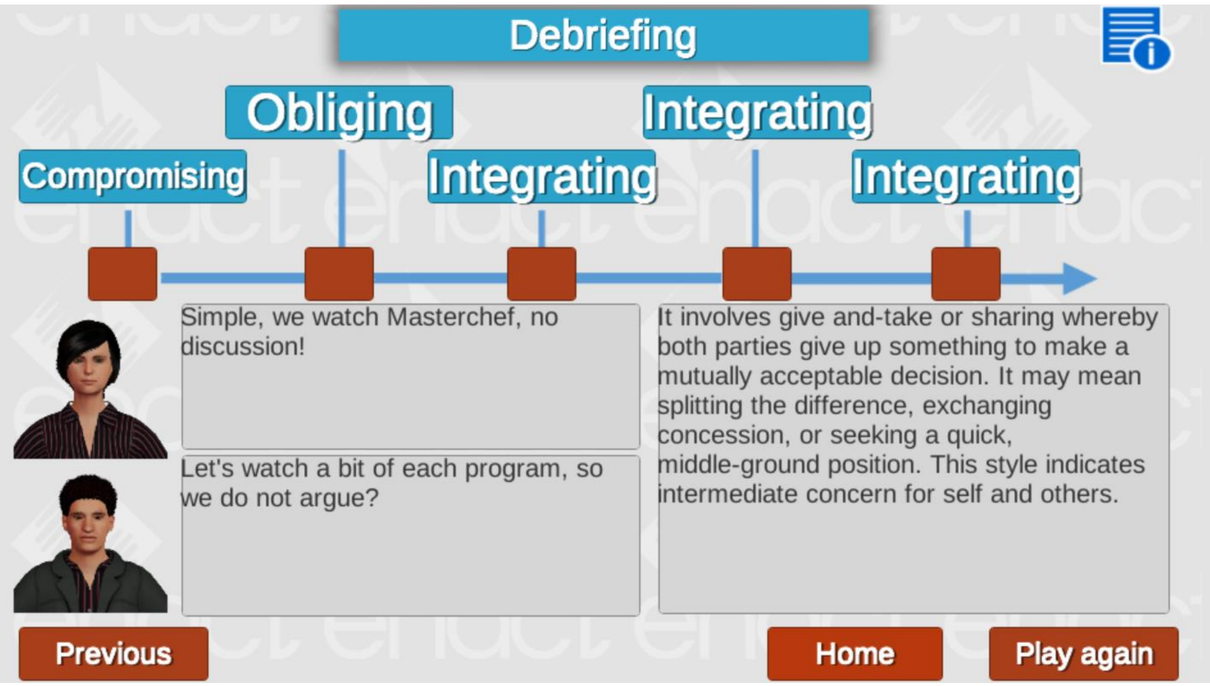


Figure 17c. ENACT's third debriefing page. The top area indicates the negotiation style used in every choice, the left-side bottom area gives a reminder of the textual content of each interaction, the righ-side bottom of the screen gives a theoretical description of the main characteristics of each Negotiation Style used (Marocco et al., 2015).

[this page intentionally left blank]

DIGITAL HUMANITY.
Do Users' Gaming Habits Affect the Perceived Human-Likeness of Virtual Agents in a Simulated Human Interaction?

SECTION 3

Empirical Works

[this page intentionally left blank]

Chapter 5

Game Habits, Blink-Related Eye Movements and the Perceived Human-Likeness of Virtual Agents in a Simulated Human Interaction

1. Introduction

1.1 Facial Expressions in Simulated Human Interactions

The necessity of addressing the blossoming social demands of a world set in the XXI century has driven attention towards effective training of Life Skills, defined as positive and versatile behaviors enabling people to successfully adapt to an ever-changing world (Jain & Anjuman, 2013). Among communication skills, effective Negotiation strategies, defined as the ability to discern the perspectives of different ethnicities, statuses, and cultural backgrounds (Dede, 2010), sustain the construction of solid interpersonal relationships, both in personal life and professional life (Buha, 2014; Casalnuovo, Vasilescu, Devanbu, & Filkov, 2015). Although there are advantages in using traditional training techniques, they are often time-consuming, expensive, and not sufficiently motivational for younger generations, and, therefore, they may not be sustainable for companies in the long run (Kron et al., 2017). In this context, Simulated Human Interactions come into play, as previous studies have suggested they improve users' competencies more efficiently and flexibly if compared to traditional teaching methods (e.g., Azadegan, Riedel, & Hauge, 2012).

Virtual Agents capable of realistically portraying emotional facial expressions are essential for the success of a technologically enhanced training, considering that facial expressions are a rich source of information during a negotiation (Cano & Sáenz, 1999; Datar, Garvin, & Cullen, 2010). However, according to the Uncanny Valley theorization, Virtual Agents whose movements, behavior, and appearance are almost but not wholly regarded as human are likely to cause a sense of repulsion or rejection in the user (Hodgins, Jörg, O'Sullivan, Park, & Mahler, 2010; Nowak & Fox 2018), to the point where the character's attractiveness and human-like appearance result to be negatively correlated (Scheneider, Wang, & Yang, 2007).

In human-to-human communication, people tend to explore the top half of the interlocutor’s face to search for emotional clues (Calder, Young, Keane, & Dean, 2000; Sullivan, Campbell, Hutton, & Ruffman, 2017; Birmingham, Svärd, Kanan, & Fischer 2018). Gazing direction and other eye dynamic movements in human-to-human communication represent valuable information and clues regarding a variety of internal states of both the talker and the listener during a conversation (e.g., Costa & Brunete, 2016; Hömke, Holler, & Levinson, 2017, 2018; Ruhland et al., 2014; Wood et al., 2015). The same attention pattern seems to be maintained when interacting with a digital human face (Tinwell, Grimshaw, Nabi, & Williams, 2011). Consequently, several studies have suggested how facial anomalies are more disturbing than even significant body motion errors (Hodgins, Jörg, O'Sullivan, Park, & Mahler, 2010; Lehmann, Roncone, Pattacini, & Metta, 2016; Donath, 2001). Among relevant eye movements for information gathering in human-to-human communication, we can find eye blinks (Costa & Brunete, 2016; Hömke et al., 2017, 2018; Ruhland et al., 2014).

1.2 Blink-Related Eye Movements in Humans

In addition to eye-wetting functions (Doane, 1980), blinking has been previously shown to accomplish social and cognitive functions during conversations (Homke, 2018; Lehmann, Pattacini & Metta, 2015). Studies have looked into the statistical distribution of eye blinking in human-to-human conversations, and results point to the fact that eye blinks frequency is far from being random. In particular, eye-blinks’ frequency seems to be related to attentional processes (Nakano, Kato, Morito, Itoi, & Kitazawa, 2013), fatigue (Johns, Tucker, Chapman, Crowley, & Michael, 2007), and lying (Burgoon, Buller, & Woodall, 1996). Neuroimaging studies have demonstrated a direct role of eye blinks in attentional engagement (Nakano, Kato, Morito, Itoi, & Kitazawa, 2013), as humans tend to blink more often when under low cognitive load (Siegle, Ichikawa, & Steinhauer 2008).

Concerning social functions, blink rate correlates positively with social group size in non-human primates (Tada, Omori, Hirokawa, Ohira, & Tomonaga, 2013), and neurological responses to blinks have been measured in an observer's brain during a conversation (Brefczynski-Lewis, Berrebi, McNeely, Prostko, & Puce, 2011). The highest blink rate in humans was recorded during conversations, rather than during other activities, such as looking at a target or reading (Doughty 2001; Dunbar, 1992). During speech production, most of the listener's eye blinks occur around the end of speaking units (Sacks, Schegloff, & Jefferson, 1974; Nakano & Kitazawa, 2010).

Blinks can also be categorized into short and long, according to a threshold of 410 ms (Homke, 2017). Long blinks usually occur during mutual gaze or in co-occurrence with feedback responses (e.g., nods), consequently affecting speakers' behavior (e.g., shorter answers are usually uttered following long blinks; Homke, 2017). For what is more, in natural human movement, the speed of upwards and downwards eyelid motions are different, with the latter phase being approximately twice as fast as the former (Evinger, Manning, & Sibony, 1991; Guitton, Simard, & Codère, 1991).

1.3 Blink-Related Eye Movements in Virtual Agents

Previous studies have demonstrated that blinking features positively affect users' reactions to Virtual Agents. For instance, virtual assistants in Microsoft Bob software for the Windows 95 operating system displayed "social idle" animations, including blinking, to make them appear more user-friendly, amusing, and natural (Winograd, 1996). Generally speaking, Agents' eye blinking behavior facilitates turn-taking in human-computer interactions (e.g., Watanabe & Higuchi, 1991) while increasing interaction comfort (DiSalvo, Gemperle, Forlizzi, & Kiesler, 2001; Dautenhahn et al., 2009; Metta et al., 2010).

Eye communicative acts, including blinking, couple a nonverbal expression with meaning as perceived by the users (Poggi, Pelachaud, & De Rosis, 2000), as blinking has been shown to accomplish both social and cognitive functions during conversations (Homke, 2018; Lehmann, Pattacini & Metta, 2015). Such a small movement seems, therefore, to influence users' impressions of the Agent, which can be perceived as concentrating heavily or day-dreaming (Corrigan, Peters, Küster, & Castellano, 2016), being nervous or friendly, and much more, depending on its blinking rate (Takashima, Omori, Yoshimoto, Itoh, Kitamura, & Kishino, 2008). For these reasons, complex automatic systems have been created for Virtual Agents to appropriately associate configurations of digital muscle group contractions with turn-taking, cognitive state, and particular emotions. Muscle contractions also include eye blinks, as fear has been shown to decrease the blinking rate, while anger and happiness to increase it (Nijholt, 2003). However, when such complex systems are not technologically reachable, a much simpler non-verbal behavior architecture based on the average human blinking rate and the average length of the inter-eye blink intervals has been previously used.

Unsurprisingly, robots with humanlike blinking behavior are perceived by users as being more intelligent compared to not-blinking or statistical blinking robots (Lehmann, Pattacini, & Metta, 2015; Roubroeks, Ham, & Midden, 2011; Roubroeks, Midden, & Ham, 2009). In previous research, users reacted to a text-only message, a blinking robot that spoke with a monotone voice, and a robot that showed head movements, gaze changes, and emotional intonation in the voice. Among the three conditions, the blinking robot provoked the lowest psychological reactance in users (Ghazali, Ham, Barakova, & Markopoulos, 2018). Nevertheless, according to a different study, when asked to assess the uncanniness of a Virtual Agent in which movements in the upper section of the face were too subtle or absent, users would especially consider uncanny virtual expressions in the lack of animations condition (Tinwell, Grimshaw, Nabi, & Williams, 2011; Dill, Flach, Hocevar, Lykawka, Musse, & Pinho, 2012).

Humanlike eye movements are of particular importance in the specific context of a Simulated Human Interaction revolving around the training of Negotiation strategies, as high levels of empathy and emotional recognition skills have been linked to effective use of conflict management strategies (Rahim, Psenicka, Polychroniou, & Zhao, 2002).

1.4 The Present Study

As previously described, ENACT's Virtual Agents display different eye expressions and gaze directions based on face-to-face observational data connected with the five Negotiation Styles theorized by Rahim (Rahim, 1983), while eye blinks were not included in the original version of the Simulated Human Interaction. Nonetheless, previous studies have both suggested that (i) an unblinking digital face may be "visually disconcerting" (e.g., Ruhland, 2014; Lehmann, Pattacini & Metta, 2015; Lehmann, Roncone, Pattacini & Metta, 2016) and (ii) blinking anthropomorphic Agents are considered to have a significantly higher agency and intelligence than the identical non-blinking forms (e.g., King & Ohya, 1996).

In natural human movement, the speed of upwards and downwards eyelid motions are different, and eye blinks frequency is far from being random, as previously stated. Therefore, complex automatic systems have been created for Virtual Agents to appropriately associate configurations of digital muscle contractions (including eye blinks) with particular emotions, turn-taking, and cognitive states (Nijholt, 2003). Nevertheless, implementing such a dynamic blinking distribution and duration in Virtual Agents of Simulated Human Interactions for educational purposes is often not possible due to a lack of time and monetary resources in the development team. In some previous instances, a system allowing Virtual Agents to blink at random times has been implemented, creating an illusion of responsive movements, as the more the on-screen agent blinks, the more it could occasionally blink in response to the subject's blink (e.g., Yoshikawa, Shinozawa, Ishiguro, Hagita & Miyamoto, 2006). Examples of the success of such a randomized system might be linked to a further important aspect of a Virtual Agent's perceived human-likeness: its agency, i.e., the user's *perception* of being interacting with a sentient being having abilities such as free will and "a mind of its own" (Guadagno, Blascovich, Bailenson, & McCall, 2007; Waytz, Cacioppo, & Epley, 2010). In fact, several studies have recorded different behaviors in users based on how they *perceived* the agent's agency independently from its being controlled by a human or being scripted (e.g., Gratch et al., 2015). Based on the literature mentioned above, two separate versions of ENACT have been created: in the first version, the Agents did not blink, while in the second version, the Agent would blink at random times.

1.5 Objectives and Hypotheses

Replicating the level of human-likeness in the Virtual Agent's appearance and behavior required to "convince users' perceptual systems that a virtual human is the real thing" (Ruhland et al., 2015) remains a contemporary challenge. Flawed Virtual Agents may negatively affect the user's perceived immersion into the Simulated Human Interaction and disrupt its simulative nature by impeding generalizability to real-life situations (Von Bergen, 2010; MacDorman & Ishiguro, 2006). Moreover, according to a recent report, players consider the quality of graphics to be the most essential aspect to consider before purchasing a video game, even above the quality of storytelling (Entertainment Software Association, 2017). The Entertainment Video game industry can increasingly rely on technological advancements and equipment, such as Motion Capture, to realistically design and portray Virtual Agents as avatars and non-playable characters. This sustains the market trend that sees video games with realistic graphics as the highest-grossing category during the last three years and more. On the contrary, advanced technology and statistical models allowing for high-fidelity face synthesis are often not available to smaller research laboratories developing Simulated Human Interactions for educational purposes.

The first objective of the present experimental study is to explore whether the introduction of a statistical blinking feature in a Virtual Agent will be enough to sustain better acceptability and quality of the interaction, therefore, constituting a possible easy win on the short development run for smaller research labs, not requiring expensive interventions in terms of money and time.

A second objective is to investigate whether the Entertainment Video Game Industry is posing higher standards for Virtual Agent human-likeness, and, therefore, if video game players who are used to interacting with state-of-the-Art Virtual Agents have harsher reactions to Simulated Human Interactions for educational purposes if compared with non-habitual video game players.

Specifically, it is hypothesized that:

Ia) Participants will evaluate the quality of the product more positively if a statistical Blink feature is added to the Virtual Agent if compared to the non-blinking Agent;

Ib) Participants will evaluate the Acceptability of the Virtual Agent's Appearance more positively if a statistical Blink feature is added to the Virtual Agent if compared to the non-blinking Agent

As previous studies have suggested that an unblinking digital face may be "visually disconcerting" (e.g., Ruhland, 2014; Lehmann, Pattacini & Metta, 2015; Lehmann, Roncone, Pattacini & Metta, 2016).

IIa) Habitual video game players will evaluate the quality of the product more negatively than non-habitual video game players;

IIb). Habitual video game players will evaluate the Acceptability of the Virtual Agent's Appearance more negatively than non-habitual video game players.

Because of their prolonged exposure to the state-of-the-art face simulations in entertainment video games involving human-Virtual Agent interactions.

2 Methods

2.1 Study Design

A 2x2 between-subjects study design (see Fig.18, following page) has been carried out entirely online due to movement limitations following the Coronavirus CoViD-19 outbreak.

The following independent variables have been considered:

1. **Game Habits:** Participants declaring a minimum of 5 hours of story-based video game-play per week were included in the habitual video game players group (VGP), participants declaring fewer hours of story-based gameplay per week were included in the non-habitual video game players (NVGP). The habitual/non-habitual group threshold, i.e., 5 hours per week, has been selected based on data retrieved from 18-35 year-olds across six countries (Germany, UK, US, France, Japan, South Korea; Limelight Networks, 2018) as well as on a variety of previous studies related to the topic (e.g., Braun, Stopfer, Müller, Beutel, & Egloff, 2016; Kim et al., 2015; Konishi et al., 2015). Although recruiting non-players, as opposed to non-habitual players, might be considered a more precise differentiation, there is an estimated 2.47 billion video game players globally (Statista, 2019) and it has become increasingly rare to encounter participants who do not play at all. Moreover, its wide use in recent literature has consolidated this threshold as the standard modality to discriminate habitual and non-habitual players (e.g., Torner, Carbonell, & Castejón, 2019; Cain, Prinzmetal, Shimamura, & Landau, 2014). The target game genre was identified in story-based games, as they are more likely to include close-ups of digital human faces and, therefore, increase the time of exposure of video game players to state-of-the-art Virtual Agents;

2. **No-blink vs. Blink**: Participants were randomly assigned to one of two different versions of ENACT, in which, respectively, the Virtual Agent confronted by the user will blink or not blink during the interaction.

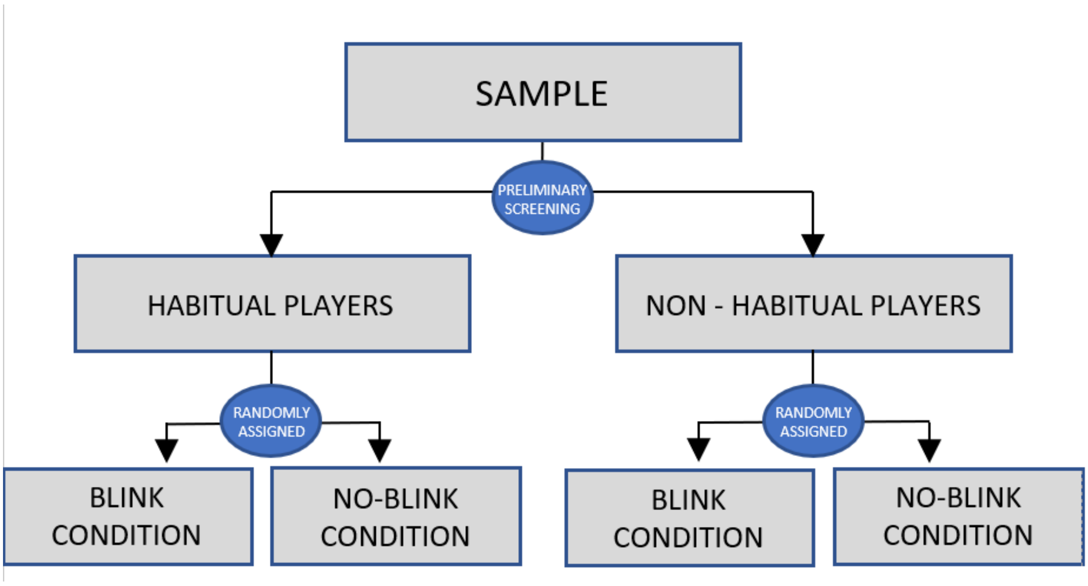


Figure 18. Graphical representation of the Study Design

2.2 Participants

A sample of 50 young male participants (mean age = 29.36; SD = ±3.67) has been recruited and divided into two experimental groups. Restriction to male participants was introduced because of the relative difficulty to find female habitual video game players, as the difference between weekly gameplay hours of habitual and non - habitual female players resulted not to be sufficiently large.

Participants declaring through a preliminary questionnaire a minimum of 5 hours of video game-play per week were included in the habitual video game players group. Participants declaring fewer hours of gameplay per week were included in the non-habitual video game players group. Participants were unaware of the study's real objectives, as it was presented as an experimental study related to conflict resolution styles. No economic rewards were provided during the research; however, feedback regarding the conflict resolution styles used while negotiating with the Virtual Agent, as provided by ENACT's second debriefing page, was provided at the end of the session.

To be included in the study, individuals had to meet the following criteria:

(1) age between 18 and 35 years old;

(2) no significant visual and auditory impairment (all with normal or corrected-to-normal visual acuity)

Before partaking, all participants were provided with written information about the study and were required to give written consent to be included. The study received ethical approval from the Ethical Committee of the University of Milano-Bicocca (protocol number 503, deliberation date December 5th, 2019).

2.3 Materials

ENACT (Marocco et al., 2015), a Simulated Human Interaction-based Negotiation training, has been used as the main object in this study. ENACT's interaction modality is based on a multiple-choice approach. During the Simulated Human Interaction, the Virtual Agent is shown on the screen's right-hand side, and their utterances are displayed in written form in a balloon above their head. The Agent communicates through a combination of four different facial expressions (neutral, happy, sad, or angry), 24 automatically implemented gestures and mouth articulations, and ten different body postures and relative gaze directions connected with the five Negotiation Styles theorized by Rahim (Rahim, 1983). The user's avatar is portrayed on the left side of the screen, in the upper corner. Below the user's avatar, there are five options through which the user can continue the interaction (Fig.19, following page). Each option is linked to a specific Negotiation Style, as theorized by Rahim (Rahim, 1983). In the selected scenario, participants are asked to play as themselves and negotiate with the Virtual Agent, their roommate Anna, which TV program they should watch that night (i.e., a news program or a cooking reality show). Participants take five conversation turns. Following social distancing rules due to the COVID-19 pandemic, participants used their personal computer (minimum requirements: IntelCore i5 with 8GB of RAM, 64-bit version of Windows 8).

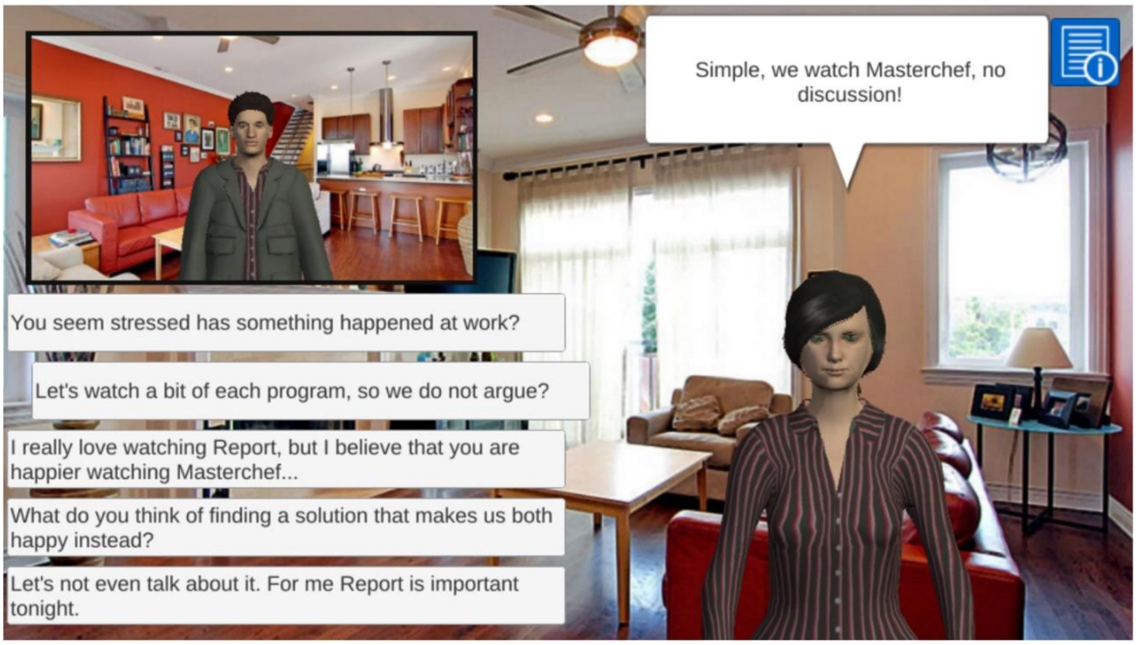


Figure 19. A screenshot of ENACT's TV program scenario used during the experimental procedure (Marocco et al., 2015).

2.4 Measures

- *Preliminary Measures of Video Game Habits*

Participants were assigned to one of the two groups according to their answer to a preliminary question related to their habitual video game habits. Participants were asked only to consider video games involving human NPCs, to measure exposure to state-of-the-Art digital simulations of human faces. Those declaring more than 5 hours of gameplay per week during the past month were included in the habitual video game players group.

- *Demographics*

Demographic information was considered through a dedicated questionnaire, including questions about the participant's gender and age.

- *Measures of Acceptability of the Virtual Agent*'s *Appearance:*

Evaluation of the physical appearance of the Agent: three items with bipolar anchors on a 9-point scale asked participants to rate the Virtual Agent Realism (as in Dill et al., 2012; Tinwell, Grimshaw, Nabi, & Williams, 2011), Likeability, and Reassurance (as in MacDorman, Green, Ho, & Koch, 2009).

Evaluation of the perceived Comfort during the Interaction: Three items related to the Comfort, Naturalness, and Ease perceived during the interaction, on a 5-point Likert scale (as in Gratch, DeVault, & Lucas, 2016).

An additional multiple-choice question was proposed to the participant. Users were asked to indicate which part of the Agent's face appeared as the most bizarre, if eyes, mouth, nose, hair, or other features (as in Dill, Flach, Hocevar, Lykawka, Musse, & Pinho, 2012), in the last case leaving a blank space for an open answer.

- *Measures of Perceived Quality of the Product*

The Game Experience Questionnaire (IJsselsteijn, de Kort, & Poels, 2013) has been used: a 33-item questionnaire on a 5-point Likert scale, measuring five subscales related to Competence, Immersion, Flow, Tension, Challenge, Negative Affect, and Positive Affect. Competence refers to the feeling of the participants of being able to accomplish the proposed tasks during the interaction; Immersion and Flow represent two different declinations of the participant's "sense of being there" (the first one related to space and the second one related to time), Tension to the anxiety experienced while interacting; Challenge to the lack of balance between the demands and the participant's skills; and, finally, Negative and Positive Affects refer to the negative and positive emotional spheres.

Likelihood to recommend: the Net Promoter Score, a single question asking the user to state, on a scale from 1 to 10, how likely they are to recommend the training to a friend or colleague interested in improving their conflict resolution style (Reichheld, 2003).

2.5 Procedure

During the first contact for recruitment, participants were preliminary asked to evaluate their mean weekly gameplay hours to be assigned to the habitual or non-habitual video game players. Once randomly assigned to the No-Blink or Blink group, participants were contacted through Google Meet while being individually seated at a desk in a quiet room in their house. Participants used their personal computers (minimum requirements: IntelCore i5 with 8GB of RAM, 64-bit version of Windows 8), and the researcher observed their actions through the screen sharing feature, following social distancing rules. After answering the demographic questionnaire and the video game habits questionnaire, participants were instructed about ENACT's structure, controls, and the interaction scenario. During the Interaction, the researcher muted their microphone and suspended communications with the participant for the Interaction duration. ENACT's audio was also muted during the interaction to prevent the risk of including confounding variables. After ending their interaction, participants completed the questionnaires related to the Agent's acceptability evaluation.

All participants were asked to complete the questionnaires related to their Player Experience and to their likelihood to recommend the training product to a friend or colleague interested in training their conflict resolution strategies. Feedback regarding the conflict resolution styles used while negotiating with the Virtual Agent was provided at the end of the session. A graphical representation of the experimental procedure can be found on the following page (Fig.20).

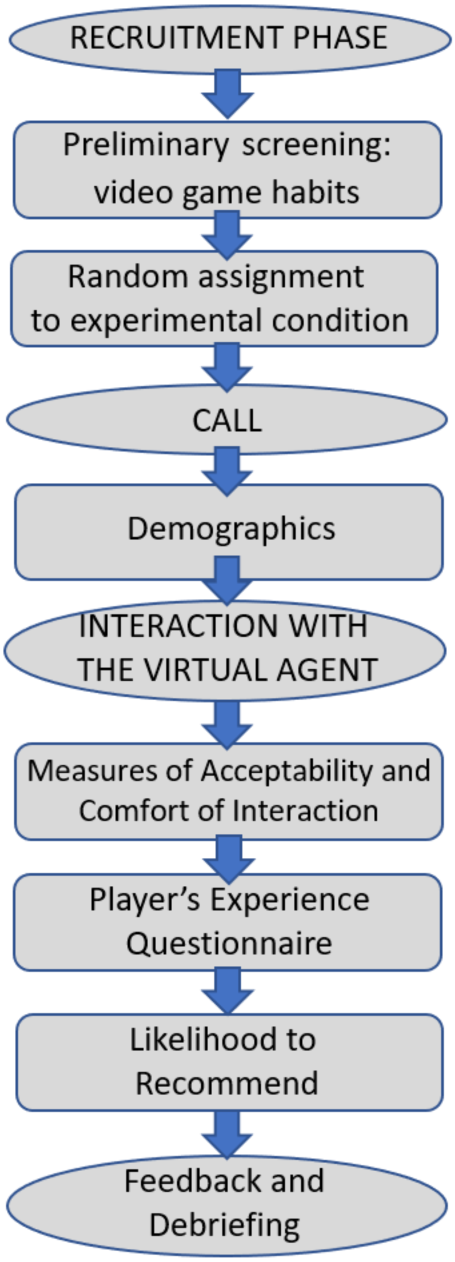


Figure 20. Graphical representation of the Experimental Procedure.

3. Results

Participants have been divided into a habitual video game players group (n = 25; mean age = 29.48; SD = ±4.14) and a non-habitual video game players group (n = 25; mean age = 29.24; SD = ±3.21), according to their declared weekly gameplay hours. The mean of weekly gameplay hours for habitual video game players was 15.64 (SD = ±6), while for non-habitual video game players was 1.83 (SD = ±1.83).

Statistical analyses were conducted to test the hypothesized main effect between game habits and the participants' player experience, likelihood to recommend, and perceived acceptability of the Virtual Agent, and between the two experimental conditions and the participants' player experience, likelihood to recommend, and perceived acceptability of the Virtual Agent. Interaction effects between the two independent variables were also explored. A one-way between subjects' ANOVA was conducted in SPSS v. 26 at the $p<.05$ level of significance.

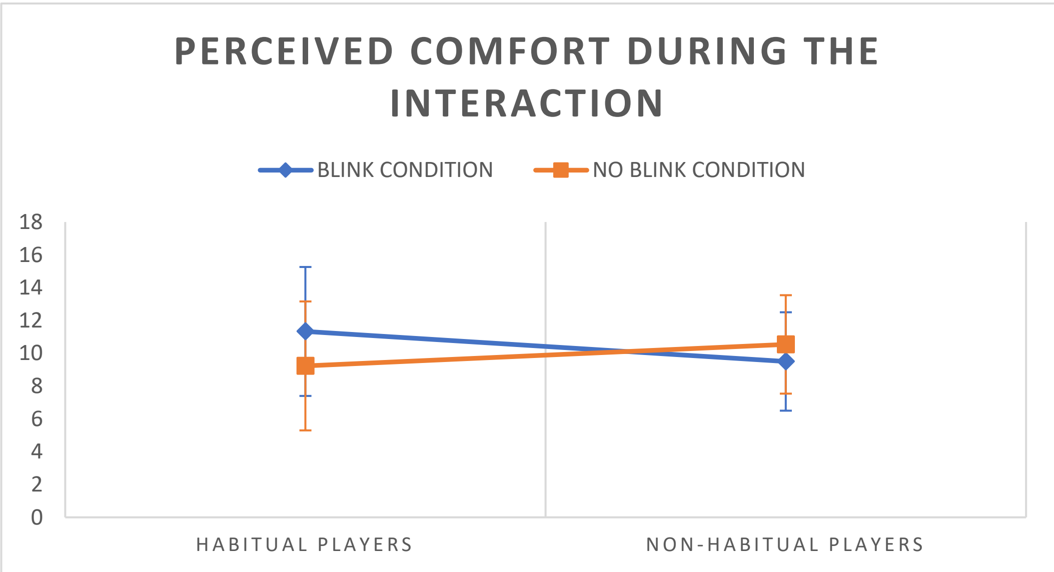
3.1 Perceived Acceptability of the Virtual Agent's Appearance

Statistical analyses revealed a significant interaction effect between experimental condition and game habits on perceived Comfort while interacting with the Agent. Results are summarized in tables 4a and 4b, page 92 of the present dissertation.

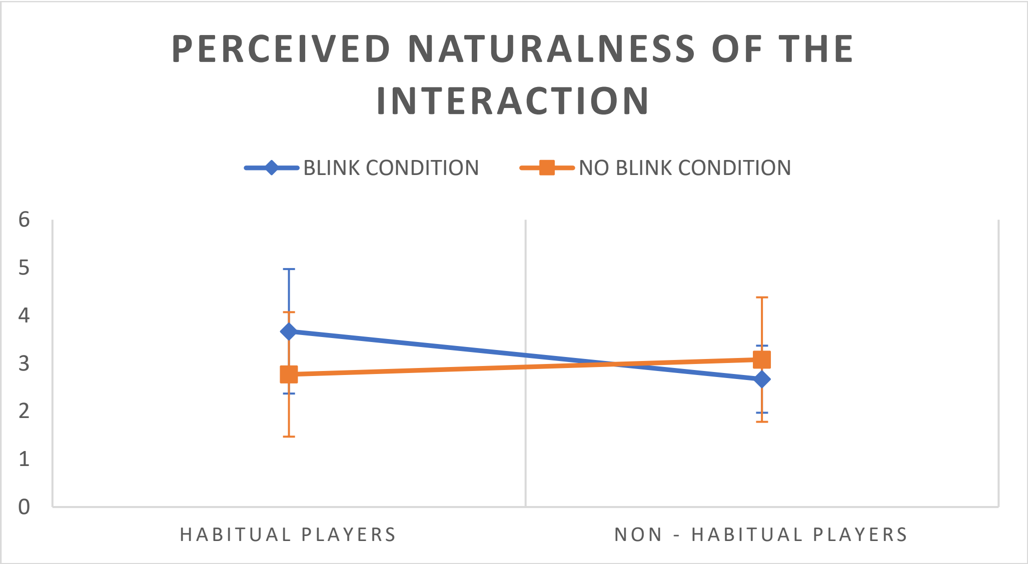
In particular:

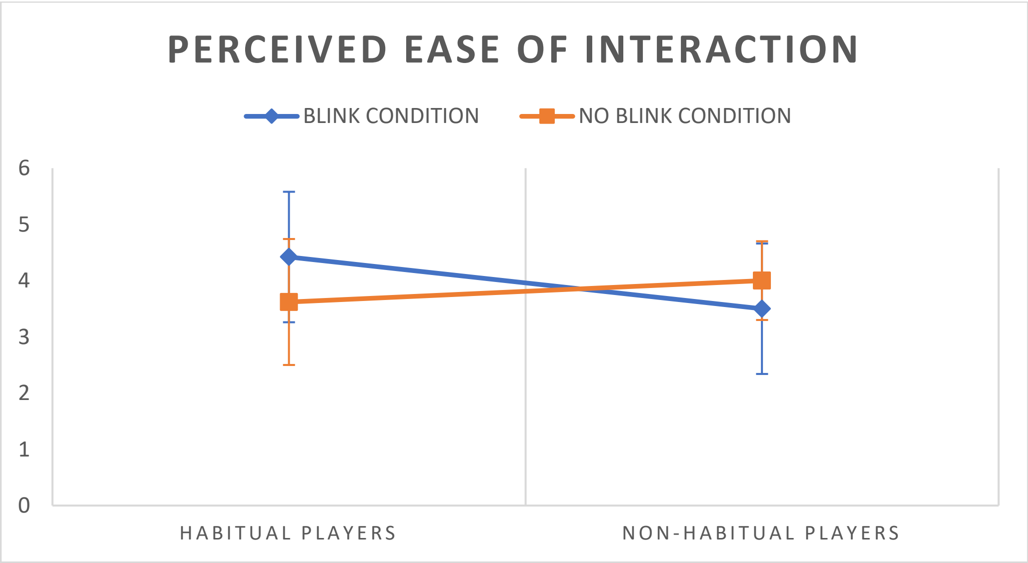
- **Evaluation of the Virtual Agent's physical appearance**: Analyses did not reveal statistically significant differences in terms of game habits [$F(1, 48) = .600$, $p = .44$] or or experimental conditions [$F(1, 48) = .48$, $p = .48$]. Interaction effects between the two variables were also explored, but analyses did not return statistically significant effects [$F(1, 48) = .020$, $p = .88$]. However, it has to be noted that there was a quasi-significant effect of the experimental condition on the *Realism* Item at the $p<.05$ level [$F(1, 48) = 3.60$, $p = .06$], with scores being significantly different in the Blink condition (mean = 4.50; SD = ±2.06) if compared to the No-Blink Condition (mean = 3.58; SD = ± 1.73).

- **Evaluation of the perceived Comfort during the interaction**: Analyses did not reveal statistically significant differences in terms of game habits [$F(1, 48) = .113$, $p = .74$] or experimental conditions [$F(1, 48) = .463$, $p = .50$]. An interaction effect between game habits and experimental condition on perceived Comfort of the Interaction was found to be close to statistical significance [$F(1, 48) = 4.03$, $p = .05$]. In particular, there was a quasi-significant interaction effect for the *Naturalness* Item at the $p<.05$ level [$F(1, 48) = 3.58$, $p = .065$]. A significant interaction effect was found for the *Ease* Item at the $p<.05$ level [$F(1, 48) = 4.76$, $p = .034$]. However, post-hoc analyses (Tukey's HSD) did not report statistically significant differences between groups. Results are presented in graphs 2 a, b, c, following pages.



Graph 2a. Comfort during the Interaction as perceived by Habitual and Non-Habitual Video Game Players in the two experimental conditions.





Graph 2b, 2c. Single-item results: Naturalness (above) and Ease (below) during the Interaction as perceived by Habitual and Non-Habitual Video Game Players in the two experimental conditions.

Measure	Habitual Players		Non-Habitual Players		$F(1, 48)$	η_p^2	Power
	M	*SD*	*M*	*SD*			
Physical Appearance	11.60	3.85	12.44	3.74	.600	.013	.118
Realism	3.92	1.77	4.12	1.71	.167	.004	.069
Likeability	3.68	1.56	4.20	1.47	.546	.012	.112
Reassurance	3.80	1.38	4.12	1.50	.598	.013	.118
Perceived Comfort	10.24	3.31	10.04	2.26	.113	.002	.062
Ease	4.00	1.19	3.76	.970	.790	.017	.141
Naturalness	3.20	1.38	2.88	1.09	1.00	.021	.165
Comfort	3.04	1.33	3.40	1.04	1.03	.022	.169

Table 4a. Means, Standard Deviations, and One-Way Analyses of the Agent's physical appearance and comfort perceived during the Interaction in terms of Game Habits.

Measure	Blink Condition		No-Blink Condition		$F(1, 48)$	η_p^2	Power
	M	*SD*	*M*	*SD*			
Physical Appearance	12.42	4.28	11.65	3.29	.600	.010	.105
Realism	4.50	2.06	3.58	1.24	3.60**	.073	.459
Likeability	3.96	1.70	4.12	1.33	.128	.003	.064
Reassurance	3.96	1.45	3.96	1.45	.000	.000	.050
Perceived Comfort	10.42	3.00	9.88	2.64	.460	.010	.102
Ease	3.96	1.23	3.81	.939	.255	.006	.079
Naturalness	3.17	1.17	2.92	1.32	.497	.011	.106
Comfort	3.29	1.27	3.14	1.15	.161	.003	.068

*Table 4b. Means, Standard Deviations, and One-Way Analyses of the Agent's physical appearance and comfort perceived during the Interaction in terms of Experimental Condition. ** indicates a quasi-statistically significant difference.*

When asked to point out the most bizarre graphical element in the Virtual Agent, both habitual and non-habitual video game players chose her eyes. On the contrary, habitual and non-habitual video game players indicated Anna's hair most frequently as the most bizarre element in the Blink condition (See Table 3).

GROUP	MOST "BIZARRE" GRAPHICAL ELEMENT
Habitual Video Game Players	**No Blink:** eyes (x5), hair (x4), mouth (x2), arm movement (x1), skin tone (x1)
	Blink: hair (x5), eyes (x3), mouth (x2), clothes (x1), nose (x1)
Non-Habitual Video Game Players	**No Blink**: eyes (x3), hair (x2), clothes (x2), bust (x2), mouth (x2)
	Blink: hair (x7), arm movement (x2), mouth (x1), eyes (x1), clothes (x1)

Table 3. The Virtual Agent's most bizarre graphical feature as indicated by habitual and non-habitual video game players in each experimental condition.

3.2 Perceived Quality of the Product

Analyses revealed a statistically significant effect of game habits on perceived Negative Affect while interacting with the Agent, while did not reveal statistically significant effects of game habits or experimental conditions on player's experience components or likelihood to recommend. Results are summarized in tables 5a and 5b, page 95 of the present dissertation. In particular:

- **Competence:** Analyses did not reveal statistically significant differences in terms of game habits [$F(1, 48) = .676$, $p = .41$] or experimental conditions [$F(1, 48) = .024$, $p = .87$]. Interaction effects between the two variables were also explored, but analyses did not return statistically significant effects [$F(1, 48) = .932$, $p = .34$];

- **Immersion:** Analyses did not reveal statistically significant differences in terms of game habits [F(1, 48) = .001, p = .98] or experimental conditions [F(1, 48) = .413, p = .52]. Interaction effects between the two variables were also explored, but analyses did not return statistically significant effects [F(1, 48) = .002, p = .97];

- **Flow:** Analyses did not reveal statistically significant differences in terms of game habits [F(1, 48) = .707, p = .40] or experimental conditions [F(1, 48) = .001, p = .97]. Interaction effects between the two variables were also explored, but analyses did not return statistically significant effects [F(1, 48) = .072, p = .78];

- **Challenge:** Analyses did not reveal statistically significant differences in terms of game habits [F(1, 48) = 1.73, p = .19] or experimental conditions [F(1, 48) = .432, p = .51]. Interaction effects between the two variables were also explored, but analyses did not return statistically significant effects [F(1, 48) = .063, p = .80];

- **Positive Affect:** Analyses did not reveal statistically significant differences in terms of game habits [F(1, 48) = .499, p = .48] or experimental conditions [F(1, 48) = .097, p = .75]. Interaction effects between the two variables were also explored, but analyses did not return statistically significant effects [F(1, 48) = .499, p = .48];

- **Negative Affect:** Analyses revealed a statistically significant difference in terms of game habits [F(1, 48) = 5.43, p = .024]. In particular, the mean score for habitual video game players (mean = 1.39; SD = ±0.48) was significantly different than the score of non-video game players (mean = 1.80; SD = ± 0.45). The effect size for this analysis (*d* = 0.88) was found to exceed Cohen's (1988) convention for a large effect (*d* = 0.80). However, no statistically significant difference was found in terms of experimental conditions [F(1, 48) = .000, p = .99]. Interaction effects between the two variables were also explored, but analyses did not return statistically significant effects [F(1, 48) = .579, p = .45];

- **Tension:** Analyses did not reveal statistically significant differences in terms of game habits [F(1, 48) = .323, p = .57] or experimental conditions [F(1, 48) = .332, p = .56.]

Interaction effects between the two variables were also explored, but analyses did not return statistically significant effects [F(1, 48) = .083, p = .77];

Concerning the Likeliness to Recommend, analyses did not reveal statistically significant differences in terms of game habits [F(1, 48) = .676, p = .41] or experimental conditions [F(1, 48) = .024, p = .87]. Interaction effects between the two variables were also explored, but analyses did not return statistically significant effects [F(1, 48) = .099, p = .75].

Measure	Habitual Players		Non-Habitual Players		*F*(1, 48)	η_p^2	Power
	M	*SD*	*M*	*SD*			
Likelihood to Recommend	7.16	2.65	7.44	1.26	.206	.004	.073
Players' Experience							
Competence	2.95	.83	3.15	.78	.676	.014	.127
Immersion	2.79	.70	2.79	.59	.001	.000	.050
Flow	3.35	.97	3.10	1.04	.707	.015	.131
Challenge	1.91	.63	2.17	.76	1.74	.036	.252
Positive Affect	2.93	.89	2.77	.59	.499	.011	.106
Negative Affect	1.49	.48	1.80	4.56	5.43*	.106	.626
Tension	1.97	1.02	2.12	.79	.323	.007	.086

*Table 5a. Means, Standard Deviations, and One-Way Analyses of the Likelihood to recommend and the players' experience components in terms of Game Habits. * indicates a statistically significant difference.*

Measure	Blink Condition		No-Blink Condition		*F*(1, 48)	η_p^2	Power
	M	*SD*	*M*	*SD*			
Likelihood to Recommend	7.29	2.40	7.31	2.06	.001	.000	.050
Players' Experience							
Competence	3.03	.97	3.06	.64	.024	.001	.053
Immersion	2.85	.76	2.73	.53	.413	.009	.096
Flow	3.23	1.09	3.22	.94	.001	.000	.050
Challenge	1.97	.74	2.11	.68	.432	.009	.099
Positive Affect	2.88	.84	2.81	.68	.970	.002	.061
Negative Affect	1.65	.57	1.64	.41	.000	.000	.050
Tension	1.97	.93	2.12	.89	.332	.007	.087

Table 5b. Means, Standard Deviations, and One-Way Analyses of the Likelihood to recommend and the players' experience components in terms of Experimental Conditions.

4. Discussion

The present experimental study supports that the introduction of a randomized blinking feature affects the Virtual Agent's perceived acceptability. Regardless of gaming habits, the blinking Virtual Agent was considered more realistic than the non-blinking Agent. Moreover, the Agent's eyes were constantly considered the most bizarre element by participants when the Agent did not blink. Users' negative attention was diverted somewhere else when the Agents displayed blink-related eye movements (i.e., hair). This seems to confirm that users were highly drawn towards examining the Agent's eye movements even without being instructed to do so, similarly to what they would do during a human-to-human conversation (Costa & Brunete, 2016; Hömke et al., 2017, 2018; Bartneck, Kanda, Ishiguro, & Hagita, 2009). More importantly, this also seems to confirm that subtle typically human characteristics such as eye blinking, even if random, can affect the user's perception of the Agents, but only to a certain extent. Blinking eyes were, in fact, still considered bizarre by some participants.

Interestingly, habitual and non-habitual video game players showed opposite tendencies in perceived Comfort during the interaction. In particular, habitual video game players reported the interaction to be easier and more natural when the Virtual Agent would blink, while non-habitual players reported preferring the non-blinking Agent in terms of ease and naturalness of the interaction. These results seem to interestingly suggest that players and non-players might experience the "too real for comfort" zone differently. A possible explanation might be found in the two experimental groups applying different standards during the Virtual Agent evaluation, as suggested before (e.g., Bartneck, Kanda, Ishiguro, & Hagita, 2009). According to a previous study, a Virtual Agent's face with limited movement was considered less human-like than a human displaying full-face movement, as expected from Mori's theorization. However, even though the limited-movement Agent failed to move his gaze between the communication partner and the environment, as expected from a real human, it was not perceived as significantly less likable than the human (Bartneck, Kanda, Ishiguro, & Hagita, 2009). By apparently contradicting Mori's original theorization of the Uncanny Valley, such results might suggest the intervention of further factors in the assumption of human-likeness.

Non-video game players might have applied standards and expectations related to the "human" category to evaluate the Virtual Agent. Therefore, the blinking movement might fall just into the Uncanny Valley and, as a consequence, they were more comfortable with the moderate realism of the Non-blinking Agent, placed “just before the Valley” (Seymour, Riemer, & Kay, 2017). On the contrary, because of their previous experience with Non-Playable Characters during gameplay, habitual video game players are likely to apply standards and expectations related to the “Virtual Agent” category when interacting with one. Expectations for this category are consequently adjusted for the possibility of imprecise movements compared with a real human, as players are aware of technological limitations, resulting in less stringent social standards in terms of Virtual Agent acceptability. Just like “we (…) do not expect our cats to follow the human standard of looking alternately at the speaker and the environment”, players might think similarly of an imprecisely blinking Virtual Agent (Bartneck, Kanda, Ishiguro, & Hagita, 2009). Therefore, Hypothesis II, stating that Virtual Agents in the No blink condition would be evaluated worse in terms of acceptability, seems to need rephrasing, and taking into consideration the target audience’s previous expertise in gaming seems to be particularly important to design Virtual Agents that avoid rousing feelings of uncanniness.

The present results seem to deny differences in the perception of the product's general quality due to its users’ game habits. However, non-video game players reported higher levels of Negative Affects (negative emotional spheres) while interacting. This might indicate that Non-habitual video game players were more sensitive to the Agent’s aggressive negotiation style (i.e., Dominating). In turn, such an increased sensitivity might be due to non-players being less used to the setting of a Simulated Human Interaction.

5. Conclusions

The following paragraphs will trace conclusions gathered from the present experimental study.

- **Random Eye Blinks in Virtual Agents Affect Their Perceived Realism:** Subtle typically human characteristics such as random eye blinking can only affect the user's perception of the Agents' realism to a certain extent. The fact that the blinking eyes were still considered bizarre by some participants suggests that the implementation of random eye blinks may not be a sufficiently efficient, cost-effective solution for independent developers seeking to increase the perceived human-likeness of the Agents. Although results from the present study are encouraging, they seem to confirm that the mere frequency of positive feedback in Virtual Agents is not as crucial as the feedback being coupled to what the speaker is doing at the moment (i.e., contingency) when it comes to creating affinity between a user and a Virtual Agent (e.g., Gratch, Wang, Gerten, Fast, & Duffy, 2007).

- **Gaming Habits May Posit Different Category Boundaries for Virtual Agents:** Game habits might not simply posit higher standards to players when evaluating a Virtual Agent's acceptability. Players and non-players might experience the Uncanny Valley differently, applying different reference systems to Virtual Agents, leading to the assignment of different social standards and normative expectations for Virtual Agents (as in Bartneck, Kanda, Ishiguro, & Hagita, 2009). This is yet another evidence pointing towards the complexity and dimensionality of the perception of an Agent's human-likeness, introducing game habits as another relevant intervening factor in the perception of uncanniness. Therefore, considering the target audience's previous gaming expertise seems to be particularly important to design Virtual Agents that avoid rousing feelings of uncanniness.

6. Limitations of the Present Study

Although carefully conducted, the present study is not without limitations. It should be noted that, between 2019 and 2020, novel coronavirus disease 2019 (COVID-19) resulted in months-long lockdowns in many countries, including Italy (Soraci et al., 2020). This unexpected situation has led to a reorganization of the present study's procedure, which has been adapted to an online setting. Several matters should be considered, such as the participants' possibly different Internet speeds while playing ENACT, leading to smoother or bumpier movements of the Virtual Agent, and differences in screen sizes, as participants played on their personal computer. Future studies should be directed towards confirming the present results in a laboratory setting.

Due to its online setting, only quantitative self-report assessments have been used in the present study. Future studies should include qualitative measures providing more in-depth and more meaningful insights from participants. For instance, the recording of spontaneous verbalizations has been widely used in previous studies, as it provides information about the honest feelings of happiness or frustration of participants without breaking immersion into the interaction (e.g., O'Hagan, 2009). Moreover, users' facial expressions can be video recorded throughout the interaction and coded according to Facial Action Units, the specific combinations of which identify seven different basic emotions: happiness, sadness, surprise, anger, fear, disgust, and contempt (Ekman, 1992). Future directions might also include psychophysiological measures (e.g., heart and breathing rate, skin conductivity), provided they are not invasive and do not excessively affect the immersion in the Simulated Human Interaction. These data are potentially very informative on the user's internal states and have been previously used in various studies related to gameplay emotions (e.g., Pallavicini, Ferrari, Pepe, Garcea, Zanacchi, & Mantovani, 2018).

The study was conducted, and the analyses were based on data from a small self-selected sample of participants. Consequently, results have indicated tendencies, and statistically significant differences were few. Further investigation of a larger sample is needed to confirm the preliminary results provided by the present study. Several participants have spontaneously reported they would have liked to spend more time using ENACT before answering the questionnaires. Therefore, the introduction of more prolonged interactions could be useful to confirm the present results.

[this page intentionally left blank]

Chapter 6

Game Habits, Gibberish Voice, and Perceived Human-Likeness of Virtual Humans in a Simulated Human Interaction Training for Negotiation Skills

1. Introduction

Simulated Human Interactions have been shown to improve users' competencies more efficiently and flexibly than traditional teaching methods (e.g., Azadegan, Riedel, & Hauge, 2012). A relevant component of effective Negotiation training is the transition from a self-centered approach to considering all the involved parties' values and needs (Monoriti & Gabellini, 2018). A vital role in improving such ability is played by the individual's social perception, defined as "the ability to understand and react appropriately to the social signals sent out by other people vocally, facially, or through body posture" (Mayer et al., 2016). In this sense, Virtual Agents capable of realistically portraying vocal non-verbal components of communication, such as pitch, volume, pace, tone, and timbre, are crucial for the success of technologically enhanced Conflict Resolution training.

However, the concept of an audio Uncanny Valley has been introduced in the last few years, according to which "the closer simulated sound comes to natural sound, the more human ears pick up on the subtle differences that mark it as strange" (Ramsey, 2014). For what is more, vocal non-verbal aspects of communication of the Agent's voice have also been shown to affect users' perception of Virtual Agents and their assumptions over their perceived personality traits. For instance, tone shallowness and lack of emotional expression negatively affected users' level of perceived friendliness (Druga, Williams, Breazeal, & Resnick, 2017).

1.1 Vocal Non-Verbal Aspects of Human-Virtual Agent Communication

As information conveyed during interactions among people is intrinsically multimodal, Virtual Agents should express themselves in an equally multimodal manner to simulate a rich and intelligent-like interaction and to make their message more understandable (Xie, Jia, Meng, Deng, & Wang, 2015; Knapp, Hall, & Horgan, 2013). Virtual Agents in the context of

Simulated Human Interactions can be voiced through different techniques: by a recording of a human voice, by voice conversion (a procedure that directly records and slightly modifies a human voice), by concatenating pieces of recorded human speech stored in databases, or through the use of a computer-synthesized voice (Wu, Ning, Zang, Jia, Meng, Meng, & Cai, 2015). Depending on Virtual Agent's voice's resulting quality, studies have indicated different trends related to users' perception of the Virtual Agent's human-likeness. For instance, an older computer-synthesized voice resulted in an inferior evaluation of the Agent as a conversational partner compared to a human voice (e.g., Mayer, 2014), even lower than the evaluation of Agents relying on purely text-based communication (Gong & Nass, 2007). On the contrary, a Virtual Agent speaking through a modern text-to-speech voice engine was rated as credible as an Agent with a human voice, outperforming an Agent speaking through an older text-to-speech voice engine (Craig & Schroeder, 2017).

Even when aware that a Virtual Agent's voice is undoubtedly synthetic, users tend to assign human properties to the Agents' voice, primarily based on non-verbal vocal communication elements (Short, 2017). These human properties can include a Virtual Agent's identity, attitudes, opinions regarding sensible subjects, and background information such as a name, nationality, or family, among others (Cohen, Cohen, Giangola, & Balogh, 2004). According to literature, non-verbal vocal communication elements such as vocal effort, nasality, and dynamic range also influence the attribution of personality traits such as emotional stability and extroversion in fellow humans and Virtual Agents speaking with a synthetic voice (Sherer, 1978; Garcia & Lopez, 2019; Nass & Gong, 2000). For what is more, features of a synthetic voice such as prosody, volume, and pace also help the user recognize the emotions "felt" by the Agent, even when the facial animations are faulty or non-existent (Hodgins, Jörg, O'Sullivan Park, & Mahler, 2010).

In the context of human-machine interactions, people are capable of identifying the emotions conveyed by the pitch of a robotic beeping, just as they can recognize the "emotions" communicated by music thanks to the track's tempo (Breazeal, 2003), hence the public success of robot "pets" such as R2-D2 or BB-8 in the *Star Wars* series.

Such considerations sustain the possibility to eliminate the semantic content while still maintaining a degree of understandability of the Agent's motives. Gibberish, a speech modality defined as "vocalizations of meaningless strings of speech sounds, and thus have no semantic meaning" (Yilmazyildiz, Verhelst, & Sahli, 2015), has been used in different experimental settings and products, including ENACT (Marocco, 2015). By only varying the Agent's speaking rate, listeners have been shown to recognize Sadness in utterances in 94% of cases, Surprise in 87% of cases, Happiness in 84% of cases, Disgust in 74% of cases, Fear in 73% of cases, and Anger in 66% of cases, the latter two often mistaken for Surprise (Yilmazyildiz, 2015).

Nonetheless, the gibberish contribution to emotional expression for virtual and robotic Agents is truly valuable only as long as non-verbal vocal communication elements, such as emphasis, prosody, and vowel to consonant ratio, are sufficiently naturally presented (Xie, Jia, Meng, Deng, & Wang, 2015). In fact, according to literature, gibberish should be created by substituting real words' vowels with different vowels and real words' consonants with other consonants; therefore, the letter swapping should be weighted on a natural language to sound sufficiently natural (Yilmazyildiz, 2015). Moreover, optimal results of perceived naturalness are obtained when both the gibberish input text and the synthesizer share the same language (Yilmazyildiz, Latacz, Mattheyses, & Verhelst, 2010). Therefore, the final expressive speech strongly depends on the text-to-speech engine quality (Yilmazyildiz et al., 2011). A successful application of such a concept is the *Animal Crossing* series (Nintendo, 2001 - 2020), a series of social simulation games in which, though distorted to appear child-like, gibberish adaptations to different languages can be heard.

For what is more, even though gibberish might substitute different languages that have similar consonant/vowel ratios, it still might not be suitable for different cultures, as different cultures often imply different non-verbal aspects. For instance, cultures differ in vocal feedback modalities, such as "hu-hu" or "hmmm" (Endrass, Schuermann, Kaufmann, Spielberg, Kniesche, & Kathmann, 2010).

1.2 The Present Study

In ENACT, while the Agents' semantic content is conveyed through written text, their prosody, pace, and tone are conveyed by a synthetic voice speaking gibberish, created from an English text source. Non-verbal information is furtherly conveyed through different balloon lines, e.g., a jagged balloon implies shouting, while a dashed one suggests whispering. Studies have indicated that the modality through which a Virtual Agent communicates highly influences users' perception of the Agent itself. For instance, an Agent communicating through voice caused higher feelings of human likeness in users if compared with a text-based Agent, even though the effect was only present when the discussion topic was utilitarian (e.g., "When did the Titanic sink?") and not hedonic (e.g., "Do you have dreams?"; Cho, Molina, & Wang, 2019). Nevertheless, much seems to depend on the quality of the synthetic voice, as Agents with synthetic voices are alternatively evaluated worse than text-based communication (Gong & Nass, 2007) or as credible as an Agent with a human voice (Craig & Schroeder, 2017). If a synthetic gibberish voice is enough to sustain Virtual Agents' acceptability is still an open question. Two separate versions of ENACT have been, therefore, created: in the first version, the Agents spoke the original gibberish accompanied by written text, while in the second version, all audio features have been deactivated, and the Virtual Agent will therefore communicate through text only.

1.3 Objectives and Hypotheses

Along the same lines of the first experimental study, the first objective is to explore whether the introduction of a gibberish form of communication in a Virtual Agent will be enough to sustain better acceptability and quality of the interaction if compared to a text-only approach, therefore constituting a possible easy win on the short development run for smaller research labs, not requiring expensive interventions in terms of money and time. The use of non-semantic speech modalities might be beneficial to cut production times and expenses in Simulated Human Interactions available in different languages, which should be otherwise voiced multiple times or involve retrieving data for speech synthesis from different language databases.

A second objective is to investigate if video game players who are used to interacting with state-of-the-Art Virtual Agents have harsher reactions to Simulated Human Interactions for educational purposes if compared with non-habitual video game players.

Specifically, it is hypothesized that:

Ia) Participants will evaluate the quality of the product more positively if the Virtual Agent communicates through Gibberish as opposed to a Text-Only form of communication;

Ib) Participants will evaluate the Acceptability of the Virtual Agent's Appearance more positively if the Virtual Agent communicates through Gibberish instead of a Text-Only form of communication.

As a Virtual Agent speaking through a modern text-to-speech voice engine was rated as credible as an Agent with a human voice, outperforming an Agent speaking through an older text-to-speech voice engine (Craig & Schroeder, 2017).

IIa) Habitual story-based video game players will evaluate the quality of the product more negatively than non-habitual video game players;

IIb) Habitual story-based video game players will evaluate the Acceptability of the Virtual Agent more negatively than non-habitual video game players;

As habitual players are highly exposed to story-driven video games in which actors provide voicing for Virtual Agents (De Castell & Jenson, 2003).

III) Interaction effects between Game Habits and the Experimental Conditions are also expected. Habitual story-based video game players will evaluate the Virtual Agent's Acceptability more negatively if Gibberish is added to its written communication than if it communicates through text only.

2 Methods

2.1 Study Design

A 2x2 between-subjects study design (see Fig. 21) has been carried out entirely online due to movement limitations following the Coronavirus CoViD-19 outbreak.

The following independent variables have been considered:

1. Game Habits: Participants declaring a minimum of 5 hours of story-based video game-play per week were included in the habitual video game players group (VGP), participants declaring fewer hours of story-based gameplay per week were included in the non-habitual video game players (NVGP). See page 82, paragraph 2.1, of the present dissertation for more details about the selected threshold. The target game genre was identified in story-based games, as they tend to include human Virtual Agents and devote much time to dialogues;

2. Text-Only vs. Gibberish: Participants were randomly assigned to one of two different versions of ENACT: in the original version, the Virtual Agent confronted by the user speaks Gibberish (Gibberish condition) during the interaction; in the modified version, all auditory elements have been deactivated (Text-Only condition).

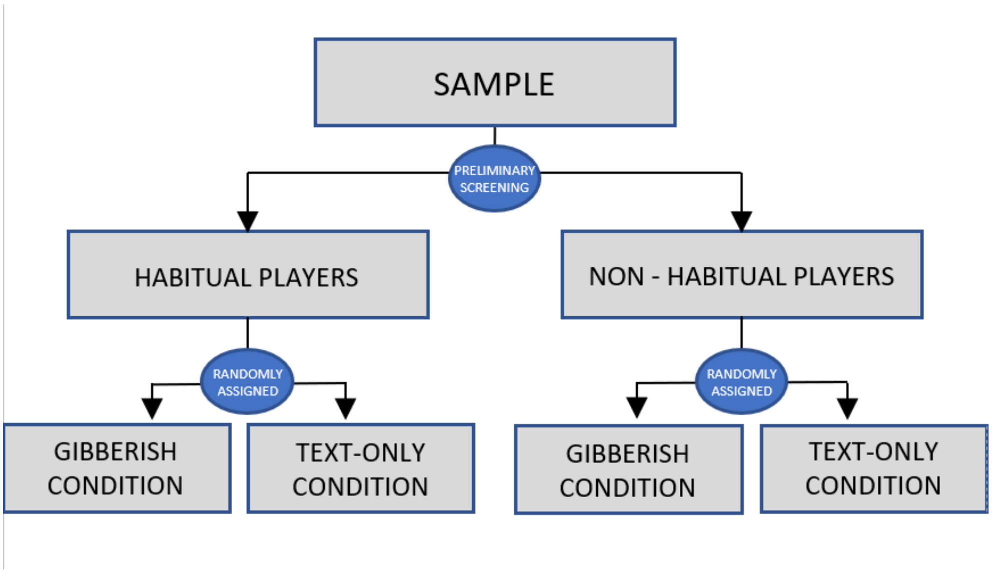


Figure 21. Graphical representation of the Study Design

2.2 Participants

A sample of 60 young male participants (mean age = 27.7; SD = ±3.88) has been recruited. Restriction to male participants was introduced because of the relative difficulty to find female habitual video game players, as the difference between weekly gameplay hours of habitual and non - habitual female players resulted not to be sufficiently large.

Participants were unaware of the study's real objectives, as it was presented as an experimental study related to conflict resolution styles. No economic rewards were provided during the research; however, feedback regarding the conflict resolution styles used while negotiating with the Virtual Agent, as provided by ENACT's second debriefing page, was provided at the end of the session. To be included in the study, individuals had to meet the following criteria:

(1) age between 18 and 35 years old;

(2) no significant visual and auditory impairment (all with normal or corrected-to-normal visual acuity)

Before partaking, all participants were provided with written information about the study and were required to give written consent to be included. The study received ethical approval from the Ethical Committee of the University of Milano-Bicocca.

2.3 Materials

Analogously to the first experimental study, the same scenario of ENACT has been used as material for this second study. Further information regarding ENACT can be found on page 84, paragraph 2.3, of the present dissertation. Following social distancing rules due to the COVID-19 pandemic, participants used their personal computer to interact with the Agent (minimum requirements: IntelCore i5 with 8GB of RAM, 64-bit version of Windows 8).

2.4 Measures

- *Measures of Video Game Habits*

Participants were assigned to the habitual or non-habitual story-based video game players groups according to their answer to a preliminary question related to their mean weekly gameplay hours in the previous four weeks. Two further questions about the language of the fruition of said video games (related to, respectively, dialogues and subtitles) were also included for habitual players to measure their familiarity with the sound of different languages.

- *Demographics*

Demographic information was considered through a dedicated questionnaire, including questions about the participant's gender and age.

- *Measures of Acceptability of the Virtual Agent's Appearance*

Eight items with bipolar anchors on a 7-point scale were used to evaluate the Virtual Agent's appearance, adapted from the Humanness and Eeriness Indexes (Ho & MacDorman, 2010). The anchors were: Artificial vs. Biological, Created by Man vs. Similar to Man, Inanimate vs. Alive, Unreal vs. Realistic, and Unlikeable vs. Likeable from the Humanness Index; Eerie vs. Reassuring, Predictable vs. Unpredictable, and Uncomfortable vs. Comfortable from the Eeriness Index. It has to be noted that a higher score in the Eeriness Index indicates higher perceived reassurance while interacting with the Virtual Agent.

- *Measures of the Virtual Agent's Voice Quality*

Three items on a 7-point Likert scale related to Voice Likeability, Voice Consistency (with the Agent's appearance), and Voice Expressiveness were included (Cabral, Cowan, Zibrek, & McDonnell, 2017).

A single 10-point Likert scale item related to the degree of resemblance of the Agent's language with a human language as opposed to a random combination of sounds was also included, as well as an open question asking which natural language the participants were reminded of while listening to the Agent's (as in Yilmazyildiz, 2014; Allison, 2018).

- *Measures of Perceived Quality of the Product*

The Game Experience Questionnaire has been used (IJsselsteijn, de Kort, & Poels, 2013): a 33-item questionnaire on a 5-point Likert scale, measuring five subscales. These are furtherly divided into Positive subscales, including Competence, Immersion, Flow, and Positive Affect, and Negative subscales, including Tension, Challenge, and Negative Affect. Further information regarding the subscales can be found on page 86, paragraph 2.4, of the present dissertation.

Likelihood to recommend: the Net Promoter Score, a single question asking the user to state, on a scale from 1 to 10, how likely they are to recommend the training to a friend or colleague interested in improving their conflict resolution style (Reichheld, 2003).

2.5 Procedures

During the first contact for recruitment, participants were preliminary asked to evaluate their mean weekly gameplay hours to be assigned to the habitual or non-habitual video game players. Once randomly assigned to the Text-Only or Gibberish group, participants were contacted through Google Meet while being individually seated at a desk in a quiet room in their house. Participants used their personal computers (minimum requirements: IntelCore i5 with 8GB of RAM, 64-bit version of Windows 8) while wearing their headphones, and the researcher observed their actions through the screen sharing feature, following social distancing rules.

After answering the demographic questionnaire, participants were instructed about ENACT's structure and controls and the interaction scenario. During the Interaction, the researcher muted their microphone and suspended communications with the participant for the Interaction duration. After ending the interaction, participants completed the questionnaires related to the Humanness and Eeriness Indexes. Participants in the Gibberish condition then answered the question related to the Virtual Agent's Voice. All participants were asked to complete the questionnaires related to their Player Experience and to their likelihood to recommend the training product to a friend or colleague interested in training their conflict resolution strategies. Finally, feedback regarding the conflict resolution styles used while negotiating with Anna, as provided by ENACT's second debriefing page, was provided at the end of the session. Feedback regarding the conflict resolution styles used while negotiating with the Virtual Agent was provided at the end of the session. A graphical representation of the experimental procedure can be found on the following page (Fig.22)

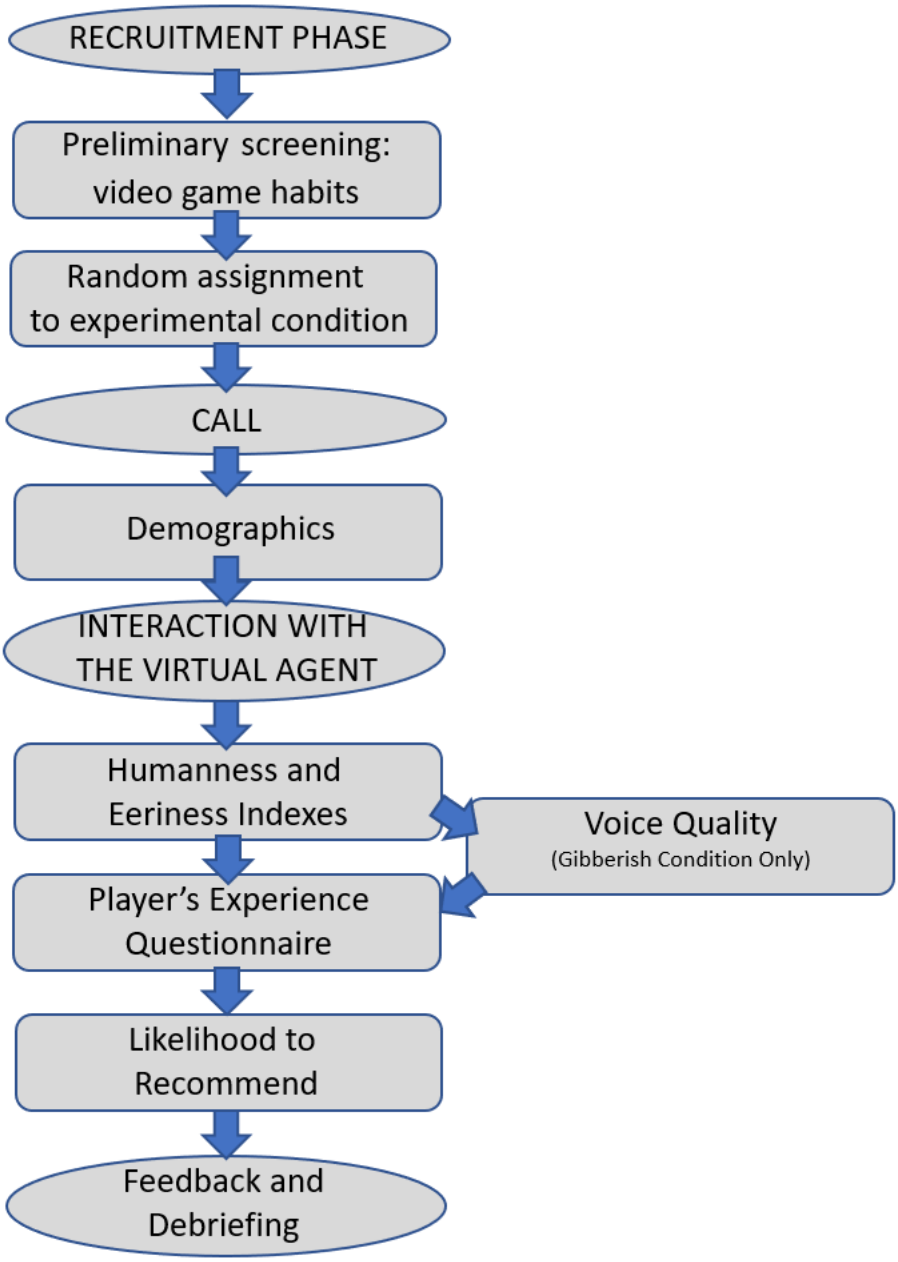


Figure 22. Graphical representation of the Experimental Procedure.

3 Results

Participants have been divided into a habitual video game players group (n = 30 mean age = 27.03; SD = ±4.44) and a non-habitual video game players group (n = 30; mean age = 28.37; SD = ±3.16), according to their declared weekly gameplay hours. The mean of weekly gameplay hours for habitual video game players was 12.46 (SD = ±6), while for non-habitual video game players was 1.83 (SD = ±3.88). Statistical analyses were conducted to test the hypothesized main effect between game habits and the participants' player experience, likelihood to recommend, and perceived acceptability of the Virtual Agent, and between the two experimental conditions and the participants' player experience, likelihood to recommend, and perceived acceptability of the Virtual Agent. Interaction effects between the two independent variables were also explored. A one-way between subjects' ANOVA was conducted in SPSS v. 26 at the $p<.05$ level of significance.

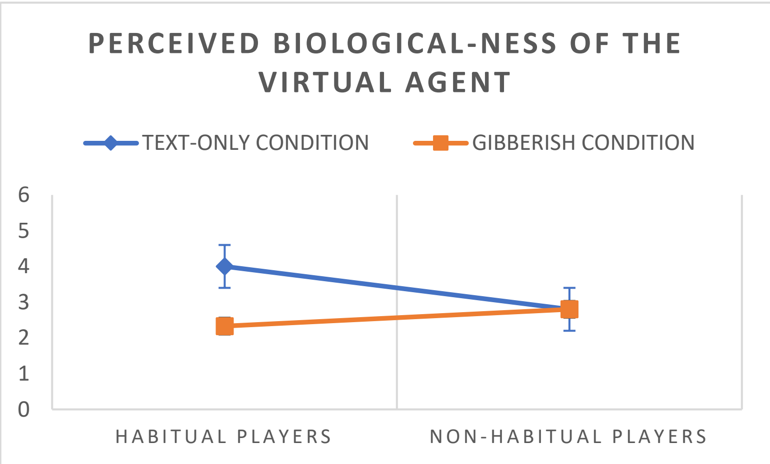
3.1 Perceived Acceptability of the Virtual Agent

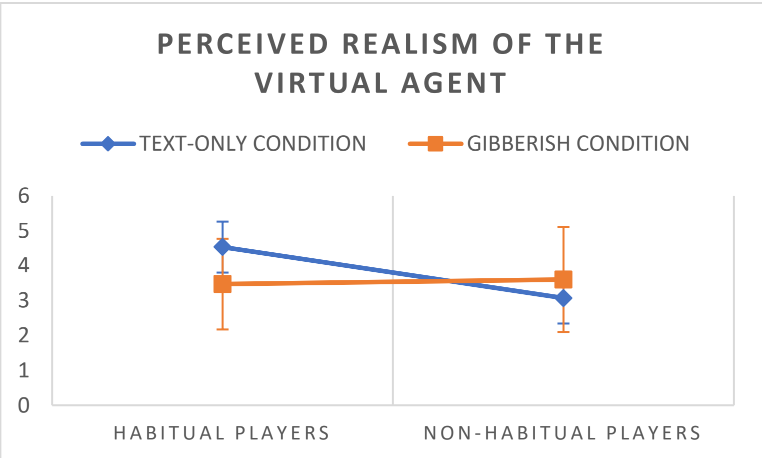
Statistical analyses revealed a statistically significant effect of game habits on the perceived Biological-ness of the Virtual Agent. Moreover, analyses showed a statistically significant interaction effect of game habits and experimental conditions on perceived Realism and the Virtual Agent's Eeriness. Results are summarized in tables 6a and 6b, page 115 of the present dissertation. In particular:

- **Evaluation of the Virtual Agent's Humanness Index:** Analyses did not reveal statistically significant differences in terms of game habits [$F(1, 58) = 1.27$, $p = .26$] or or experimental conditions [$F(1, 58) = 1.58$, $p = .21$]. Interaction effects between the two variables were also explored, but analyses did not return statistically significant effects [$F(1, 58) = 2.74$, $p = .10$].

  However, it has to be noted that analyses on the Biological Item of the Humanness Index revealed a statistically significant difference in terms of game habits [$F(1, 58) = 5.94$, $p = .018$] and a significant interaction effect [$F(1, 58) = 5.94$, $p = .018$]. Scores were significantly different in habitual players (mean = 3.17; SD = ±1.59) if compared to the Non-habitual players (mean = 2.80; SD = ± 1.24), although the effect size for this analysis (*d* = 0.30) was found to meet Cohen's (1988) convention for a small effect (*d* = 0.20), and in habitual players in the Gibberish Condition (mean = 4.00; SD = ± 1.73) if compared to habitual players in the Text-Only Condition (mean = 2.33; SD = ± .90), according to Tukey's HSD (See Graph 3a). The effect size for this analysis (*d* = 1.21) was found to exceed Cohen's (1988) convention for a large effect (*d* = 0.80).

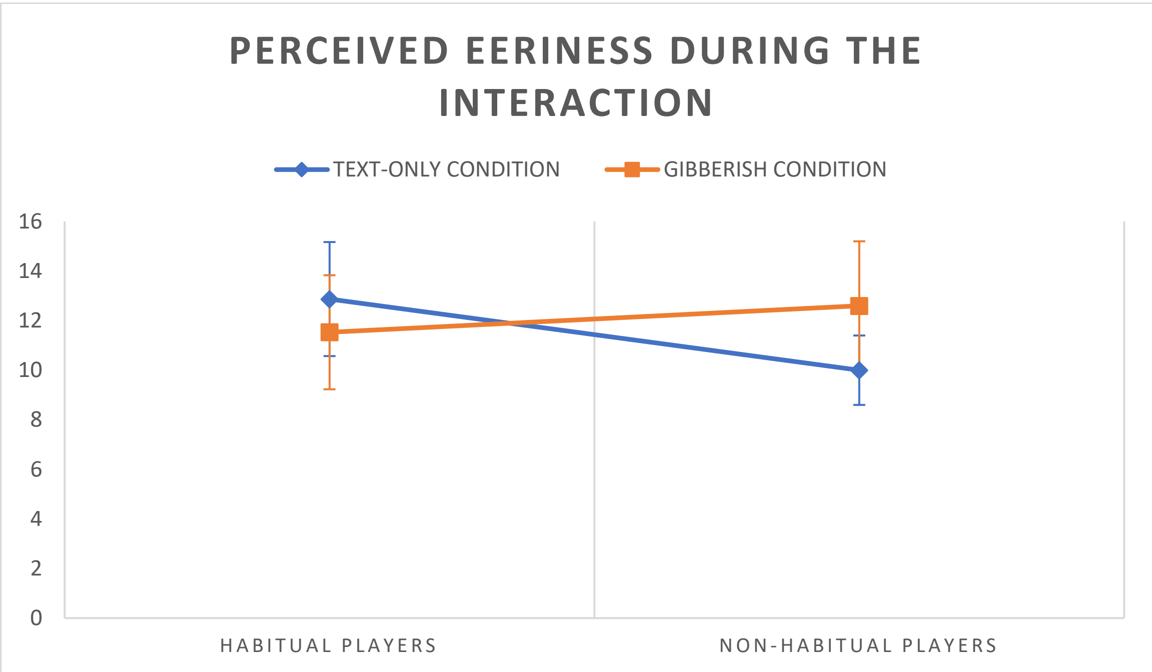
Moreover, analyses on the Realism Item of the Humanness Index revealed a statistically significant interaction effect [F(1, 58) = 4.97, p = .030]. Scores were significantly different in habitual players in the Text-Only Condition (mean = 4.53; SD = ±1.40) if compared to the Non-habitual players in the same Condition (mean = 3.07; SD = ± 1.03), according to Tukey's HSD (See Graph 3b). The effect size for this analysis (*d* = 1.18) was found to exceed Cohen's (1988) convention for a large effect (*d* = 0.80).





Graph 3a, b. Single-item results: Biological-ness (above) and Realism (below) as perceived by Habitual and Non-Habitual Video Game Players in the two experimental conditions.

- **Evaluation of the Virtual Agent's Eeriness:** Analyses did not reveal statistically significant differences in terms of game habits [$F(1, 58) = 2.40$, $p = .13$] or experimental conditions [$F(1, 58) = 1.19$, $p = .28$]. However, an interaction effect between game habits and experimental condition was found to be statistically significant [$F(1, 58) = 11.47$, $p = .001$]. Scores were significantly different in non-habitual players in the Text-Only Condition (mean = 10.00; SD = ±1.41) if compared to the Gibberish Condition (mean = 12.60; SD = ± 2.66). The effect size for this analysis (d = 1.18) was found to exceed Cohen's (1988) convention for a large effect (d = 0.80). Moreover, scores were significantly different in habitual video game players in the Text-Only Condition (mean = 12.87; SD = ±2.38) if compared to Non-habitual video game players in the same Condition (mean = 10.00; SD = ±1.41) according to Tukey's HSD (See Graph 4). The effect size for this analysis (d = 1.22) was found to exceed Cohen's (1988) convention for a large effect (d = 0.80).



Graph 4. Eeriness as perceived by Habitual and Non-Habitual Video Game Players in the two experimental conditions. Note that higher scores indicate higher perceived reassurance while interacting with the Virtual Agent.

Measure	Habitual Players		Non-Habitual Players		$F(1, 58)$	η_p^2	Power
	M	*SD*	*M*	*SD*			
Humanness	17.47	5.98	15.77	5.93	1.27	.022	.198
Biological	3.27	1.59	2.80	1.24	1.15	.020	.184
Similar to Man	3.40	1.43	3.27	1.84	.09	.002	.061
Alive	3.63	1.49	3.23	1.67	.93	.016	.157
Realistic	4.00	1.46	3.33	1.39	3.45**	.058	.447
Likeable	3.27	1.53	3.33	1.32	.03	.001	.054
Eeriness	12.20	2.41	11.30	2.48	2.40	.041	.332
Reassuring	3.73	1.17	3.7	.91	4.97*	.082	.591
Predictable	3.60	1.69	3.93	1.59	.61	.011	.120
Comfortable	4.87	1.47	4.20	1.66	2.82	.048	.379

*Table 6a. Means, Standard Deviations, and One-Way Analyses of the Humanness and Eeriness Index of the Virtual Agent in terms of Game Habits. * indicates a statistically significant difference. ** indicates a quasi-statistically significant difference.*

Measure	Gibberish Condition		Text-Only Condition		$F(1, 58)$	η_p^2	Power
	M	*SD*	*M*	*SD*			
Humanness	15.67	5.72	17.57	6.16	1.58	.028	.236
Biological	2.57	1.13	3.40	1.59	5.94*	.096	.669
Similar to Man	3.13	1.52	3.53	1.74	.89	.016	.153
Alive	3.27	1.62	3.60	1.57	.64	.011	.124
Realistic	3.53	1.50	3.80	1.42	.55	.010	.113
Likeable	3.37	1.45	3.23	1.40	.13	.002	.065
Eeriness	12.07	2.52	11.43	2.42	1.19	.021	.189
Reassuring	3.70	.98	3.20	1.13	3.87*	.065	.489
Predictable	3.63	1.69	3.90	1.60	.39	.007	.095
Comfortable	4.73	1.59	4.33	1.60	1.01	.018	.168

*Table 6b. Means, Standard Deviations, and One-Way Analyses of the Humanness and Eeriness Index of the Virtual Agent in terms of Experimental Condition. * indicates a statistically significant difference.*

Participants found the quality of the Virtual Agent's Voice to be middling, in terms of Voice Likeability (mean = 3.33/7; SD = ±1.42) Voice Consistency with the Agent's appearance (mean = 4.2/7; SD = ±1.83), Voice Expressiveness (mean = 3.6/7; SD = ±1.64), and degree of resemblance of the Agent's language with a human language (mean = 5.76/10; SD = ±2.17). Results are summarized in Table 8, following page.

Regardless of gaming habits, most participants in both groups reported that the gibberish reminded them of English (11 players; 8 non-players). Differences were found in the languages stated by the other participants. Most players (11) reported habitually set their videogame language to English, while the others (4) reported playing in Italian. Interestingly, while several habitual players who reported playing games dubbed in English stated the Gibberish reminded them of non-English languages, all habitual players who reported mostly playing Italian-dubbed video games identified the Gibberish as resembling English (see Table 7).

Participants in Gibberish Condition	Language of habitual fruition of videogames	Identified language when listening to Gibberish			Total
		English	Other Languages		
Habitual Video Game Players	English	7	French	1	15
			Unintelligible	1	
			Esperanto	1	
			Japanese	1	
	Italian	4			
Non-habitual Video Game Players		8	Russian	2	15
			Italian	1	
			Hindi	1	
			Chinese	1	
			Dutch	1	
			Danish	1	

Table 7. Results from an open question asking which natural language the participants were reminded of while listening to the Agent's (as in Yilmazyildiz, 2014; Allison, 2018).

Measure	Habitual Players		Non-Habitual Players		*F*(1, 58)	p
	M	*SD*	*M*	*SD*		
Voice Quality						
Likeability	3.60	1.24	3.06	1.57	1.05	.31
Consistency	4.53	1.50	3.93	2.12	.79	.37
Expressiveness	3.80	1.32	3.53	1.95	.19	.66
Language Familiarity	6.33	1.79	5.2	2.43	2.12	.15

Table 8. Means, Standard Deviations, and One-Way Analyses of the Agent's Voice Quality and Language Familiarity in terms of Game Habits.

3.2 Perceived Quality of the Product

Statistical analyses revealed a statistically significant effect of experimental conditions on perceived Flow while interacting with the Virtual Agent and an effect at the margins of statistical significance of game habits on Likelihood to Recommend. Results are summarized in tables 9a and 9b, page 119 of the present dissertation. In particular,

Concerning the Game Experience subscales,

- **Competence:** Analyses did not reveal statistically significant differences in terms of game habits [F(1, 58) = 2.36, p = .12] or experimental conditions [F(1, 48) = .000, p = 1]. Interaction effects between the two variables were also explored, but analyses only reveal a certain trend towards significance [F(1, 58) = 3.05, p = .086];

  Immersion: Analyses did not reveal statistically significant differences in terms of game habits [F(1, 58) = .921, p = .34] or experimental conditions [F(1, 58) = .374, p = .54]. Interaction effects between the two variables were also explored, but analyses did not return statistically significant effects [F(1, 58) = 1.17, p = .28];

- **Flow:** Analyses did not reveal statistically significant differences in terms of game habits [F(1, 58) = 1.09, p = .30]. However, a statistically significant difference was found in terms of experimental conditions [F(1, 58) = 4.81, p = .032]. In particular, the mean score for the Gibberish Condition (mean = 3.40; SD = ±0.74) was significantly different than the score for the Text-Only Condition (mean = 2.97; SD = ± 0.77). The effect size for this analysis (*d* = 0.88) was found to exceed Cohen's (1988) convention for a medium effect (*d* = 0.57). Interaction effects between the two variables were also explored, but analyses did not return statistically significant effects [F(1, 58) = .192, p = .66];

- **Challenge:** Analyses did not reveal statistically significant differences in terms of game habits [F(1, 58) = .034, p = .85] or experimental conditions [F(1, 58) = .49, p = .49]. Interaction effects between the two variables were also explored and analyses returned a statistically significant effect [F(1, 58) = 4.10, p = .048]. However, post-hoc analyses (Tukey's HSD) did not report statistically significant differences between groups;

- **Positive Affect:** Analyses did not reveal statistically significant differences in terms of game habits [F(1, 58) = .456, p = .50] or experimental conditions [F(1, 58) = .456, p = .50]. Interaction effects between the two variables were also explored, but analyses did not return statistically significant effects [F(1, 58) = 2.80, p = .10];

- **Negative Affect:** Analyses did not reveal statistically significant differences in terms of game habits [F(1, 58) = .023, p = .88] or experimental conditions [F(1, 58) = .009, p = .93]. Interaction effects between the two variables were also explored, but analyses did not return statistically significant effects [F(1, 58) = 2.28, p = .136];

- **Tension:** Analyses did not reveal statistically significant differences in terms of game habits [F(1, 58) = .323, p = .57] or experimental conditions [F(1, 58) = .332, p = .56.] Interaction effects between the two variables were also explored, but analyses did not return statistically significant effects [F(1, 58) = .083, p = .77];

Concerning the Likeliness to Recommend, analyses reveal a difference at the margins statistical significance in terms of game habits [F(1, 58) = 3.2, p = .07], but did not reveal significant differences in terms of experimental conditions [F(1, 48) = .800, p = .37]. Interaction effects between the two variables were also explored, but analyses did not return statistically significant effects [F(1, 58) = 1.15, p = .28].

Measure	Habitual Players		Non-Habitual Players		$F(1, 58)$	η_p^2	Power
	M	*SD*	*M*	*SD*			
Likelihood to Recommend	8.23	1.45	7.57	1.43	3.20	.054	.420
Players' Experience							
Competence	3.44	.57	3.15	.88	2.36	.041	.328
Immersion	3.12	.78	2.92	.82	.92	.016	.157
Flow	3.29	.82	3.08	.74	1.09	.019	.177
Challenge	2.21	.78	2.17	.64	.34	.001	.054
Positive Affect	3.25	1.00	3.10	.76	.45	.008	.102
Negative Affect	1.54	.56	1.65	.57	.63	.011	.122
Tension							

Table 9a. Means, Standard Deviations, and One-Way Analyses of the Likelihood to recommend and the players' experience components in terms of Game Habits.

Measure	Gibberish Condition		Text-Only Condition		$F(1, 58)$	η_p^2	Power
	M	*SD*	*M*	*SD*			
Likelihood to Recommend	8.07	1.38	7.73	1.55	.37	.014	.142
Players' Experience							
Competence	3.30	.71	3.30	.80	.00	.000	.050
Immersion	3.08	.76	2.96	.85	.37	.007	.092
Flow	3.41	.74	2.97	.77	4.81*	0.79	.578
Challenge	2.25	.72	2.13	.70	.49	.009	.106
Positive Affect	3.10	.98	3.25	.78	.45	.008	.102
Negative Affect	1.60	.55	1.60	.59	.00	.000	.050
Tension							

*Table 9b. Means, Standard Deviations, and One-Way Analyses of the Likelihood to recommend and the players' experience components in terms of Experimental Conditions. * indicates a statistically significant difference.*

4. Discussion

Results support that the introduction of Gibberish does affect the Virtual Agent's perceived acceptability. Regardless of game habits, participants considered the Virtual Agent to be more biological-looking in the Text-Only Condition. They also found the Agent more reassuring when speaking Gibberish and reported higher Flow levels while interacting in the same condition. Following previous literature, much seems to depend on the quality of the synthetic voice: Agents with particularly mechanical synthetic voices are evaluated worse than text-based communication (Gong & Nass, 2007), while a modern text-to-speech voice engine was rated as credible as an Agent with a human voice, outperforming an Agent speaking through an older text-to-speech voice engine (Craig & Schroeder, 2017). It can be inferred that the quality of the Gibberish voice was deemed inadequate by participants, as suggested by the middling evaluation of the Agent's voice quality. Moreover, these results seem to be in line with the Uncanny Valley's classical theorization, as increasing human-likeness is associated with more comfort only until it exceeds a certain threshold. At that point, there is a sharp drop in the perceived familiarity of the Agent before becoming better again (Kätsyri, Förger, Mäkäräinen, & Takala, 2015).

Interestingly, habitual and non-habitual players showed opposite tendencies regarding perceived reassurance and realism of the Virtual Agent. In particular, habitual video game players reported the Agent to be more reassuring and biological-looking when the Virtual Agent would not audibly speak, while non-habitual players reported higher scores for the Virtual Agent speaking Gibberish. These results seem to suggest, once again, that players and non-players might experience the "too real for comfort" zone differently. While non-habitual players might expect a Virtual Agent to speak, whichever the quality of its voice, habitual players' adverse reaction to the Gibberish might be due to its perceived insufficient quality if compared to what is provided by the entertainment industry, increasingly relying on actors' performances to convey visually credible characters (Domsch, 2017). Interestingly, a recent study suggested that users prefer a more mechanical feminine voice than a good quality masculine voice. The reason behind this unexpected answer was thought to be the habit of having a female voice as a Virtual Assistant (Perez & Saffon, 2018), demonstrating the mediation role of exposure habits in uncanniness perception. Therefore, Hypothesis II, stating that Virtual Agents in the Text-Only condition would be evaluated worse in terms of acceptability, seems once again to need rephrasing, and taking into consideration the target audience's previous expertise in gaming seems to be particularly important to design Virtual Agents that avoid rousing feelings of uncanniness. Along the same lines of the previous study, present results seem to deny differences in the perception of the product's quality due to its users' game habits.

5. Conclusions

The following paragraphs will trace conclusions gathered from the present experimental study.

The Quality of the Virtual Agent's Gibberish Affects Their Perceived Realism and Reassurance: Confirming previous literature, Agents with mechanical-sounding synthetic voices are evaluated worse than text-based communication (see Gong & Nass, 2007). Consequently, the mere addition of a Gibberish feature cannot be considered as a possible easy win on the short development run for smaller research labs, which, instead, should prefer a Text-Only approach when modern text-to-speech voice engines are not available.

Gaming Habits May Posit Higher Standards for Virtual Agents' Voices: Game habits might posit higher standards to players when evaluating a Virtual Agent's voice acceptability. Due to their previous experiences, players and non-players might experience the Audio Uncanny Valley differently when it comes to tolerance to a synthetic voice's quality in terms of non-semantic elements. This suggests the perception of an Agent's human-likeness is a complex and dimensional matter, adding previous game experience as an intervening factor in the perception of uncanniness, similarly to what has been suggested for Virtual Assistants' voices (Perez & Saffon, 2018).

6. Limitations of the Present Study

Although carefully conducted, the present study is not without limitations.

The unexpected situation due to the COVID-19 pandemic has led to a reorganization of the present study's procedure, which has been adapted to an online setting. Several matters should be considered, such as the participants' possibly different Internet speeds while playing ENACT, leading to the Virtual Agent's smoother or bumpier movements. It must also be noted that the development work necessary to create other versions of ENACT has been abruptly halted for the same reason. Consequently, the Text-Only condition was obtained by asking participants to deactivate audio features of their Internet browser. As for the Gibberish condition, ENACT's background music could not be removed without intervention from the development team. Therefore, the music's presence during the Gibberish condition interaction might have acted as a confounding variable. Future studies should be directed towards confirming the present results after removing the background music.

Regarding selected measures, the same limitations as the first experimental study apply to the second one. Future studies should include qualitative measures providing more profound and meaningful insights from participants, as well as psychophysiological measures (e.g., heart and breathing rate, skin conductivity), provided they are not invasive, excessively affecting the immersion in the Simulated Human Interaction (see page 99, paragraph 6 of the present Dissertation for more details).

Finally, the study was conducted, and the analyses were based on data from a small self-selected sample of participants, mainly recruited online, from a population of Ph.D. Candidates. Further investigation of a larger sample is needed to confirm the preliminary results provided by the present study.

Chapter 7
Conclusions and Future Directions

7.1 Conclusions

During the last decades, Virtual Agents have not only been increasingly used in a variety of areas but also in a variety of roles. For instance, they can act as mentors and deliverers of notions in the context of Simulated Human Interactions to enhance effective communication strategies. Nevertheless, since their visual appearance and behavior are intrinsically based on technological resources and design aspects (Straßmann & Krämer, 2017), replicating the level of human-likeness required to "convince users' perceptual systems that a virtual human is the real thing" (Ruhland et al., 2015) still remains a challenge for Entertainment and Education alike. Mori's theorization is known for having systematized the Uncanny Valley theory basics, here defined as the abrupt decline in humans' perceived familiarity of a stimulus when it nears perfection (Mori, 1970; Mori, MacDorman, & Kageki, 2012). In particular, the Uncanny Valley effect refers to the observer's unpleasant impression of a virtual being with an almost, but not entirely, realistic human form (Seyama & Nagayama, 2007). Previous literature has described several intervening factors in the perception of uncanniness, including the Agent's static and dynamic features, but also individual differences in the degree of predisposition to anthropomorphize an Agent (e.g., Epley, Waytz, & Cacioppo, 2007; Kätsyri, Förger, Mäkäräinen, & Takala, 2015). During the last decades, video games have been representing an entertainment source for a growing number of people, and this dissertation's objective has been to confirm whether game habits might be considered among intervening factors. The video game industry has been driving technological innovation allowing for high-fidelity face and voice synthesis of Virtual Agents in Entertainment products. Such technologies are often not available to smaller research laboratories relying on limited resources. Therefore, the present dissertation has also explored the possibility of identifying "easy wins" on the short development run, essential elements that do not require expensive interventions in terms of money and time but can increase the perception of the Virtual Agent's quality.

However, there is a vast amount of knowledge about Virtual Agents from various sources across a range of disciplines (e.g., anatomy, psychology, neuroscience, and social sciences) that might make up for the technological and economic disadvantages of small research labs. Therefore, the present dissertation has explored the possibility of identifying "easy wins" on the short development run, essential elements that do not require expensive interventions in terms of money and time but can increase the perception of the Virtual Agent's quality.

ENACT (Marocco, Pacella, Dell'Aquila, & Di Ferdinando, 2015), an online Simulated Human Interaction for the training of Negotiation strategies, has been used as the main object of this dissertation. In ENACT, trainees take five conversation turns with a Virtual Agent that communicates through a combination of four different facial expressions, 24 gestures, and ten different body postures and gaze directions. The present dissertation includes two experimental studies, exploring the effectiveness of low budget implementations of Virtual Agent's features, i.e., random eye blinks and spoken gibberish accompanying written communication. In the specific context of ENACT, the mere introduction of eye blinks in random moments of the interaction with a Virtual Agent seem to moderately affect the user's perception of the Agents' realism. Nonetheless, the simple frequency of positive feedback in Virtual Agents might not be as crucial as its contingency in building a rapport between the user and the Virtual Agent (e.g., Gratch, Wang, Gerten, Fast, & Duffy, 2007). Moreover, in cases when modern text-to-speech voice engines are not available, it seems advisable to rely on a text-only form of communication for Virtual Agents instead of a gibberish-based communication.

Though no one expects research laboratories to be able to compete with the monetary and time resources of the entertainment video game industry, being aware of such implicit and explicit rules of human-likeness is likely to allow educational products to be appropriately quality controlled in an era in which users are becoming more and more demanding. When perfection is not achievable, theoretical and practical knowledge can allow designers to know when and how to hide imperfections in animations, textures, and design effectively. To quote a successful example, Stephen Spielberg's famous *Jurassic Park* scene in which a Tyrannosaurus Rex attacks the main characters was filmed in 1993, yet it still stands the test of time, nearly three decades later. Due to the technological limitations of their time, Academy Award winner visual effects artists from *Industrial Light & Magic* were, in fact, fully aware that the dinosaur's skin would have looked excessively smooth and shiny to be considered realistic. Therefore, they decided to set the scene at nighttime, during a storm: darkness helped to hide the smoothness, and rain and lightning justified the excessive shininess of the animal's skin (Fig. 23a, following page; Pueringer, 2019c). On the contrary, designers of the infamous video game

Mass Effect: Andromeda deliberately decided to include a scene in which the main character pours herself a glass of wine, though unnecessary for the plot, despite obvious flaws in the animation of liquid substances (Fig. 23b).

Fig. 23a. The T-Rex scene in Jurassic Park (Spielberg, 1993), created by the VFX company Industrial Light & Magic in which darkness helped to hide the smoothness, and rain and lightning justified the excessive shininess of the animal's skin.

Fig. 23b. The faulty wine-pouring scene in Mass Effect: Andromeda (BioWare, 2017), in which the wine does not display appropriate reflective properties and appears to be disconnected from the bottle.

It has to be furtherly noted that the contemporary Entertainment Video Game Industry has encouraged the acceleration of technological progress related to the acquisition of visual and auditory data from real-life situations, using sophisticated systems and real actors as reference (Batt, 2015). The present dissertation suggests that video game habits might have a mediating role in the perception of Virtual Agents' qualities. Nevertheless, game habits might not simply posit higher standards to players.

When it comes to evaluating a Virtual Agent's voice acceptability, a synthetic gibberish seems to be less tolerated by habitual players as opposed to-non players, the latter expecting a human-like being to be able to speak, regardless of the quality of the voice. Nevertheless, when it comes to visual aspects of a Virtual Agent's face, the blinking movement might fall into the Uncanny Valley only for non-habitual players, while habitual players appreciated it.

Therefore, players and non-players might experience the "too real for comfort" zone differently, consequently assigning different expected social standards and normative expectations to Virtual Agents they interact with. Such results confirm that the perception of an Agent's human-likeness is a complex and dimensional matter, therefore including previous gaming literacy into the factors intervening in the perception of uncanniness.

Considering the target audience's previous gaming expertise seems to be particularly important to design Virtual Agents that avoid rousing feelings of uncanniness. In conclusion, researchers urgently need to communicate with industry professionals, as rapid technological advancements pose different standards to an increasingly larger population of habitual players, and future studies should be directed towards further investigating this complex relationship.

7.2 Future Studies

Building on the present results, future studies might be conducted after modifying the blinks' randomization to create a more accurate illusion of responsiveness. Although ENACT's multiple-choice interaction modality might not allow for a real synchronization of blinks between the Agent and the user, it would be worth exploring the possibility of increasing blinks during crucial moments of the Virtual Agent's conversational turns. Literature has previously stated that blink frequency tends to increase around the end of speaking units (Sacks, Schegloff, Jefferson, 1974; Nakano & Kitazawa, 2010) and when fatigued (Johns, Tucker, Chapman, Crowley, Michael, 2007) or lying (Burgoon, Buller, Woodall, 1996).

Future studies may also explore the effects of the coupling between blinking behavior and gaze behaviors connected to interest expression, visual attention, and social cognition on the Agent’s perceived human-likeness. Gaze, among other non-verbal components of communication, is an essential aspect of human-to-human interactions as an indicator of interest and a source of information for the interlocutor (Kevin, Pai, & Kunze, 2018). First of all, in human-to-human communication, gazing away from or towards the interlocutor can function as a signal to hand over the turn or avoid the turn to be taken over. (Nijholt, 2003). As for emotional content, people who are more likely to engage in eye contact are perceived as more likable and attractive than people who tend to avert their gaze to avoid eye contact (Mason, Tatkow, & Macrae, 2005).

Moreover, emotion interpretation seems to vary as a function of gaze direction in human-to-human conversations, especially in situations in which additional interpretation by the observer is required (e.g., an angry face with averted gaze or a fearful face with a direct gaze; Adams & Kleck, 2003). People are, therefore, capable of quickly notice if a person is looking at them. This has been demonstrated by the fact that, during visual search tasks, participants took less time to locate people whose gaze was directed at them rather than people whose eyes were averted (Conty, Tijus, Hugueville, Coelho, & George, 2006; Senju, Hasegawa, & Tojo, 2005; Frischen, Bayliss, & Tipper, 2007).

Previous literature has already suggested that Virtual Agents displaying oriented gaze in connection to current events are perceived as more credible by interacting users (Gatica-Perez, 2009; Peters, Pelachaud, Bevacqua, Mancini, Poggi, 2005). Implementing appropriate gaze behavior in connection with turn-taking heavily affects users’ satisfaction with the interaction. For instance, users who interacted with a gaze-appropriate Agent gave a better evaluation of the Agent’s personality and the naturality of its head movements compared to an Agent that showed minimal gaze change. Moreover, they had faster task-completion times and better evaluated the Simulated Human Interaction’s usability (Heylen, Van Es, Nijholt, & van Dijk, 2002).

Agents displaying adequate proxemic and gaze behavior also seem to encourage users to show stronger gaze and proxemic responses (Kolkmeier, Vroon, & Heylen, 2016), suggesting once again that, provided the level of human-likeness is acceptable, social behaviors in Simulated Human Interactions can be consistent with human-to-human interactions. Interestingly, proxemic and gaze behaviors seem to affect the Agents' personality as perceived by the users. Agents that exhibited more directed gaze and reduced interpersonal distance were evaluated with higher intimacy scores (Kolkmeier, Vroon, & Heylen, 2016).

Simulated Human Interactions with eye-tracking integrated models capable of reacting in real-time to the user's gaze by moving the Agent's accordingly have also been created during recent years. Users interacting with a gaze aware Virtual Agent gave higher evaluations of Social Presence, Rapport, Engagement, Social Attraction, and Perception of Story (Kevin, Pai, & Kunze, 2018). They also rated the Agents' head and eye movements as more natural (as opposed to artificial; Krejsa, Kerou, & Liarokapis, 2018). This also seems to be valid for passing crowds of Virtual Agents. Dynamic gaze behavior leading the virtual passers-by to look at the user when they enter the Agent's field of view increases both its behavioral plausibility and the general quality of the virtual experience (Narang, Best, Randhavane, Shapiro, & Manocha, 2016). Virtual Agents in ENACT are presently capable of 10 different body postures and relative gaze directions connected with the five Negotiation Styles theorized by Rahim (Rahim, 1983), based on face-to-face observational data (Marocco et al., 2015). It would be interesting to investigate if the Agents' present gazing behavior is enough to create the illusion of being dynamically responsive to the environment and, therefore, to be perceived as human-like.

As for the quality of the Virtual Agent's voice, future studies might be conducted to explore more deeply users' perception of several qualities of the gibberish voice, such as pitch, volume, pace, tone, and timbre. Moreover, giving the Virtual Agent a human voice might introduce a further exciting topic: consistency between the Agent's voice and appearance. According to the literature, adverse reactions seem to be expected if there is a mismatch between the Virtual Agent's appearance and voice. For instance, pairing a synthetic child-like voice with a male adult appearance ensured that a social robot fell into the Uncanny Valley (Zlotowski et al., 2015). As the entertainment industry increasingly relies on the same actor as the motion capture reference and the Virtual Agent's voice, habitual video game players might be more sensitive to mismatches between the Virtual Agent's appearance and its voice.

Moreover, it would be interesting to explore the consequences of the (accidental or intentional) introduction of "quirks", habits, and verbal idiosyncrasies in the Virtual Agent's voice. Several studies have suggested how Virtual Agents' pronunciation and prosody errors might be viewed positively by users, as a flawless synthetic voice could be perceived as "too perfect" and, therefore, alienating (Scott, Ashby, & Hanna, 2020). On the contrary, mistakes might be considered as patterns of strength and weaknesses that outline someone's personality, evoking feelings of connection and engagement in listeners (Plantec, Mauldin, Romero, & McBride, 2004).

Finally, another characteristic that has been explored for its mediating effect in user perception is the gender of the Virtual Agent's voice. According to a recent study, users prefer a more mechanical feminine voice than a good quality masculine voice. The reason behind this unexpected selection was thought to be the habit of having a female voice as a Virtual Assistant, therefore affecting users' preferences in exciting ways (Perez & Saffon, 2018). It would be worth exploring if such results are confirmed for embodied Virtual Agents.

Epilogue

.

The fundamental goal in creating Digital Humanity should be not as much creating human-like Virtual Agents that are realistic in an absolute sense, but rather Virtual Agents which are perceived to be believable in the digital world they live in (Chowanda, Flintham, Blanchfield, & Valstar, 2016; Monnin, 2010), appearing and acting within a defined set of rules that apply within that same world (Brown, 2012).

In this sense, by being nothing other than "simulacra of real relationships" (Domsch, 2017), the idea of believability does not necessarily come from a realistic Agent, but from one "*that provides the illusion of life, and thus permits the audience's suspension of disbelief*" (Bates, 1994).

[this page intentionally left blank]

Epilogue Art: Detroit Become Human artwork from Quantic Dream (2018)

Bibliography

Adams, R.B., & Kleck, R.E. (2003). Perceived gaze direction and the processing of facial displays of emotion. *Psychological Science*. 2003; 14:644–647.

Adolphs, R. (2002). Recognizing emotion from facial expressions: psychological and neurological mechanisms. *Behav. Cogn. Neurosci. Rev.* 1, 21–62.

Akiwa, Y., Suga, Y., Ogata, T., & Sugano, S. (2004) Imitation based Human-Robot Interaction - Roles of Joint Attention and Motion Prediction. in *13th IEEE International Workshop on Robot and Human Interactive Communication*. Kurashiki, Japan.

Almeida, P. L., Moita, M., Magalhães, R., Santos, A., & Moreira, L. (2011). Serious Games as an onboarding organizational tool for new employees. In *6th Iberian Conference on Information Systems and Technologies (CISTI 2011)*. IEEE.

Andrews, T. J., Schluppeck, D., Homfray, D., Matthews, P., & Blakemore, C. (2002). Activity in the fusiform gyrus predicts conscious perception of Rubin's vase–face illusion. *Neuroimage*, *17*(2).

Anolli, L., & Mantovani, F. (2011). *Come funziona la nostra mente. Apprendimento, simulazione e Serious Games*. Il Mulino.

Asch, S. E. (1956). Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychological Monographs: General and Applied*, 70(9).

Bales, R. F., & Strodtbeck, F. L. (1951). Phases in group problem-solving. *The Journal of Abnormal and Social Psychology*, *46*(4).

Bar-On, R. (2004). The Bar-On Emotional Quotient Inventory (EQ-i): Rationale, description and summary of psychometric properties. In G. Geher (Ed.), *Measuring emotional intelligence: Common ground and controversy* (p. 115–145). Nova Science Publishers.

Bartneck, C., Kanda, T., Ishiguro, H., & Hagita, N. (2009). My robotic doppelgänger-A critical look at the uncanny valley. In *RO-MAN 2009-The 18th IEEE International Symposium on Robot and Human Interactive Communication*.

Bates, J. (1994). The role of emotion in believable agents. *Communications of the ACM*, *37*(7).

Batt, S.E. (January 2015) Are cutting-edge graphics a double-edged sword? In *Paste*, retrieved from https://www.pastemagazine.com/games/how-important-are-cutting-edge-graphics/. Last accessed on Jan 18th, 2018.

Bedwell, W. L., Salas, E., & Fiore, S. M. (2011). Developing the 21st century (and beyond) workforce: A review of interpersonal skills and measurement strategies. In *NRC Workshop on Assessing 21st Century Skills.*

Ben-Yoav, 0., & Pruitt, D. G. (1984a). Accountability to constituents: A two-edged sword. *Organizational Behavior and Human Performance,* 34.

Ben-Yoav, 0., & Pruitt, D. G. (1984b). Resistance to yielding and the expectation of cooperative future interaction in Negotiation. *Journal of Experimental Social Psychology*, 34.

Benitez, M., Medina, F. J., & Munduate, L. (2018). Buffering relationship conflict consequences in teams working in real organizations. *International Journal of Conflict Management.*

Bentham, J. (1843). *The works* (Vol. 7). Tait.

Berg, J. H., & Piner, K. (1990). *Social relationships and the lack of social relationship* in Personal relationships and social support, W. Duck and R. C. Silver, Eds. London: Sage.

Birmingham, E., Svärd, J., Kanan, C., & Fischer, H. (2018). Exploring emotional expression recognition in aging adults using the moving window technique. *PLoS One* 3:10.

Black, P., & Wiliam, D. (2005). *Inside the black box: Raising standards through classroom assessment*. Granada Learning.

Blascovich, J., & Bailenson, J. (2011). *Infinite reality: Avatars, eternal life, new worlds, and the dawn of the virtual revolution.* William Morrow & Co.

Brack, J., & Kelly, K. (2012). Maximizing millennials in the workplace. *UNC Executive Development*, *22*(1).

Breazeal, C. (2003). Emotion and sociable humanoid robots. *International journal of human-computer studies*, *59*(1-2).

Brefczynski-Lewis, J.A., Berrebi, M., McNeely, M., Prostko, A., & Puce, A. (2011). In the blink of an eye: Neural responses elicited to viewing the eye blinks of another individual, Frontiers in Human *Neuroscience,* 5.

Brown, D. (2012). *The suspension of disbelief in videogames* (Doctoral dissertation, Brunel University School of Arts PhD Theses).

Buha, P. (2014). Employee Onboarding: how does proper onboarding improve the effectiveness of new hires, and what are the best practices. *Academia.edu*.

Burgoon J., Buller D., &Woodall W. (1996). *Nonverbal Communication: The Unspoken Dialogue.* McGraw-Hill Higher Education.

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).

Cain, M. S., Prinzmetal, W., Shimamura, A. P., & Landau, A. N. (2014). Improved control of exogenous attention in action video game players. *Frontiers in psychology, 5*.

Calder, A. J., Young, A. W., Keane, J., & Dean, M. (2000). Configural information in facial expression perception. *J. Exp. Psychol. Hum. Percept. Perform.* 26.

Cano, J. L., & Sáenz, M. J. (1999). Development of a project simulation game. *Project Management*, *5*(1).

Carnevale, P. J. & Pruitt D. J. (1992). Negotiation and Mediation. *Annual Reviews of Psychology,* 43.

Casalnuovo, C., Vasilescu, B., Devanbu, P., & Filkov, V. (2015). Developer onboarding in GitHub: the role of prior social links and language experience. In *Proceedings of the 2015 10th Joint Meeting on Foundations of Software Engineering*.

Cassell, J. (2000). Nudge nudge wink wink: Elements of face-to-face conversation for embodied conversational agents. *Embodied conversational agents, 1*.

Castagna, M. (2014). *Lezione*. In Quaglino, G. P. Formazione. I metodi. Milano: Raffaello Cortina Editori.

Chaminade, T., Hodgins, J., & Kawato, M. (2007). Anthropomorphism influences perception of computer-animated characters' actions. *Social cognitive and affective neuroscience*, *2*(3).

Cheetham, M., Suter, P., & Jäncke, L. (2011). The human likeness dimension of the "uncanny valley hypothesis": behavioral and functional MRI findings. *Frontiers in human neuroscience*, *5*.

Cho, E., Molina, M. D., & Wang, J. (2019). The effects of modality, device, and task differences on perceived human likeness of voice-activated virtual assistants. *Cyberpsychology, Behavior, and Social Networking*, *22*(8).

Chowanda, A., Blanchfield, P., Flintham, M., & Valstar, M. (2016). Computational models of emotion, personality, and social relationships for interactions in games. In *The 2016 International Conference on Autonomous Agents & Multiagent Systems*.

Clark, N., & Rutter, D. (1985) Social categorization, visual cues and social judgments. *European Journal of Social Psychology*, 15.

Cohen, J. (1988). The effect size. *Statistical power analysis for the behavioral sciences*, 77-83.

Cohen, M. H., Giangola, J. P., & Balogh, J. (2004). *Voice user interface design*. Addison-Wesley Professional.

Coleman, P. T., & Lim, Y. Y. J. (2001). A systematic approach to evaluating the effects of collaborative negotiation training on individuals and groups. *Negotiation Journal*, *17*(4).

Conty L., Tijus C., Hugueville L., Coelho E., & George N. (2006). Searching for asymmetries in the detection of gaze contact versus averted gaze under different head views: A behavioural study. *Spatial Vision*, 19.

Corrigan, L. J., Peters, C., Küster, D., & Castellano, G. (2016). Engagement perception and generation for social robots and virtual agents. In *Toward Robotic Socially Believable Behaving Systems-Volume I*. Springer, Cham.

Costa, S., Brunete, A., Bae, B. C., & Mavridis, N. (2018). Emotional storytelling using virtual and robotic agents. *International Journal of Humanoid Robotics*, *15*(03).

Coulson, M., Barnett, J., Ferguson, C. J., & Gould, R. L. (2012). Real feelings for virtual people: Emotional attachments and interpersonal attraction in video games. *Psychology of Popular Media Culture*, 1(3).

Craig, S. D., & Schroeder, N. L. (2017). Reconsidering the voice effect when learning from a virtual human. *Computers & Education*, 114.

Currall, S. C., Friedman, R. A., Tidd, S. T. & Tsai, J. C. (2000). What goes around comes around. The impact of personal conflict style on work conflict and stress. *The International Journal of Conflict Management*, 11(1).

Dalibard, S., Magnenat-Talmann, N., & Thalmann, D. (2012). Anthropomorphism of artificial agents: a comparative survey of expressive design and motion of virtual. *Characters and Social Robots.*

Damiano L., Dumouchel P., & Lehmann H. (2014), Towards Human-Robot Affective Co-evolution. *International Journal of Social Robotics*.

Datar, S. M., Garvin, D. A., Cullen, P. G., & Cullen, P. (2010). *Rethinking the MBA: Business education at a crossroads*. Harvard Business Press.

Dautenhahn, K., Nehaniv, C. L., Walters, M. L., Robins, B., Kose-Bagci, H., Mirza, N.A., & Blow, M. (2009). KASPAR–a minimally expressive humanoid robot for human–robot interaction research. *Applied Bionics and Biomechanics*, 6(3-4).

De Castell, S., & Jenson, J. (2003). OP-ED serious play. *J. Curriculum Studies*, *35*(6).

de Gelder, B., Teunisse, J. P., & Benson, P. J. (1997). Categorical perception of facial expressions: Categories and their internal structure. *Cognition & Emotion*, *11*(1).

Deal, J. J., Altman, D. G., & Rogelberg, S. G. (2010). Millennials at work: What we know and what we need to do (if anything). *Journal of Business and Psychology*, *25*(2).

Dede, C. (2010). Comparing frameworks for 21st century skills. *21st century skills: Rethinking how students learn*, *20*.

Deng, Z., Lewis, J. P., & Neumann, U. (2005). Automated eye motion using texture synthesis. *IEEE Computer Graphics and Applications*, *25*(2).

DiSalvo, C., Gemperle, F., Forlizzi, J., & Kiesler, S. (2001). All robots are not created equal: The design and perception of humanoid robot heads. In *Proc. Designing Interactive Systems*.

Dill, V., Flach, L. M., Hocevar, R., Lykawka, C., Musse, S. R., & Pinho, M. S. (2012, September). Evaluation of the uncanny valley in CG characters. In *International Conference on Intelligent Virtual Agents*. Springer, Berlin, Heidelberg.

Doane M.G. (1980). Interaction of eyelids and tears in corneal wetting and the dynamics of the normal human eyeblink. *American journal of ophthalmology* 1;89(4).

Domsch, S. (2017). Dialogue in video games. *Dialogue across Media, 28*.

Donath, J. (2001). Mediated faces. In *International Conference on Cognitive Technology* (pp. 373-390). Springer, Berlin, Heidelberg.

Doughty M.J. (2001) Consideration of three types of spontaneous eyeblink activity in normal humans: during reading and video display terminal use, in primary gaze, and while in conversation. *Optometry and Vision Science*, 1;78.

Druga, S., Williams, R. Breazeal, C., & Resnick, M. (2017). Hey Google, is it OK if I eat you? in: *Proceedings of the 2017 Conference on Interaction Design and Children*.

Dunbar R.I. (1992) Neocortex size as a constraint on group size in primates. *Journal of human evolution* 22(6).

Dyck, M., Winbeck, M., Leiberg, S., Chen, Y., Gur, R. C., & Mathiak, K. (2008). Recognition profile of emotions in natural and virtual faces. *PLoS One*, *3*(11).

Ekman, P. (1992). *Are there basic emotions?*

Elfenbein, H. A., Der Foo, M., White, J., Tan, H. H., & Aik, V. C. (2007). Reading your counterpart: The benefit of emotion recognition accuracy for effectiveness in negotiation. *Journal of Nonverbal Behavior*, 31(4).

Endrass, T., Schuermann, B., Kaufmann, C., Spielberg, R., Kniesche, R., & Kathmann, N. (2010). Performance monitoring and error significance in patients with obsessive-compulsive disorder. *Biological psychology*, *84*(2).

Entertainment Software Association, & Entertainment Software Association. (2017). *Sales, demographic and usage data: Essential facts about the computer and video game industry*. Washington, DC.

Epley, N., Akalis, S., Waytz, A., & Cacioppo, J. T. (2008). Creating social connection through inferential reproduction: Loneliness and perceived agency in gadgets, gods, and greyhounds. *Psychological science*, *19*(2).

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

Espinoza, C. (2012). *Millennial integration: Challenges millennials face in the workplace and what they can do about them* (Doctoral dissertation, Antioch University).

European Commission (2018). *Alfred: a virtual assistant helping older people stay active*. Retrieved from https://ec.europa.eu/digital-single-market/en/news/alfred-virtual-assistant-helping-older-people-stay-active.

Evinger C., Manning K. A., & Sibony P. A. (1991). Eyelid movements. mechanisms and normal data. *Investigative Ophthalmology & Visual Science* 32, 2.

Faigin, G. (2012). *The artist's complete guide to facial expression*. Watson-Guptill.

Farrell, L., & Hurt, A. C. (2014). Training the Millennial Generation: Implications for Organizational Climate. *E Journal of Organizational Learning & Leadership*, *12*(1).

Feldman, N. H., Griffiths, T. L. & Morgan, J. L. (2009). The influence of categories on perception: explaining the perceptual magnet effect as optimal statistical inference. *Psychological Review* 116(4).

Fenzel, J. L. (2013). *Examining generational differences in the workplace: Work centrality, narcissism, and their relation to employee work engagement.*

Fernet, C., Austin, S., & Vallerand, R. J. (2012). The effects of work motivation on employee exhaustion and commitment: An extension of the JD-R model. *Work & Stress*, *26*(3).

Ferris, G. R., Judge, T. A., Rowland, K. M., & Fitzgibbons, D. E. (1994). Subordinate influence and the performance evaluation process: Test of a model. *Organizational Behavior and Human Decision Processes*, 58.

Festinger, L. (1962). *A theory of cognitive dissonance* (Vol. 2). Stanford university press.

Fisher, R., Ury, W. & Patton, B. (1991). *Getting to yes Negotiating Agreement Without Giving in*. New York: Viking.

Foley, J. D., Van Dam, A., Feiner, S. K., & Hughes, J. F. (1995). *Computer Graphics: Principles and Practice*.

Freud, S. (1919). *The uncanny*. Fantastic literature: A critical reader.

Frischen, A., Bayliss, A. P., & Tipper, S. P. (2007). Gaze cueing of attention: visual attention, social cognition, and individual differences. *Psychological bulletin*, 133(4).

Funk, T. (2018). A Musical Suite Composed by an Electronic Brain: Reexamining the *Illiac Suite* and the Legacy of Lejaren A. Hiller Jr. *Leonardo Music Journal* 28.

Garcia, M. P., & Lopez, S. S. (2019). Exploring the Uncanny Valley Theory in the Constructs of a Virtual Assistant Personality. In *Proceedings of SAI Intelligent Systems Conference*. Springer, Cham.

Gatica-Perez D.: Modeling interest in face-to-face conversations from multimodal nonverbal behaviour. In *MultiModal Signal Processing: Methods and Techniques to Build Multimodal Interactive Systems*, Thiran J., Bourlard H., Marqués F., (Eds.), EURASIP and Academic Press series in signal and image processing. Elsevier Science.

Gee, F. C., Browne, W. N., & Kawamura, K. (2005). Uncanny valley revisited. In *ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005.*

Ghazali, A. S., Ham, J., Barakova, E., & Markopoulos, P. (2018). The influence of social cues in persuasive social robots on psychological reactance and compliance. *Computers in Human Behavior*, *87*.

Gilbert, J. (2011). The Millennials: A new generation of employees, a new set of engagement policies The Workplace. *Ivey Business Journal.*

Gibson, L. A., & Sodeman, W. A. (2014). Millennials and technology: Addressing the communication gap in education and practice. *Organization Development Journal*, 32(4).

Go, E., Sundar, S.S. (2019). Humanizing chatbots: the effects of visual identity and conversational cues on humanness perceptions. *Computers in Human Behavior,* 97.

Gobbini, M. I., and Haxby, J. V. (2007). Neural systems for recognition of familiar faces. *Neuropsychologia* 45.

Goleman, D. (2001). Emotional intelligence: Issues in paradigm building. *The emotionally intelligent workplace*, *13*, 26.

Gong, L., & Nass, C. (2007). When a talking-face computer agent is half-human and half-humanoid: Human identity and consistency preference. *Human communication research*, *33*(2).

Gorman, P., Nelson, T., & Glassman, A. (2004). The Millennial generation: A strategic opportunity. *Organizational Analysis*, *12*(3).

Graen, G. B., & Schiemann, W. A. (2013). Leadership-motivated excellence theory: an extension of LMX. *Journal of Managerial Psychology*.

Grant, W., Wassenhove, V. & Poeppel, D. (2004) ‘Detection of auditory (cross spectral) and auditory–visual (cross-modal) synchrony’, *Speech Communication*, Vol. 44.

Gratch, J., Wang, N., Gerten, J., Fast, E., & Duffy, R. (2007). Creating rapport with virtual agents. In *International workshop on intelligent virtual agents*. Springer, Berlin, Heidelberg.

Gratch, J., DeVault, D., & Lucas, G. (2016). The benefits of virtual humans for teaching negotiation. In *International Conference on Intelligent Virtual Agents* (pp. 283-294). Springer, Cham.

Gratch, J., DeVault, D., Lucas, G. M., & Marsella, S. (2015). Negotiation as a challenge problem for virtual humans. In *International Conference on Intelligent Virtual Agents*. Springer, Cham.

Greco, M., & Murgia, G. (2007). Improving negotiation skills through an online business game. In *Proceedings of the European Conference on Game Based Learning*.

Grimshaw, M. (2009). *The audio Uncanny Valley: Sound, fear and the horror game*.

Gross, M. A. & Guerrero, L. (2000). Managing conflict appropriately and effectively. An application of the competence model to Rahim's organizational conflict styles. *International Journal of Conflict Management*, 11 (3).

Guadagno, R. E., Blascovich, J., Bailenson, J. N., & McCall, C. (2007). Virtual humans and persuasion: The effects of agency and behavioral realism. *Media Psychology, 10*(1).

Guitton D., Simard R., Codère F. (1991) Upper eyelid movements measured with a search coil during blinks and vertical saccades. *Investigative Ophthalmology & Visual Science* 32.

Gursoy, D., Chi, C. G. Q., & Karadag, E. (2013). Generational differences in work values and attitudes among frontline and service contact employees. *International Journal of Hospitality Management, 32*.

Hanson, D., Olney, A., Pereira, S., & Zielke, M. (2005). Upending the uncanny valley. In *AAAI* (Vol. 5).

Hart, H. (2012). When will a motion capture actor win an Oscar? In *Wired magazine*, Retrieved from https://www.wired.com/2012/01/andy-serkis-oscars/

Hartmann, B., Mancini, M., & Pelachaud, C. (2005). Implementing expressive gesture synthesis for embodied conversational agents. In *International Gesture Workshop*. Springer, Berlin, Heidelberg.

Haslam, N., Loughnan, S., Kashima, Y., & Bain, P. (2008). Attributing and denying humanness to others. *European review of social psychology, 19*(1).

Haxby, J. V., Hoffman, E. A., and Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends Cogn. Sci.* 4.

Hershatter, A., & Epstein, M. (2010). Millennials and the world of work: An organization and management perspective. *Journal of business and psychology*, *25*(2).

Heylen, D., Van Es, I., Nijholt, A., & van Dijk, B. (2002). Experimenting with the gaze of a conversational agent. In *Proceedings international CLASS workshop on natural, intelligent and effective interaction in multimodal dialogue systems*.

Hiltz, S. R., Johnson, K., & Turoff, M. (1986). Experiments in group decision making communication process and outcome in face-to-face versus computerized conferences. *Human communication research*, *13*(2).

Hirokawa, R. Y. (1983). Group communication and problem-solving effectiveness II: An exploratory investigation of procedural functions. *Western Journal of Communication (includes Communication Reports)*, *47*(1).

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).

Ho, C. C., MacDorman, K. F., & Pramono, Z. D. (2008). Human emotion and the uncanny valley: a GLM, MDS, and Isomap analysis of robot video ratings. In *200 3rd ACM/IEEE International Conference on Human-Robot Interaction (HRI)*.

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).

Hömke, P., Holler, J., & Levinson, S. C. (2017). Eye blinking as addressee feedback in face-to-face conversation. *Research on Language and Social Interaction*, *50*(1).

Hömke, P., Holler, J., & Levinson, S. C. (2018). Eye blinks are perceived as communicative signals in human face-to-face interaction. *PloS one*, *13*(12).

Hopkins, M. M., & Yonker, R. D. (2015). Managing conflict with emotional intelligence: Abilities that make a difference. *Journal of Management Development.*

Huyler, D., Pierre, Y., Ding, W., & Norelus, A. (2015). Millennials in the workplace: positioning companies for future success. *SFERC 2015*, *114*.

IJsselsteijn, W. A., de Kort, Y. A., & Poels, K. (2013). The game experience questionnaire. *Eindhoven: Technische Universiteit Eindhoven*.

INACSL Standards Committee (2016). INACSL standards of best practice: Simulation SM Glossary Clinical Simulation in *Nursing*, 12.

ISFE (2019). The Video Game Industry. Retrieved from www.isfe.eu, last accessed Dec 27th, 2020.

Ishai, A. (2008). Let’s face it: it’s a cortical network. *NeuroImage* 40.

Jain, S., & Anjuman, A.S.S. (2013). Facilitating the acquisition of soft skills through training. *IUP Journal of Soft Skills*, 7(2).

Jennings, A. T. (2000). Hiring generation-X. *Journal of Accountancy*, *189*(2).

Jentsch (1906) *On the psychology of the uncanny*. R. Sellars, Trans., Angelaki 1(2).

Jin, J., & Rounds, J. (2012). Stability and change in work values: A meta-analysis of longitudinal studies. *Journal of Vocational Behavior*, *80*(2).

Johns M., Tucker A., Chapman R., Crowley K., Michael N. (2007). Monitoring eye and eyelid movements by infrared reflectance oculography to measure drowsiness in drivers. *Somnologie - Schlafforschung und Schlafmedizin* 11.

Johnson, R. D., Veltri, N. F., & Hornik, S. (2008). Attributions of responsibility toward computing technology: The role of interface social cues and user gender. *Intl. Journal of Human–Computer Interaction*, *24*(6).

Kaifi, B. A., Nafei, W. A., Khanfar, N. M., & Kaifi, M. M. (2012). A multi-generational workforce: Managing and understanding millennials. *International Journal of Business and Management*, *7*(24).

Kanai, R., and Rees, G. (2011). The structural basis of inter-individual differences in human behaviour and cognition. *Nat. Rev. Neurosci.* 12.

Kang, M. (2009) ‘The ambivalent power of the robot’, Antennae, *The Journal of Nature in Visual Culture*, 1(9).

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*.

Kelley, H. H., & Stahelski, A. J. (1970). Social interaction basis of cooperators' and competitors' beliefs about others. *Journal of Personality and Social Psychology*, 16.

Kevin, S., Pai, Y. S., & Kunze, K. (2018). Virtual gaze: exploring use of gaze as rich interaction method with virtual agent in interactive virtual reality content. In *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology*.

King, W. J., & Ohya, J. (1996). The representation of agents: Anthropomorphism, agency, and intelligence. In *Conference Companion on Human Factors in Computing Systems*.

Kivunja, C. (2015). Teaching students to learn and to work well with 21st century skills: Unpacking the career and life skills domain of the new learning paradigm. *International Journal of Higher Education*, *4*(1).

Klehm, O., Rousselle, F., & Papas, M. (2015). Recent Advances in Facial Appearance Capture. *Computer Graphics Forum* 34, 2.

Kolkmeier, J., Vroon, J., & Heylen, D. (2016). Interacting with virtual agents in shared space: Single and joint effects of gaze and proxemics. In *International Conference on Intelligent Virtual Agents*. Springer, Cham.

Knapp, M. L., Hall, J. A., & Horgan, T. G. (2013). *Nonverbal communication in human interaction*. Cengage Learning.

Krahn, H. J., & Galambos, N. L. (2014). Work values and beliefs of 'Generation X'and 'Generation Y'. *Journal of Youth Studies, 17*(1).

Krejsa, J., Kerou, B., & Liarokapis, F. (2018). A model for eye and head motion for virtual agents. In 2018 10th International Conference on Virtual Worlds and Games for Serious Applications (VS-Games). *IEEE*.

Kron, F. W., Fetters, M. D., Scerbo, M. W., White, C. B., Lypson, M. L., Padilla, M. A., ... & Guetterman, T. C. (2017). Using a computer simulation for teaching communication skills: A blinded multisite mixed methods randomized controlled trial. *Patient education and counseling*, *100*(4).

Kuhn, T. I. M., & Poole, M. S. (2000). Do conflict management styles affect group decision making? Evidence from a longitudinal field study. *Human communication research, 26*(4).

Kuron, L. K., Lyons, S. T., Schweitzer, L., & Ng, E. S. (2015). Millennials' work values: differences across the school to work transition. *Personnel Review*.

Le, B. H., Ma, X., & Deng, Z. (2012). Live speech driven head-and-eye motion generators. *IEEE transactions on visualization and computer graphics*, *18*(11).

Lee, S. P., Badler, J. B., & Badler, N. I. (2002). Eyes alive. In *Proceedings of the 29th annual conference on Computer graphics and interactive techniques*.

Leeb, R. T., & Rejskind, F. G. (2004). Here's looking at you, kid! A longitudinal study of perceived gender differences in mutual gaze behavior in young infants. *Sex Roles*, *50*(1-2).

Lehmann, H., Pattacini, U., & Metta, G. (2015). Blink-sync: Mediating human-robot social dynamics with naturalistic blinking behavior. In *Workshop on Behavior Coordination Between Animals, Humans and Robots; 10th ACM/IEEE International Conference on Human-Robot Interaction.*

Lehmann, H., Roncone, A., Pattacini, U., & Metta, G. (2016). Physiologically inspired blinking behavior for a humanoid robot. In *International Conference on Social Robotics*. Springer, Cham.

Lemaignan, S., Fink, J., Dillenbourg, P., & Braboszcz, C. (2014). The cognitive correlates of anthropomorphism. In *2014 Human-Robot Interaction Conference, Workshop" HRI: a bridge between Robotics and Neuroscience"*.

Lewin, D., & Gollan, P. J. (Eds.). (2019). *Advances in Industrial and Labor Relations*. Emerald Group Publishing.

Li, Z., & Mao, X. (2012). Emotional eye movement generation based on Geneva emotion wheel for virtual agents. *Journal of Visual Languages & Computing*, *23*(5).

Locke, E. A., Sirota, D., & Wolfson, A. D. (1976). An experimental case study of the successes and failures of job enrichment in a government agency. *Journal of Applied Psychology*, *61*(6).

Macaluso, E., George, N., Dolan, R., Spence, C. & Driver, J. (2004) 'Spatial and temporal factors during processing of audiovisual speech: a PET study', *NeuroImage*, 21(2).

MacDorman, K. F. (2006) Subjective ratings of robot video clips of human likeness, familiarity, and eeriness: An exploration of the uncanny valley. *ICCS/cogSci-2006 long symposium: Toward Social Mechanisms of android science,* 26-29.

MacDorman, K. F., & Chattopadhyay, D. (2016). Reducing consistency in human realism increases the uncanny valley effect; increasing category uncertainty does not. *Cognition*, *146*.

MacDorman, K. F., Green, R. D., Ho, C. C., & Koch, C. T. (2009). Too real for comfort? Uncanny responses to computer generated faces. *Computers in human behavior*, *25*(3).

Marocco, D., Pacella, D., Dell'Aquila, E., & Di Ferdinando, A. (2015). Grounding serious game design on scientific findings: the case of ENACT on soft skills training and assessment. In *Design for Teaching and Learning in a Networked World*. Springer, Cham.

Martin, C. A. (2005). From high maintenance to high productivity. What managers need to know about Generation Y. *Industrial and Commercial Training*, *37*(1).

Mason M.F., Tatkow E.P., & Macrae C.N. (2005). The look of love: Gaze shifts and person perception. Psychological Science, 16.

Mattys, S., Bernstein, L.E., Edward, T. & Auer, J. (2000) When lipreading words is as accurate as listening. *139th ASA Meeting*, Atlanta, Georgia.

Mayer, R. E. (2014). Principles based on social cues in multimedia learning: Personalization, voice, image, and embodiment principles. *The Cambridge handbook of multimedia learning*, *16*.

Mayer, J. D., Caruso, D. R., & Salovey, P. (2016). The ability model of emotional intelligence: principles and updates. *Emot. Rev.* 8.

McCrindle, M., & Wolfinger, E. (2010). Generations defined. *Ethos*, *18*(1).

McDonnell, R., Breidt, M., & Bülthoff, H. H. (2012). Render me real? Investigating the effect of render style on the perception of animated virtual humans. *ACM Transactions on Graphics (TOG)*, *31*(4).

McGurk, H. and MacDonald, J. (1976) Hearing lips and seeing voices, *Nature* 1(264).

McTear, M., Callejas, Z. & Griol, D. (2016) *The Conversational Interface: Talking to Smart Devices*, Springer.

Meng, M., Cherian, T., Singal, G., & Sinha, P. (2012) Lateralization of face processing in the human brain. *Proceedings of the Royal Society B: Biological Sciences 279*, 1735.

Metta, G., Natale, L., Nori, F., Sandini, G., Vernon, D., Fadiga, L., Von Hofsten, C., Rosander, K., Lopes, M., Santos-Victor, J. & Bernardino, A., (2010). The iCub humanoid robot: An open-systems platform for research in cognitive development. *Neural Networks*, 23(8).

Mezzaroba, O., & Silveira, V.O.D. (2018). The principle of the dignity of human person: A reading of the effectiveness of citizenship and human rights through the challenges put forward by globalization. *Revista de Investigações Constitucionais*, *5*(1).

Minato, M. Shirmada, H. Ishiguro, and S. Itakura. (2004). Development of an Android Robot for Studying Human-Robot Interaction. In *17th International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems*. Ottawa, Canada: Springer-Verlag GmbH.

Monnin, A. (2010). Humanity and Digital Characters in Virtual Worlds: Crossing the Fictional Boundaries.

Monoriti, A. & Gabellini, R. (2018). *NegoziAzione. Il manuale dell'interazione umana*. Milano: Giuffrè Francis Lefebvre.

Moore, R. K. (2012). A Bayesian explanation of the 'Uncanny Valley' effect and related psychological phenomena. *Scientific reports*, *2*(1).

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

Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley [from the field]. *IEEE Robotics & Automation Magazine*, *19*(2).

Myers, K. K., & Sadaghiani, K. (2010). Millennials in the workplace: A communication perspective on millennials' organizational relationships and performance. *Journal of business and psychology*, *25*(2).

Nakamura, Y. (January 23, 2019). *Peak Video Game? Top Analyst Sees Industry Slumping in 2019*. Bloomberg L.P. Retrieved January 29, 2019.

Nakano T., Kato M., Morito Y., Itoi S., & Kitazawa S. (2013). Blink-related momentary activation of the default mode network while viewing videos. *Proceedings of the National Academy of Sciences*, 110(2).

Nakano T., & Kitazawa S. (2010) Eyeblink entrainment at breakpoints of speech. *Experimental Brain Research* 205(4).

Narang, S., Best, A., Randhavane, T., Shapiro, A., & Manocha, D. (2016). PedVR: Simulating gaze-based interactions between a real user and virtual crowds. In *Proceedings of the 22nd ACM conference on virtual reality software and technology*.

Nass, C., & Gong, L. (2000). Speech interfaces from an evolutionary perspective. *Communications of the ACM*, *43*(9), 36-43.

Nass, C., & Lee, K. M. (2001). Does computer-synthesized speech manifest personality? Experimental tests of recognition, similarity-attraction, and consistency-attraction. *Journal of experimental psychology: applied*, *7*(3), 171.

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of social issues*, *56*(1).

Nijholt, A. (2003). Disappearing computers, social actors and embodied agents. In *Proceedings. 2003 International Conference on Cyberworlds* IEEE.

Novick, D., & Gris, I. (2014). Building rapport between human and ECA: A pilot study. In *International Conference on Human-Computer Interaction*. Springer, Cham.

Nowak, K. L., & Fox, J. (2018). Avatars and computer-mediated communication: a review of the definitions, uses, and effects of digital representations. *Review of Communication Research*, *6*.

O'Hagan, M. (2009). Towards a cross-cultural game design: an explorative study in understanding the player experience of a localised Japanese video game. *The Journal of Specialised Translation*, 11.

Ohman, A. (1993). Fear and anxiety: Evolutionary, cognitive and clinical perspectives. *Handbook of Emotions*.

Özçelik, G. (2015). Engagement and retention of the millennial generation in the workplace through internal branding. *International Journal of Business and Management*, *10*(3).

Pallavicini, F., Ferrari, A., Pepe, A., Garcea, G., Zanacchi, A., & Mantovani, F. (2018). Effectiveness of virtual reality survival horror games for the emotional elicitation: Preliminary insights using Resident Evil 7: Biohazard. In *International Conference on Universal Access in Human-Computer Interaction*. Springer, Cham.

Perez, M., & Saffon, S (2018). Building trust between users and data driven virtual assistants. In: *14th International Conference on Artificial Intelligence Applications and Innovations* (AIAI).

Peters C., Pelachaud C., Bevacqua E., Mancini M., & Poggi I. (2005) A model of attention and interest using gaze behavior. In *Intelligent Virtual Agents, vol. 3661 of Lecture Notes in Computer Science*. Springer Berlin Heidelberg.

Pisoni, D. B., & Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Perception & psychophysics*, *15*(2).

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).

Plantec, P. M., Mauldin, M. L., Romero, J. S., & McBride, A. J. (2004). Virtual Humans. *U.S. Patent No. 6,826,540*. Washington, DC: U.S. Patent and Trademark Office.

Plutchik, R. (1980). A general psychoevolutionary theory of emotion. In *Theories of emotion*. Academic press.

Poggi, I., Pelachaud, C., & De Rosis, F. (2000). Eye communication in a conversational 3D synthetic agent. *AI communications*, *13*(3).

Pollick, F.E. (2010) *In search of the uncanny valley*. Berlin Heidelberg: Springer

Popescu, M. M., Romero, M., & Usart, M. (2013). Serious games for serious learning using SG for business, management and defence education. *Int. J. Comput. Sci. Res. Appl*, *3*(1).

Pruitt, D. G. (1983). Strategic choice in Negotiation. *American Behavioral Scientist*, 27.

Pruitt, D. G. & Carnevale, P. J. (1993). *Negotiation in social conflict*. Pacific Groves (CA): Brooks/Cole.

Pruitt, D. G., Carnevale, P. J., Ben-Yoav, 0., Nochajski, T. H. & Van Slyck, M. (1983). Incentives for cooperation in integrative bargaining. In R. Tietz (a cura di), *Aspiration levels in bargaining and economic decision making*. Berlin: Springer.

Pueringer, N. (2019a, May 25). *VFX Artists React to MARVEL Bad & Great CGi*. YouTube. https://www.youtube.com/watch?v=HWnRuPZ1Exg&t=48s

Pueringer, N. (2019b, October 19) *VFX Artists React to Bad & Great CGi 11*, retrieved from https://www.youtube.com/watch?v=erY_8JRgUiI

Qiu, L., & Bensabat, I. (2005). The Effects of Text-to-speech Voice and 3D Avatars on Consumer Trust in the Design of Live Help Interface of Electronic Commerce. In *Proceedings of the 10th Americas Conference on Information Systems*. New York, USA.

Rahim, M. A. (1983). A measure of styles of handling interpersonal conflict. *Academy of Management Journal*, 26.

Rahim, M. A. & Bonoma, T. V. (1979). Managing organizational conflict: A model for diagnosis and intervention. *Psychological Reports*, 44.

Rahim, M. A. & Buntzman, G. F. (1990). Supervisory power bases, styles of handling conflict with subordinates, and subordinate compliance and satisfaction. *Journal of Psychology*, 123.

Rahim, M. A., Garrett, & J. E., Buntzman, G. F. (1992). Ethics of managing interpersonal conflict in organizations. *J Bus Ethics*, 11.

Rahim, M. A., Psenicka, C., Polychroniou, P., & Zhao, J. H. (2002). A model of emotional intelligence and conflict management strategies: A study in seven countries. *International Journal of Organizational Analysis*, 10(4).

Raiffa, H. (1982). *The art and science of Negotiation*. Cambridge (MA): University Press.

Ramsey. F. (2014). Surround sound - an imitation of reality or an illusion?. *Society of Acoustic Engineers (Audio Engineering Society).*

Raybourn, E. M. (2007). Applying simulation experience design methods to creating serious game-based adaptive training systems. *Interacting with computers, 19*(2).

Reichheld, F. F. (2003). The one number you need to grow. *Harvard business review*, *81*(12).

Richmond, V.P., Wagner, J. P. & McCroskey, J. C. (1983). The impact of perceptions of Leadership style, use of power, and conflict management style on organizational outcomes. *Communication Quarterly*, 31.

Riemer, K., & Kay, J. (2019). Mapping beyond the uncanny valley: A Delphi study on aiding adoption of realistic digital faces. In *Proceedings of the 52nd Hawaii International Conference on System Sciences*.

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

Robinson, S., & Stubberud, H. A. (2012). Millennial workforce: Communicating and multitasking. *International Journal of Management & Information Systems (IJMIS)*, *16*(4).

Rosenthal-von der Pütten, A. M., Krämer, N. C., Maderwald, S., Brand, M., & Grabenhorst, F. (2019). Neural Mechanisms for Accepting and Rejecting Artificial Social Partners in the Uncanny Valley. *Journal of Neuroscience*, *39*(33).

Ross Jr, W. H., Pollman, W., Perry, D., Welty, J., & Jones, K. (2001). Interactive video negotiator training: A preliminary evaluation of the mcgill negotiation simulator. *Simulation & Gaming, 32*(4).

Roubroeks, M., Ham, J., & Midden, C. (2011). When artificial social agents try to persuade people: The role of social agency on the occurrence of psychological reactance. *International Journal of Social Robotics*, 3(2).

Roubroeks, M., Midden, C., & Ham, J. (2009). Does it make a difference who tells you what to do?: Exploring the effect of social agency on psychological reactance. *Proceedings of the 4th international conference on persuasive technology*. ACM.

Ruhland, K., Andrist, S., Badler, J., Peters, C., Badler, N., Gleicher, M., ... & Mcdonnell, R. (2014). Look me in the eyes: A survey of eye and gaze animation for Virtual Agents and artificial systems.

Ruhland, K., Peters, C. E., Andrist, S., Badler, J. B., Badler, N. I., Gleicher, M., ... & McDonnell, R. (2015). A review of eye gaze in virtual agents, social robotics and hci: Behaviour generation, user interaction and perception. In *Computer graphics forum* (Vol. 34, No. 6).

Ruijten, P. A. (2015). On the perceived human-likeness of virtual health agents: towards a generalized measurement of anthropomorphism. In *Intelligent Virtual Agents-workshop on virtual health agents, At Delft*.

Ruijten, P. A., Bouten, D. H., Rouschop, D. C., Ham, J., & Midden, C. J. (2014, March). Introducing a rasch-type anthropomorphism scale. In *2014 9th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE.

Ruijten, P. A., de Kort, Y. A., & Kosnar, P. (2012). Bridging the gap between the home and the lab: a qualitative study of acceptance of an avatar feedback system. In *International Conference on Persuasive Technology*. Springer, Berlin, Heidelberg.

Roubroeks, M., Ham, J., & Midden, C. (2011). When artificial social agents try to persuade people: The role of social agency on the occurrence of psychological reactance. *International Journal of Social Robotics*, *3*(2).

Rumiati, R. & Pietroni, D. (2001). *La negoziazione. Psicologia della trattativa: come trasformare un conflitto in opportunità di sviluppo personale, organizzativo e sociale*. Milano: Raffaello Cortina Editore.

Sakata, K., Takubo, T., Inoue, K., Nonaka, S., Mae, Y., & Arai, T. (2004) Psychological Evaluation on Shape and Motions. iIn *13th IEEE International Workshop on Robot and Human Interactive Communication*. Kurashiki, Japan.

Sacks H, Schegloff EA, & Jefferson G. (1974). A simplest systematics for the organization of turn-taking for conversation. Language. 1974;50(4).

Sasaki, K., Ihaya, K., & Yamada, Y. (2017). Avoidance of novelty contributes to the uncanny valley. *Frontiers in Psychology, 8*.

Saygin, A. P., 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 and affective neuroscience, 7*(4).

Schlegel, K., Mehu, M., van Peer, J. M., & Scherer, K. R. (2018). Sense and sensibility: The role of cognitive and emotional intelligence in negotiation. *Journal of Research in Personality*, 74.

Schröter, F., & Thon, J. N. (2018). *Video Game Characters-Theory and Analysis*.

Scott, K. M., Ashby, S., & Hanna, J. (2020). " Human, All Too Human": NOAA Weather Radio and the Emotional Impact of Synthetic Voices. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*.

Senju A, Tojo Y, Yaguchi K, & Hasegawa T. (2005). Deviant gaze processing in children with autism: An ERP study. Neuropsychologia, 43.

Serzhanova, V. (2011). The Influence of the European Integration and Globalization Processes on a Contemporary State. *Annales Universitatis Apulensis Series Jurisprudentia*, *14*.

Seyama, J. I., & Nagayama, R. S. (2007). The uncanny valley: Effect of realism on the impression of artificial human faces. *Presence: Teleoperators and virtual environments, 16*(4).

Seymour, M., Riemer, K., & Kay, J. (2017). Interactive Realistic Digital Avatars-Revisiting the Uncanny Valley.

Shanahan, M. (2015). *The Technological Singularity*. The MIT Press Essential Knowledge series.

Sherer, K.R. (1978). Personality inference from voice quality- The loud voice of extroversion. *European Journal of Social Psychology,* 8(4).

Shimada, M., Minato, T., Itakura, S., & Ishiguro, H. (2018). Uncanny valley of androids and the lateral inhibition hypothesis. In *Geminoid Studies*. Springer, Singapore.

Short, T. X. (2017). Designing stronger AI personalities. In *Thirteenth Artificial Intelligence and Interactive Digital Entertainment Conference*.

Sicart, M. (2008). Defining game mechanics. *Game Studies, 8*(2).

Siegle G.J., Ichikawa N., Steinhauer S. (2008) Blink before and after you think: blinks occur prior to and following cognitive load indexed by pupillary responses. *Psychophysiology* 1;45(5).

Sobieraj, S., & Krämer, N. C. (2014). What is beautiful in cyberspace? Communication with attractive avatars. In *International Conference on Social Computing and Social Media*. Springer, Cham.

Soraci, P., Ferrari, A., Abbiati, F. A., Del Fante, E., De Pace, R., Urso, A., & Griffiths, M. D. (2020). Validation and Psychometric Evaluation of the Italian Version of the Fear of COVID-19 Scale. *International journal of mental health and addiction*.

Statista (2019). The video game industry in Europe. Retrieved from https://www.statista.com/statistics/788396/leading-gaming-markets-by-revenue-in-europe/. Last accessed Dec 27th, 2020.

Steckenfinger, S. A., & Ghazanfar, A. A. (2009). Monkey visual behavior falls into the uncanny valley. *Proceedings of the National Academy of Sciences, 106*(43).

Stein, J. P., & Ohler, P. (2018). Uncanny… But Convincing? Inconsistency Between a Virtual Agent's Facial Proportions and Vocal Realism Reduces Its Credibility and Attractiveness, but Not Its Persuasive Success. *Interacting with Computers*, 30(6).

Stern, S.E., Mullennix, J.W. &Yaroslavsky, I. (2006) Persuasion and social perception of human vs. synthetic voice across person as source and computer as source conditions. *Int. J. Hum. Comput. Stud.*, 64.

Straßmann, C., & Krämer, N. C. (2017). A categorization of Virtual Agent appearances and a qualitative study on age-related user preferences. In *International Conference on Intelligent Virtual Agents*. Springer, Cham.

Straßmann, C., & Krämer, N. C. (2018). A Two-Study Approach to Explore the Effect of User Characteristics on Users' Perception and Evaluation of a Virtual Assistant's Appearance. *Multimodal Technologies and Interaction*, 2(4).

Sullivan, S., Campbell, A., Hutton, S. B., & Ruffman, T. (2017). What's good for the goose is not good for the gander: age and gender differences in scanning emotion faces. *J. Gerontol. B Psychol. Sci. Soc. Sci.* 72.

Szerszen, K. A. (2011). *The Audio/Visual mismatch and the Uncanny Valley: An investigation using a mismatch in the human realism of facial and vocal aspects of stimuli* (Doctoral dissertation).

Tada, H, Omori Y, Hirokawa K, Ohira H, Tomonaga M. Eye-blink behaviors in 71 species of primates. *PloS one*. 2013 May 31;8(5).

Tajfel, H. (1981). Human groups and social categories: Studies in social psychology. *Cup Archive.*

Takashima, K., Omori, Y., Yoshimoto, Y., Itoh, Y., Kitamura, Y., Kishino, F. (2008) Effects of avatar's blinking animation on person impressions. In: *Proceedings of Graphics Interface*,. Canadian Information Processing Society, Toronto.

Tavinor, G. (2017). What's my motivation? Video games and interpretative performance. *The Journal of Aesthetics and Art Criticism*, *75*(1).

Taylor, K. A., Mesmer-Magnus, J., & Burns, T. M. (2008). Teaching the art of negotiation: Improving students' negotiating confidence and perceptions of effectiveness. *Journal of Education for Business*, *83*(3).

Terada, K., Jing, L., & Yamada, S. (2015, April). Effects of agent appearance on customer buying motivations on online shopping sites. In *Proceedings of the 33rd annual acm conference extended abstracts on human factors in computing systems*.

Thomas, K. W. (1992). *Conflict and Negotiation processes in organizations*. In M. D. Dunnette e L. M. Hough (a cura di), Handbook of industrial & organizational psychology: Vol. 3. Palo Alto, CA: Consulting Psychologists Press.

Thompson, J. C., Trafton, J. G., & McKnight, P. (2011). The perception of humanness from the movements of synthetic agents. *Perception*, *40*(6).

Tinwell, A., Grimshaw, M., & Nabi, D. A. (2015). The effect of onset asynchrony in audio-visual speech and the Uncanny Valley in virtual characters. *International Journal of Mechanisms and Robotic Systems*, *2*(2).

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).

Tinwell, A., Grimshaw, M., & Williams, A. (2010). Uncanny behaviour in survival horror games. *Journal of Gaming & Virtual Worlds*, *2*(1).

Torner, H. P., Carbonell, X., & Castejón, M. (2019). A comparative analysis of the processing speed between video game players and non-players. *Aloma: Revista de Psicologia, Ciències de l'Educació i de l'Esport, 37*(1).

Twenge, J. M. (2010). A review of the empirical evidence on generational differences in work attitudes. *Journal of Business and Psychology*, *25*(2).

Tynjälä, P., Slotte, V., Nieminen, J., Lonka, K., & Olkinuora, E. (2006). From university to working life: Graduates’ workplace skills in practice. *Higher education and working life: Collaborations, confrontations and challenges*.

Ury W. (2007). *Getting Past No. Negotiation in difficult situations*. New York: Bantam Books.

Van De Vliert E. & Hordijk, J. W. (1989). A Theoretical Position of Compromising Among Other Styles of Conflict Management. *The Journal of Social Psychology*, 129 (5).

Van Keuren, J. (2006). *Web-Based Instruction: A Practical Guide for Online Courses*. Rowman & Littlefield Education. 15200 NBN Way, PO Box 191, Blue Ridge Summit, PA 17214-0191.

Van Niekerk, W. M., De Klerk, M., & Pires-Putter, J. K. (2017). Conflict management styles and work-related well-being among higher education employees. *Journal of Psychology in Africa*, 27(3).

Vincent, K. (2005). Social and emotional wellbeing: Sew what. *Education Connect, 1*.

Von Bergen, J.M. (2010) Queasy about Avatars and Hiring Employees (online) http://www.philly.com/philly/blogs/jobs/Queasy_about_avatars_and_hiring_employees_.html

Von Neumann, J. e Morgenstern, O. (1944). *Theory of Games and Economic Behavior*. Princeton: Princeton University Press.

Voorveld, H. A., & Araujo, T. (2020). How Social Cues in Virtual Assistants Influence Concerns and Persuasion: The Role of Voice and a Human Name. *Cyberpsychology, Behavior, and Social Networking*.

Vuilleumier, P., and Pourtois, G. (2007). Distributed and interactive brain mechanisms during emotion face perception: evidence from functional neuroimaging. *Neuropsychologia* 45.

Wang, S., Lilienfeld, S. O., & Rochat, P. (2015). The uncanny valley: Existence and explanations. *Review of General Psychology, 19*(4).

Watanabe, T., & Higuchi, A., 1991. Facial Expression Graphics Feedback for Improving the Smoothness of Human Speech Input to Computers. *Advances in Human Factors/Ergonomics* 18.

Wayne, S.J., Liden, R. C., Graf, L.K., & Ferris, G. R. (1997). The role of upward influence tactics in human resource decisions. *Personnel Psychology*, 50.

Waytz, A., Cacioppo, J., & Epley, N. (2010). Who sees human? The stability and importance of individual differences in anthropomorphism. *Perspectives on Psychological Science*, *5*(3).

Wilmot, W. W. & Hocker, J. L. (2001). *Interpersonal conflict*. New York (NY): McGraw-Hill.

Winograd, T. (1996). Profile: Microsoft Bob. In Winograd, T. (Ed.) *Bringing Design to Software*. New York: ACM Press.

Wood, E., Baltrusaitis, T., Zhang, X., Sugano, Y., Robinson, P., & Bulling, A. (2015). Rendering of eyes for eye-shape registration and gaze estimation. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 3756-3764).

Wu, Z., Ning, Y., Zang, X., Jia, J., Meng, F., Meng, H., & Cai, L. (2015). Generating emphatic speech with hidden Markov model for expressive speech synthesis. *Multimedia Tools and Applications*, *74*(22).

Wyn, J. (2009). *Touching the future: Building skills for life and work*.

Xie, L., Jia, J., Meng, H., Deng, Z., & Wang, L. (2015). Expressive talking avatar synthesis and animation. *Multimedia Tools and Applications*, *74*(22).

Yaghoubzadeh, R., Kramer, M., Pitsch, K., & Kopp, S. (2013). Virtual Agents as daily assistants for elderly or cognitively impaired people. In *International workshop on intelligent Virtual Agents*. Springer, Berlin, Heidelberg.

Yamada, Y., Kawabe, T., and Ihaya, K. (2012). Can you eat it? A link between categorization difficulty and food likability. *Adv. Cogn. Psychol.* 8.

Yamada, Y., Sasaki, K., Kunieda, S., and Wada, Y. (2014). Scents boost preference for novel fruits. *Appetite* 81, 102–107.

Yee, N., Bailenson, J. N., & Rickertsen, K. (2007). A meta-analysis of the impact of the inclusion and realism of human-like faces on user experiences in interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems*.

Yilmazyildiz, S., Henderickx, D., Vanderborght, B., Verhelst, W., Soetens, E., & Lefeber, D. (2011). EMOGIB: emotional gibberish speech database for affective human-robot interaction. In *International Conference on Affective Computing and Intelligent Interaction*. Springer, Berlin, Heidelberg.

Yilmazyildiz, S., Latacz, L., Mattheyses, W., & Verhelst, W. (2010). Expressive gibberish speech synthesis for affective human-computer interaction. In *International Conference on Text, Speech and Dialogue*. Springer, Berlin, Heidelberg.

Yilmazyildiz, S., Verhelst, W., & Sahli, H. (2015). Gibberish speech as a tool for the study of affective expressiveness for robotic agents. *Multimedia Tools and Applications*, *74*(22).

Yoshikawa, Y., Shinozawa, K., Ishiguro, H., Hagita, N., Miyamoto, T. (2006). The effects of responsive eye movement and blinking behavior in a communication robot, in: *Intelligent Robots and Systems, IEEE/RSJ International Conference on IEEE*.

Yukl, G., & Tracey, J. B. (1992). Consequences of influence tactics used with subordinates, peers, and the boss. *Journal of Applied Psychology*, 77.

Zgodavova, K., Kisela, M., & Sutoova, A. (2016). Intelligent approaches for an organisation's management system change. *The TQM Journal*, *28*(5).

Zhao S. (2006). Humanoid social robots as a medium of communication. *New Media & Society,* 8.

Zlotowski, J. A., Sumioka, H., Nishio, S., Glas, D. F., Bartneck, C., & Ishiguro, H. (2015). Persistence of the uncanny valley: the influence of repeated interactions and a robot's attitude on its perception. *Frontiers in psychology*, *6*.

[this page intentionally left blank]

If you can't fly, then run
(Today we will survive)
If you can't run, then walk
(Today we will survive)
If you can't walk, then crawl
Even if you have to crawl, gear up

Special Thanks to

My colleagues and friends, G., A., J., C., E., G., M., D., T., and S.
My dear PhD companions M., G., D., I., E., R., L., and P.
My biggest fan and supporter L.
And J.H., for being proud of me.

[this page intentionally left blank]