Michele Fanelli

A Study on the Relationship between Social Presence, Emotional Facial Behaviour and Flow in Multiplayer Game Experiences

MASTER THESIS

submitted in fulfilment of the requirements for the degree of

Master of Science

Programme: Master's programme Game Studies and Engineering

Alpen-Adria-Universität Klagenfurt

Evaluator
Assoc. Prof. Dipl.-Ing. Dr. Mathias Lux
Alpen-Adria-Universität Klagenfurt
Institut für Informationstechnologie

Klagenfurt, April 2024

Affidavit

I hereby declare in lieu of an oath that

- the submitted academic paper is entirely my own work and that no auxiliary materials have been used other than those indicated,
- I have fully disclosed all assistance received from third parties during the process of writing the thesis, including any significant advice from supervisors,
- any contents taken from the works of third parties or my own works that have been included either literally or in spirit have been appropriately marked and the respective source of the information has been clearly identified with precise bibliographical references (e.g. in footnotes),
- I have fully and truthfully declared the use of generative models (Artificial Intelligence, e.g. ChatGPT, Grammarly Go, Midjourney) including the product version,
- to date, I have not submitted this paper to an examining authority either in Austria or abroad and that
- when passing on copies of the academic thesis (e.g. in printed or digital form), I will ensure that each copy is fully consistent with the submitted digital version.

I am aware that a declaration contrary to the facts will have legal consequences.

m.p. Place, date

Table of Contents

Table of Figures

Table of Tables

Table of Equations

Introduction

This master thesis work was inspired by an encountered GDC talk on *Portal 2* (Valve 2011), “Portal 2: Creating a Sequel to a Game That Doesn't Need One” (Wolpaw and Faliszek 2016), where the main writers of the game, Eric Wolpaw and Chet Faliszek, recount the designing and writing process of Portal 2. In the second half of the talk, the writers make some interesting claims:

> So there's this strange thing that happens in co-op. We get people in separate rooms again, they can't see each other. We sometimes test them with no microphones, so they couldn't talk to each other. There's no way they can communicate with each other, and yet they laughed, they smiled, they joked. Right? **They were having all these regular things you have when you're next to someone, all these regular social norms, they carry through into the game, even though they didn’t have the communication and the other person can’t see their faces, smiling or laughing.** So this is an important takeaway from anyone, anyone out there writing a game. If you're going to do a comedy game, make it co-op, because otherwise you're gonna have years of misery and self-doubt as you watch it playtested. (Faliszek 2016, 30:55 own emphasis)

Figure 1: A playtester keeping an immobile face while playing the single-player mode of Portal 2. Own creenshot taken at 30:20 from “Portal 2: Creating a Sequel to a Game That Doesn't Need One” (Wolpaw and Faliszek 2016).

In this extract, the writer describes the playtesting phase of the multiplayer mode of the game, during which the playtesters were isolated in separate rooms, playing with each other with no visual, textual or audio communication, except for the mechanics offered by the game (such as the avatars waving their hands and pointing to a direction). As described by Faliszek, albeit in the absence of any verbal communication, the playtesters were observed laughing, smiling and joking, all reactions that the same persons were not having while playtesting the single-player mode of the game. While playing the single-player mode, even though *Portal 2* is a comedic game and the playtesters were reporting finding it deeply humorous, they were not laughing, but instead, as shown in the GDC talk, the faces were immobile, like frozen (*Figure 1*).

It is not unusual to see a person maintain a straight serious face while focusing on an activity, it is instead a clear indicator that the person is entering the flow state. The flow is a mental state that, as will be discussed more in detail in a later section, makes the person completely absorbed and engaged in a specific activity (Csikszentmihalyi 1991). During the flow state, all the thoughts and emotions irrelevant to the activity are removed, as all mental energies are redirected to the task at hand. While in this condition, the individual is happy, efficient effortlessly focused, and intrinsically motivated (Moneta and Csikszentmihalyi 1999). In addition, the subject experiences some common effects known as flow characteristics, including a sense of control, loss of self-consciousness, concentration and time distortion (Nakamura and Csikszentmihalyi 2002). Time distortion, for instance, is caused by the phenomenon of the mind redeploying to the engaged activity the mental energies that are usually allocated to the function of paying attention to the time passing (Csikszentmihalyi 2008). If the activity that the individual is trying to perform does not require any form of social interaction, then the same mechanism occurs with facial expressions: the mind withdraws any kind of facial emotional behaviour in order to redeploy the mental resources to the activity. In light of these factors, it is not a surprise to see a straight face during the playtesting of a comedic game. The naturally emerging interrogative is “Why would the playtesters smile knowing that no other person could see them? Does social interaction, or the idea of interacting socially, change the flow state accordingly?”

Partly, these open questions don’t fall flat on the academic literature, as digital games have long been studied for their inherent capability of being enablers of social interaction (Bryce and Rutter 2003). This quality of games to naturally enable and fuel social interaction is believed to tap into the hardcoded and intrinsically rewarding human desire for connection (Aitken and Trevarthen 1997, Ryan, Rigby and Przybylski 2006). Social interaction within the boundaries of games has been identified as a primary draw and major motivator for the players (Sherry et al. 2006). Playing together leads to greater enjoyment (Gajadhar, deKort and IJsselsteijn 2008), higher immersion (Cairns et al.

2013), more emotional involvement and a more vivid feeling of being inside the game (Ravaja et al. 2006). As will be discussed in detail in the following section, plenty of studies have been conducted on the measurable qualities of social play, including experiments on the relative position of players (Cairns et al. 2013), on the aesthetics of the game (Lin 2013), on the player's grade of familiarity (Ravaja et al. 2006), on the game goal structure (Waddel and Peng 2014), on the game outcome (Kivikangas and Ravaja 2013), and on the influence of an observer (Thorsteinsson and James 1999).

To the principal investigator's knowledge, the connection between flow and social play has barely been brushed by researchers (Hsu and Lu 2004, Labonté-LeMoyne et al. 2016), with most game scholars even considering social play or more generally social interaction to be a diversion from the traditional formulations of flow and immersion (de Kort, IJsselsteijn and Poels 2007; Ermi and Mayra 2005; Brown and Cairns 2004; Csikszentmihalyi 1991). These contradictions sparked further interest in the mentioned open questions and eventually served as inspiration for the composition of this master's thesis work. An initial effort to formalize Faliszek's claims was made:

> Even with scarce mediated communication, in the presence of another person in the virtual environment, players' behaviour changes as there was colocation with said person (affecting consequentially the flow state).

These claims were taken and restructured into two distinct research questions, to which a third one was subsequently added:

> Is it possible to perceive social presence in a multiplayer game experience even with close to no communication?
>
> Is there a relationship between perceived social presence and a more visible facial emotional behaviour?
>
> Is it possible for a player to perceive both social presence and the state of flow?

The term social presence refers to the subjective experience of being present with a 'real' person and having access to their thoughts and emotions (Biocca 1997). In this case, it connotates the subjective perception of being in the presence of a real person in the same room. Social presence has long been a research focus in networked communications systems (Biocca and Harms 2002) since it was found that a more vivid perception tends to lead to greater enjoyment and social influence (Oh, Bailenson, and Welch 2018, 28). Social presence, the "sense of being with another" (Biocca, Harms and Burgoon 2003, 456), is considered to be a subset of the larger subjective experience known as presence, which denotes the subjective perception of feeling inside the mediated virtual environment (Slater and Wilbur 1997).

It was originally believed that presence could only be obtained through pervasive and immersive systems, making it dependent on the quality of the medium (Short, Williams and Christie 1976, 65). More recent research ditched this line of thought in favour of a view of social presence more dependent on individual factors rather than contextual factors (Walther 1992). Among the studies identified in a recent comprehensive systematic literature review (Oh, Bailenson, and Welch 2018), few researchers found evidence of the connection between social presence and the immersive properties of the medium, as most claimed the opposite (20). These results seem to offer a working ground for the first research question of this master thesis, as it would test if it is possible at all to perceive social presence without any form of communication besides the visual representation of the playing character controlled by the other person. The second research question concerns instead the effects of social presence, i.e. how the behaviour of the individual changes when perceiving to be in the presence of another sentient social entity. Interestingly it seems that the vast majority of the research corpus focuses on which factors contribute to perceiving social presence, rather than what effects social presence has. The second proposed research question intervenes in attempting to fill this research gap. The same can be said for the third research question which, as previously mentioned, finds fertile ground.

This master thesis study attempts to find an answer to the three above-mentioned research questions. It does so with an experiment in which participants are isolated in different rooms with no possible communication during the session, and are asked to play a multiplayer game experience, first against a bot (a game avatar controlled by the computer) and then against a known person playing in another room. Multiple introspective subjective measurements are administered to the participants in the form of self-reports. Before the game session, the participants are instructed to fill out a Hexad-12 questionnaire (Krath et al. 2023) in order to determine the player type, then after each session, the participants are given to compile a custom version of the Game Experience Questionnaire (Poels, De Kort and Ijsselsteikn 2007), containing the items from the scales measuring immersion, flow and social presence. Recordings of the in-game behaviour and of the facial behaviour are both taken during each game session, with the intent of studying the former through qualitative methods such as observations, and the latter quantitatively and statistically through the application of the OpenFace software, capable of extrapolating the numerical values of the facial action units from the recorded data. The goals of the study are the following: (1) identifying a statistically significant difference in the self-reported social presence scores between the computer-controlled agent setting and the human-controlled avatar setting, (2) identifying a statistically significant relationship between the self-reported social presence scores and the correspondent facial action units activation values, and (3) identifying a statistically significant relationship between the self-reported social presence and the flow scores. This manuscript will first provide an in-depth review of the academic knowledge of social play, social presence, flow, player personality, facial action units and game research methodologies, then it will describe in detail how the gathered knowledge is applied to design the most effective experiment, and lastly, it will report the results of all the statistical and interpretative analysis conducted over the data, discussing it comprehensively before concluding.

Literature Review

This master's thesis work starts with a comprehensive inquiry into the existing scholarly literature, aimed at gathering pertaining academic knowledge to better comprehend the psychophysiological processes that will be investigated in the study. Through the application of acquired knowledge, it will be possible to meticulously design an experiment model characterized by both efficiency and scientific rigour, and not overlapping with any antecedent experiments investigating the same psychophysiological processes.

Social Play

The influence of social dynamics on digital game experiences has garnered significant attention in academic literature, with studies pointing at digital games' natural capability to bring and create opportunities for social interaction (Bryce and Rutter 2003). Naturalistic observations in home environments spotlight such reflections, showing a propensity for collaborative play, even in single-player games (Carr et al. 2004). This way in which social dynamics influence and shape the digital game experiences is commonly referred to as social play. Isbister (2010) defines social play as any situation in which more than one player actively engages with a game at once. The definition includes all forms of multiplayer gameplay, such as playing co-located or online, playing in turn or asynchronously on the same device, and even playing alone in front of an audience. As studied, the effects of social play are measurable both in the case of co-located co-players and in the case of remote connection between the players (Cairns et al. 2013). The social setting is defined by context-related factors and by the game elements and mechanics, with opportunities for verbal or non-verbal communication also contributing to shaping the social context (de Kort and Ijsselsteijn 2008; Emmerich and Masuch 2016). The aesthetics of the game have been shown as well to influence the social behaviour of players (Lin 2013). Social play has been identified as a primary draw and major motivator for people to play digital games (Sherry et al. 2006), probably due to its natural capability to satisfy a wide-reaching human need to feel connected to one another (Ryan, Rigby and Przybylski 2006), tapping on human's nature of social beings, with intrinsically rewarding, hardcoded, motivation for social interaction (Aitken and Trevarthen 1997).

It has been studied extensively in the last twenty years how, apart from being a major motivator, social play represents also a key factor in the shaping of the game experience. Evidence is accumulating that playing socially leads to greater enjoyment (Gajadhar, deKort and IJsselsteijn 2008), with some studies even claiming that playing socially is more immersive than playing alone (Cairns et al. 2013). Game players report more positive experiences when playing with other humans compared to the computer (Lim and Reeves 2010), with the specificity of social-competitive situations such as playing against an opponent (Ravaja et al. 2006). Furthermore, the elicitations of different game experiences based on the nature of the partner are further widened in the multifacetedness of the grade of familiarity, such as in the case of the different reactions when playing with friends and with strangers (Ravaja et al. 2006; Peng and Hsieh 2012). Similarly, other studies explore the effects of the relationship type (friend vs stranger) in cooperative and competitive multiplayer contexts, revealing insights into the negative emotions experienced during social play, indicating that while hostility remains low across conditions, cooperative goal structures foster the desire for even more cooperation (Waddel and Peng 2014).

Another popular object of analysis within game experience research is the game outcome, which seems to present non-trivial relationships with the game experience. According to the study of Kivikangas and Ravaja (2013), failure in multiplayer games doesn't necessarily elicit negative responses from players, instead, it occasionally elicits positive responses, as the researchers found that the players of a first-person shooter game experienced increased positive emotions and arousal, and decreased negative emotions when their character is killed or injured. The authors of the study employed physiological techniques to measure players' emotional reactions in front of different game outcomes, and noticed how participants were responding positively to victory, but also to defeat, and showed negative emotions mostly only when defeating a friend.

The effects of social play are not limited to the different emotions elicited against the different playing partners, since, as given by Ibister's definition, social play occurs also in scenarios of players engaged with digital games in front of an audience. Laboratory research on social support shows that the presence of a supportive other attenuates physiological responses to behavioural changes, as performing a task in front of an observer increases sympathetic arousal compared to performing a task in the absence of the observer (Thorsteinsson and James 1999). Exacerbating these findings, other studies show as well how in the presence of an observer, individuals experience increased emotional arousal due to the inevitable high task performance evaluation potential given by the social situation (Ravaja et al. 2006).

Moved by the realization that gaming is as often about social interaction as it is interaction with the game content, De Kort, IJsselsteijin and Poels argue that digital gaming technology, due to its inherent capability to provide social play, should be regarded as social presence technology, "as it provides a setting and medium for interacting with others at a distance (e.g. in online gaming) and augments and enlivens communication in co-located settings" (2007, 2). Contemporary research tends to focus a lot on the social context of play, overlooking the interaction between the game itself and the social experience of players, thus not considering how, regardless of the design of the game, the act of play itself may be the main influence of social experiences (Hudson and Cairns 2016, 1).

Social play seems to be a key factor in affecting the game experience. This thesis experiment aims to study the degree of this influence, specifically on known phenomena such as immersion and flow, as well as on some physiological reactions, such as the involuntary activation of facial action units.

Social Presence

The notion of Social presence refers to the subjective experience of being present with a "real" person and having access to their thoughts and emotions (Biocca 1997). As such, offering high levels of social presence has long been one of the primary goals of networked communications systems (Biocca and Harms 2002). Social Presence also depends on how strongly one feels that they're communicating with an intelligent being aware of their presence (Biocca 1997). "[M]ultiple studies show that the vivid perceptions of another person often lead to greater enjoyment and social influence in neutral and positive contexts" (Oh, Bailenson, and Welch 2018, 28).

In the works of Biocca, Harms and Burgoon (2003), we can see a conceptual subdivision of social presence into three dimensions. The first dimension is copresence, which is the sensation of being in contact with another person. The feeling can be based on a sensory representation ("I can see the other person"), but it might also be the result of a strong mental representation of the other ("I can imagine the other person") (Ekman et al. 2012, 330). The dimension of psychological involvement refers instead to a level of personal understanding, perceived emotional engagement, and attention between participants. "Psychological involvement relates to the level of emotional and mental connectedness between people but is also influenced by the participants' subjective evaluations of the situation ("I can understand your feelings" or "I respond to how you feel")" (Ekman et al. 2012, 330). Psychological involvement is not only limited to positive interactions, as it emerges as well during conflicts where the participants interact on a mental or emotional level. Lastly, the dimension of behavioural involvement concerns situations in which people perform similar tasks or if the participants need to coordinate their behaviour in response to the other person's actions (e.g. playing a multiplayer game or taking turns in a conversation). In the first studies conducted on the field, it was believed that these dimensions were three definable levels of the perceived social presence, starting with a low-level perception of other social entities, followed by psychological involvement, then behavioural engagement, ending on mutual co-presence (Biocca, Harms and Gregg, 2001; Biocca and Harms, 2002). More recent studies reconceptualize the three perceptions as three separate scales (De Kort, Ijsselsteijin and Poels 2007)

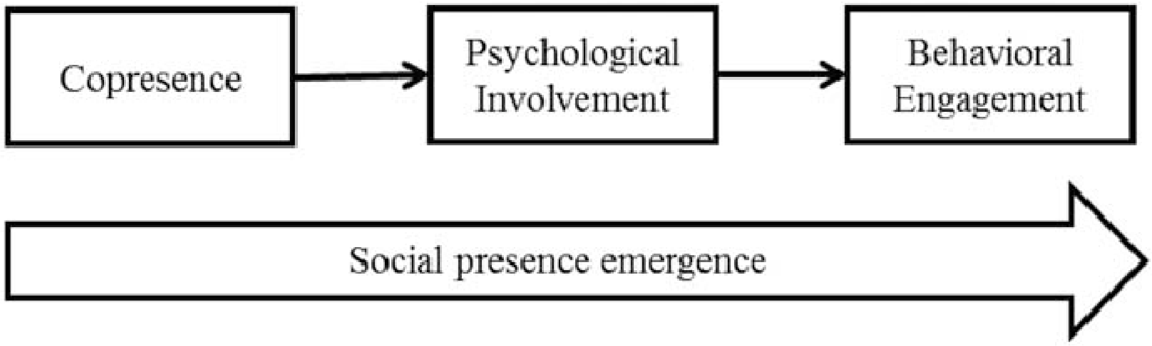


Figure 2: Copresence, psychological involvement and behavioural engagement were originally conceptualised as three definable levels of social presence in ordinal succession, emerging from greater levels of social presence. "Social presence emergence as a process (adapted from Biocca, Harms, & Gregg, 2001)" (Kohonen-Aho and Alin 2015).

In their extensive Systematic Review of Social Presence, Oh, Bailenson, and Welch (2018) analyzed 233 different findings from 152 studies that investigated the factors, characteristics, and qualities of social presence. This subsection elaborates on many of the grouped findings highlighted in their work, since as claimed by the authors, using consciously these results could maximize the amount of social presence the users can feel while interacting in the virtual environment (ibid, 2). Among the analyzed studies, some point out that social presence can be influenced by contextual and individual factors, i.e. the communication context and the individual traits of the person, which can impact the perception of the psychological distance between the interactants (e.g., Siriaraya and Ang 2012; Kang and Gratch 2014; Verhagen et al. 2014). Oh, Bailenson and Welch's Systematic Literature Review (2018) discusses the predictors of social presence and groups them into three overarching categories: immersive qualities, contextual properties and individual traits.

The first analysed topic shows us that immersion is not a relevant factor in inducing social presence. In this field, researchers tend either to use their terms and definitions or to borrow concepts that they adapt to their study. As such we can see a misleading use of terms, first of all, the concepts immersion and presence which are often used interchangeably. Oh, Bailenson and Welch (2018) start their work with an attempt to clarify such confusion, by defining immersion as the medium's technological capacity to generate realistic experiences that can remove people from their physical reality (cf. Slater and Wilbur 1997), and presence as the subjective experience of actually being in the mediated virtual environment (cf. Slater and Wilbur 1997; Witmer and Singer 1998;

Walther and Parks 2002). Therefore, while immersion is a characteristic of those media capable of delivering “an inclusive, extensive, surrounding and vivid illusion of reality to the senses of a human participant” (Oh, Bailenson, and Welch 2018, 604), presence is a personal instinctual experience, which is needed to fully experience a virtual environment (Cummings et al. 2012).

Furthermore, presence can be divided into three distinct subcategories: telepresence (also called spatial presence), self-presence, and social presence (Lee 2004). Telepresence is “the extent to which one feels present in the mediated environment, rather than in the immediate physical environment” (Steuer 1992, 75). Self-presence is the extent to which the “virtual self is experienced as the actual self ” (Aymerich-Franch et al. 2012, 1). Self-presence differs from telepresence for not concerning the vividly of the perceived surroundings, but the personal connection the user feels to their virtual body, their emotions and identity (Ratan and Hasler 2009). Finally, social presence refers to the “sense of being with another” (Biocca et al. 2003, 456) and is dependent on the subjective perception to have “access to the intelligence, intentions, and sensory impressions of another” (Biocca 1997, 22). Unlike the previous two features which only concern the user and the system, social presence requires the user to be co-present with another entity that somehow appears sentient (Oh, Bailenson, and Welch 2018, 2). Social presence is a key component of all virtual environments designed to mediate people, as without it the mediated other is perceived as an artificial entity and not as a social being (Lee et al. 2006).

The research on social presence ideally splits into two clusters. On one side, the research stems from the belief that social presence is an inherent quality of the communication medium, which would therefore justify the researchers to focus on studying how the modality and the immersion influence the perceived social presence. The other side instead, brought forth by more recent studies, aims to broaden the views and consider the influence of contextual and individual factors. Some authors of the first side (Short, Williams and Christie 1976) argue that some media are inherently more capable than others to deliver social cues, emphasizing that social presence is a “quality of the medium” (65). In line with the Media Richness Theory (Daft and Lengel 1986), studies on the impact of the medium on social presence, analyze how social presence changes according to the medium when all other circumstances are equal (e.g., Axelsson et al. 2001; Moreno and Mayer 2004; Zhan and Mei 2013). On the other side, Walther (1992) opposes these medium-centric views and argues that individuals are capable of adapting to different communication media. This point of view is known as Social Information Processing Theory (SIPT) and it claims that social presence is dependent on the interactants rather than the medium itself, with environments with few verbal clues or no verbal cues at all (such as in text-based Computer-Mediated Communication) being able to produce the same amount of intimacy as face-to-face (FtF) communication, even though with more time. According to Walther (1996), people communicating through text-based CMC can eventually reach even higher levels of Social Presence than FtF interactants. While SIPT posits that Social Presence is not only determined by the medium characteristics, it does not deny its influence, nor the inherent differences among media types.

> "When individuals are only given limited communication options (e.g. short timespan, specific task type, etc) it is probable that the technological features of the environment will influence the level of Social Presence a person feels. At the same time, however, this perspective offers a more nuanced view of social presence; while the immersive qualities (i.e., computer system's technological capacity to deliver a vivid experience) can impact social presence, individual communication strategies as well as contextual differences have a significant effect on social presence." (Oh, Bailenson, and Welch 2018, 3)

Aligning with the traditions of Social Presence Theory (Short, Williams and Christie 1976) and Media Richness Theory (Daft and Lengel 1986) most scholars research the impact of medium modalities on social presence, mostly comparing CMC, FtF and various audiovisual modalities. FtF is considered to be the gold standard for social presence (Biocca et al. 2001), with most studies unsurprisingly confirming that lower levels of social presence are perceived in CMC (Cortese and Seo 2012; Zhan and Mei 2013; Biocca et al. 2001), and many other studies confirming as well that text-based CMC elicits lower social presence than most other modalities (Bente et al. 2008; Appel et al. 2012; Kim et al. 2014). However, granting more time to CMC interactants leads to equally desirable communication outcomes as the FtF counterparts, as the 2 months employed by Francescato et al. (2006) was sufficient for both groups of participants to adapt their communication strategies to the given platforms.

> "While the literature shows a general consensus that immersive virtual environments are more likely to generate greater feelings of telepresence compared to non-immersive virtual platforms (Cummings and Bailenson 2016), this does not appear to be the case for social presence. Among the 10 studies that we identified, only two studies found significant differences in social presence between an immersive platform and a non-immersive one (Schroeder et al. 2001; Heldal et al. 2005)." (Oh, Bailenson, and Welch 2018, 20)

One of the unique features that impact the perception of Social Presence in virtual environments is the visual representation of the communication partner, as most of the current evidence indicates that people feel higher levels of social presence when a visual representation is available (Harris et al. 2009; Oh, Bailenson, and Welch 2018, 20). Among photographic, anthropomorphic and behavioural realism (respectively how realistic it looks, how human-like it looks, and how much it behaves as a real human by blinking and breathing), behavioural realism seems to be the strongest predictor of Social Presence, with interactants perceiving higher levels of Social Presence when the virtual agents act in a perceived natural way such as nodding, blushing (Pan et al. 2008) and maintaining a natural amount of eye contact (Bente et al. 2008).

Shifting the focus of the study from the immersive qualities to the individual and contextual factors, as claimed by the researchers of the second side, social presence starts to be considered a subjective experience influenced both by the perceived physical and psychological distance between the interactants. Multiple studies found differences in agency depending on whether the virtual human was a human-controlled avatar or a computer-controlled agent. These studies tend to inform the interactant on the nature of the virtual human (actual person or computerized character) before the interaction. Most of these studies found that the interactants feel higher levels of Social Presence when the virtual human is thought to be controlled by an actual person rather than by a computer program (Appel et al. 2012), avatars require a lower threshold of realism compared to autonomous agents (Blascovich et al. 2002), and players exhibit higher psychological arousal when believing that their opponent in a game is an avatar rather than an agent (Lim and Reeves 2010). Interestingly, while most studies published before 2010 resulted in such claims, most of the studies published afterwards struggle to reproduce this positive relationship between social presence and agency. Oh, Bailenson and Welch (2018, 23) reflect on the possibility that more techno-domesticated users started to develop different expectations regarding how an avatar should behave in virtual environments, which makes it easier to detect an agent, and thus more difficult to perceive it as an another intelligent being.

Multiple studies consistently showed a positive relationship between physical proximity and the perception of social presence, with interactants who had completed an activity in the same room perceiving more social presence than interactants in different locations. The review authors ponder on whether the social presence was affected by the physical closeness or by the ability to see each other. Croes et al. (2016) found that both physical co-location and visibility (non-anonymity) were powerful predictors of Social Presence, thus inducing the authors to conjecture that the positive association between Social Presence and physical proximity stems from the combination of physical co-location and visibility. There is cogent evidence that the physical closeness between the interactants contributes to the perceived physiological distance and the perceived social presence, but according to Oh, Bailenson and Welch (2018), this effect is likely influenced by visibility.

Some studies explored the influence of the task type on the perceived social presence between the interactants. In one study the participants felt lower levels of social presence when observing their partner play a multiplayer game, compared to when their partner observed them or when they played together (Herrewijin and Poels 2015). "Just as nonverbal cues that implied the virtual human's awareness of the participant increased social presence (e.g., Bente et al., 2008), people may feel higher levels of social presence when the given task requires the virtual human to pay attention to and accommodate their behavior." (Oh, Bailenson, and Welch 2018, 24)

More recent studies suggest that putting the interactants in a context where the individuals are exposed to cues that indicate a social context (conversations, partners, groups) can lead to heightened levels of Social Presence (Choi and Kwak 2017; Lee and Nass 2004; Lee et al. 2005).

Social Presence in Games

As previously mentioned, social presence has long been one of the primary goals of networked communications systems since its perception leads to greater immersion and enjoyment. This being the case, the connection between social presence and digital games is inevitable, as digital games have as their direct aim enjoyment and immersion, seemingly inherently suitable for hosting and boosting social interaction. Inside digital games, social presence can be seen as the result of the complex interplay between reciprocal behaviour, communication and social context features (de Kort and Ijsselsteijn 2008). Schouten (2014) argues that in digital games, "social presence is the result of being in a social setting. The more opportunities for social interaction the setting has, the higher the degree of social presence will be."

Ravaja et al. (2006) examined whether the nature of the opponent (computer, friend or stranger) influences psychological aspects such as spatial presence and emotional responses when playing video games. The authors define spatial presence as the number of human senses stimulated by the medium together with the sensation of being inside the simulated world, which would be a conjunction of immersion and telepresence as defined by Oh, Bailenson and Welch (2018). The study results showed that playing with another human being, compared to playing against a computer, elicits higher spatial presence, engagement, anticipated threats, post-game challenge appraisals, psychological arousal, and more positively valenced emotional responses. In addition, playing against a friend elicits higher spatial presence, engagement, and self-reported physiological arousal compared to playing against a stranger. While the authors aimed to find connections with spatial presence and not social presence, the results of the study are still meaningful in crafting an experiment on social presence, as these different perceptions and states of mind are not entirely distinct but rather interconnected.

Ekman et al. (2012) utilized psychophysiological methodologies to investigate the experience of playing digital games socially. Their study introduces Physiological Linkage as a physiological metric, examining its possible connection with the perceived social presence. As the authors remark, while psychophysiology measurements do not directly reveal subjective experiences, they serve as proxies for underlying processes, offering quantitative indicators of emotion, attention, motivation, attitudes, and unconscious cognitive processes. Physiological Linkage, as a psychophysiological measurement, attempts to quantify the degree of synchronization among participants' physiological responses, addressing an open research question concerning the relationship between social presence and psychophysiology. Drawing from Dimensional Emotion Theory, which posits emotions along valence (how positive or negative the emotion is) and arousal (bodily activation) dimensions, the conceptualization of Physiological Linkage aligns with theories of emotional contagion, Theory of Mind, and mirror neuron system. According to Ekman et al., gaming activities, being driven and motivated by the game itself, may lead to heightened levels of shared actions among individuals engaging in the same game. For instance, if multiple individuals are executing identical physical movements, such as performing synchronized steps on a dance mat, their physiological responses are likely to exhibit parallels (Ekman et al. 2012, 11). Physiological Linkage is based on this theoretical synchronization of the physiological responses, and calculated on the joint samples. Another of their studies (Chanel, Kivikangas, and Ravaja 2012) highlighted a predictive relationship between physiological linkage and the Social Presence in Gaming Questionnaire (de Kort, IJsselsteijn, and Poels 2007), including a predictive relationship between social empathy and smiling, and between higher levels of reported social negative feelings and asynchronously frowning (Ekman et al. 2012, 12).

The Social Presence in Gaming Questionnaire (SPGQ), developed by de Kort, IJsselsteijn, and Poels (2007), serves as a widely utilized tool for assessing and measuring social presence in digital game research. Building upon the notion that digital gaming technology functions as social presence technology, enabling mediated interaction and enlivening communication in co-located settings, the authors aimed to improve research and investigation on the field by providing a tailored form of measurement. Inspired by the Networked Minds Measure of Social Presence (NMMSP) by Biocca, Harms, and Gregg (2001), the SPGQ integrates the three known dimensions: co-presence, physiological involvement, and behavioural involvement, along with additional items measuring positive and negative feelings. Through validation with 191 participants in a self-selected game context, the questionnaire revealed three subscales: Psychological Involvement (Empathy), Psychological Involvement (Negative feelings), and Behavioral Involvement. Modest but significant correlations were found between the Behavioural Involvement scale and the other two, while no correlation was found between the two Psychological Involvement scales, showing that the two measure different dimensions. The correlation found brings the authors to argue that “the more ‘socially present’ the other person is, the stronger their mutual influence on each other’s feelings, both positively and negatively toned” (De Kort, IJsselsteijin and Poels 2007, 5). Moreover, players exhibited higher empathy with longer gameplay sessions, and Psychological Involvement (Empathy) was observed to increase together with the increase in social presence and relatedness. Contrary to Biocca et al.’s conceptualizations, behavioural involvement remained consistently high across conditions, even within settings with low social presence and relatedness, emphasizing the natural influence that players’ virtual actions have on each other within the gaming context. These findings oppose conceptualization claiming that Behavioural Involvement needs the highest levels of social presence (Biocca, Harms and Gregg 2001; Biocca, Harms and Burgoon 2003), since, while this dimension could be hard to reach in most teleconferencing systems, in games it is naturally high. Co-presence did not emerge as a distinct dimension, empathy levels were higher for co-located opponents compared to mediated ones, while negative feelings showed no significant differences. Overall, co-located settings yielded the highest scores on social presence measures.

As social presence is an important aspect to take into consideration in analyzing game experience, it is even more important when analyzing game experience in team-based games, as multiple studies underline how the stronger motivators for playing these games are social reasons, not only competitiveness and challenge but also the possibility of cooperation and communication (Frostling-Henningsson 2009; Jansz and Tanis 2007). Hudson and Cairns (2016) conducted two studies aimed at exploring how the experience of playing influences social presence, inquiring how the effect of winning and losing influences the perceived social presence between co-players in team-based digital games. Building on the research conducted by Kivikangas and Ravaja (2013), which investigates the impact of failure on the gaming experience, Hudson and Cairns sought to examine the influence of a shift in context to a public setting. Specifically, they investigated how the gaming experience is altered when failure becomes a collective rather than an individual experience, thereby exploring the dynamics of team failure.

> [L]osing a game as a team has a degree of team responsibility and therefore could still be a full playing experience. Alternatively, it may reduce the feeling of team cohesion and hence reduce the sense of social

> connection that players seek. Conversely, winning a game may enhance the social experience and thus be a way to greatly enhance the value of the social play. Framed this way, playing socially in a public team could be a form of gamble with the social experience, where the winners take more away from the experience than the losers. Currently very little is known about this aspect of social playing. (Hudson and Cairns 2016, 2)

Expanding upon the Social Presence in Gaming Questionnaire, Hudson and Cairns designed a custom version with specific scales that distinguish between competitive social presence between opponents and cooperative social presence between teammates, the Competitive and Cooperative Presence in Gaming Questionnaire (CCPIG). The results of their first study provided evidence that losing does not influence competitive social presence, but it does reduce cooperative social presence. The results of the second study found that winning leads to an increased perception of cooperative social presence, but not significant differences in competitive social presence. In some instances, the authors remark, there was a significant difference in cooperative social presence between winning and losing. This, according to the authors, could point to a strong disconnection of the individual players from their teams, which rather than suggesting a causal relationship of the game outcome to social presence, seems to indicate the opposite: the disconnection between the players undermined the team cohesion and led it to failure. Cooperative social presence was found to be substantially higher in the winning players, but competitive social presence was found to have a smaller difference between the winning and losing players, only just achieving significance. Therefore, the amount of perceived social presence may impact not only enjoyment and engagement, as seen in the previous studies, but also the degree to which the individual is capable of coordinating with the team, thus affecting the outcome of the game session. The authors, however, remark that the players' recalled experience may be very well influenced by the game outcome, with players reporting more involvement in winning than losing, regardless of the actual felt experience during the play session. As more research on the topic accumulates, the multifaceted effects of social presence become more and more deserving of academic attention.

Flow

Flow is a highly desirable mental state in which a person is simultaneously cognitively efficient, happy, effortlessly focused, and intrinsically motivated (Moneta and Csikszentmihalyi 1999). Csikszentmihalyi (1975) initially introduced the concept of the flow state by investigating individuals engaged in diverse activities such as rock climbing, chess, and dance. Subsequently, flow theory has found application across various other domains, initially in sports, following the performance of peak athletes, then art, work, and education, and finally human-computer interaction and game experience research (Kiili et al. 2012). Flow is a state of mind characterized by complete absorption or engagement in a specific activity, during which individuals deliberately exclude irrelevant thoughts and emotions (Csikszentmihalyi 1991). In the moment the individual is in the flow state, they are experiencing an optimal experience, as they attain a positive psychological state wherein they are fully involved in the goal-driven activity in a way that it becomes all-encompassing, rendering other activities inconsequential. The heightened level of concentration is frequently concomitant with a sense of efficacy in the ongoing activity, a perception of success, and often, emotions characterized by inner peace, joy, or wonder. This state produces pleasant experiences and as such it is a provider of intrinsic motivation, making the activity worth pursuing even in the absence of extrinsic motivators such as rewards or purposes. This kind of intrinsic motivation has been a strong focus of research for human-computer interaction scholars and serious game researchers alike since this kind of motivation can be the key to engaging users with gameful systems (Perttula et al. 2017).

Later on, Nakamura and Csikszentmihalyi (2002) delineated nine distinct flow dimensions that constitute the flow experience, categorized into two groups: flow conditions and flow characteristics (Swann et al 2012). The first group, the flow conditions, consist of the antecedents of flow, i.e. the prerequisites for entering the specific state of mind. Flow conditions include challenge-skill balance, clear goals and immediate feedback. The perceived balance between the owned skill and the challenge provided by the activity is believed to be the central precondition for the onset of the flow experience. Flow characteristics, on the other hand, represent the subjective feelings of an individual during a flow state. Such characteristics include a strong sense of control, action-awareness merging, loss of self-consciousness, concentration and timelessness. Notably, it has been contended that the combination of these first eight dimensions culminates in a flow marked by intrinsic motivation, an autotelic experience, denoted as the ninth dimension. This autotelic experience is what brings enjoyment and represents the provider of the intrinsic reward. Despite the ongoing lack of definitional consensus regarding the categorization of these dimensions into conditions and characteristics within flow research, Csikszentmihalyi (1991) claims that individuals, when asked to reflect on their flow experiences, consistently reference some, if not all, of the above mentioned nine flow dimensions.

Studies investigating flow in a broad context have revealed that engaging the user in challenging tasks that require utilizing a high skill level, leads to deep concentration, absorption and immersion. Furthermore, the phenomenon of flow has been linked to various aspects such as learning, talent development, academic achievement and creative accomplishments within a professional setting (Csikszentmihalyi 1996; Csikszentmihalyi et al. 1993). In addition, as previously mentioned, flow has been an important focus of academic research in the field of game research, encompassing topics of peak performance, engagement, immersion and social interaction. Intrinsic motivation is the key focus of researchers and developers in the field of serious games, as these gameful solutions aim to be able to provide the right amount of cognitive or physical investment in order to provide the challenge-skill balance flow condition (Perttula et al. 2017).

As the challenge-skill balance is believed to be the most powerful flow predictor, it has been a priority even for game designers and developers, who aim to create highly challenging activities to make the players utilise their skills at their maximum, as in a "serious play" or "playful work" (Csikszentmihalyi and Schneider 2000) fashion. To be more accurate, the predictor doesn't simply require the two values to meet, but to balance each other both at medium to high levels. In Csikszentmihalyi's (1991) conceptualization of the subjective experience of flow, the most central condition for the flow experience to occur is that the individual uses a high level of skill to perform a highly challenging activity. The activity thus needs to be not too easy for the individual's skills, nor not too difficult. The goal needs to be reachable, and the player needs to have a reasonable chance of success, but only achievable by putting meaningful effort into it. Therefore if the high skill level meets the high challenge level, the individual is in flow, but if the balance is not met, the individual enters another state. If both challenge and skill are low, the individual gets apathetic and is disinterested in the activity. Similarly, if the skill level is high but the challenge level is low, the individual enters a relaxation state. Instead, if the challenge level is too high for the skill level the individual experiences anxiety (*Figure 3)*.

Some researchers studying the state of mind on a more psychophysiological level, suggest that the state of flow is characterized by an optimal neurophysiological state, which can measured through brain activity. Nacke, Stellmach and Lindley (2010) found a relationship between the degree of immersion of a player and a specific frequency in brain activity, generally associated with a decrease in mental workload. Léger et al. (2014) found that the flow state is connected as well with other brain activity frequencies, associated with a more relaxed and more automatic neurophysiological state. Neurophysiological activity is also believed to be affected by social interaction (Babiloni and Astolfi 2014) which, considering the rising presence of social interactions in games, points to the likability of flow and brain activity to be modulated by social interaction.

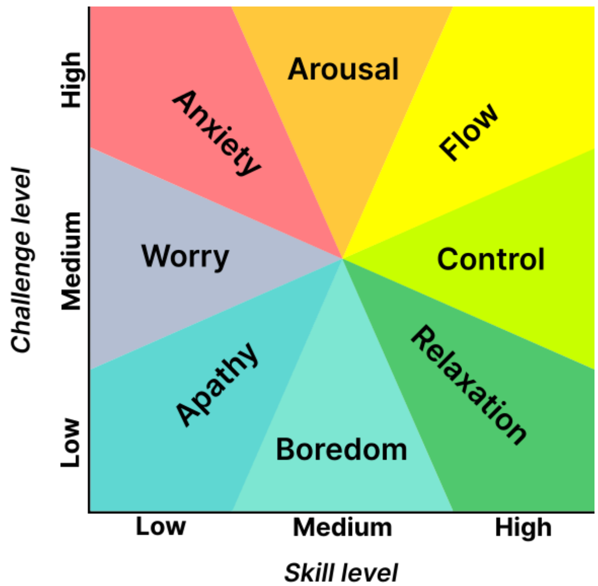


Figure 3: Mental state in terms of challenge level and skill level, according to Csikszentmihalyi's flow model. Beatson, Oliver. Wikipedia "Flow (psychology)". Accessed April 11, 2024.

Most scholars in Game Studies acknowledge the relevance of social factors but avoid incorporating them in their theoretical models, as social interaction and social context clash with the conceptualization of flow and immersion, which tends to be central in most formulations and models of game experience (Ermi and Mayra 2005; Brown and Cairns 2004; Sweester and Wyeth 2005).

In their study, Labonté-LeMoyne et al. (2016) employ electroencephalography (EEG) techniques to evaluate the flow state of participants playing a multiplayer game with a partner in the same room. The study, building on Bastarache-Roberge's (2015) research suggesting that the perceived flow state could be contingent on the psychophysiological characteristics of the player's partner, aims to explore the extent of the contingency between the psychophysiological characteristics of the two players. Starting on the hypothesis that there is a relationship between the state of flow and the neurophysiological activity of the counterpart, the authors develop an empirical study focusing on the use of neurophysiological techniques to fully capture both the explicit (self-reported) and implicit (unconscious and automatic) correlates of flow. The results show that the neurophysiological measures of each player are linked to their self-perceived flow state and the evidence of a relationship between a player's neurophysiological activity and the partner's flow state.

Game Research Methodologies

This thesis experiment consists of a quantitative and qualitative measurement of how social play affects a game experience, by identifying and measuring visible physiological aspects as well as introspective psychological aspects. To design and construct the model for the experimentation, it is needed to dwell on the study of the research methodologies of game experience or player experience research.

Game experience or payer experience is known to be a complex and multifaceted construct, which has been researched extensively in its whole and its parts. By mentioning its parts we refer to all those largely researched topics that focus on smaller aspects, such as social presence, affect and flow (Bernhaupt 2010; Ijsselsteijn et al. 2008). Player experience research has advanced greatly in the last decades, mostly with an increasingly prominent focus on Human-Computer Interaction (HCI), its techniques and their adaption to game-specific approaches (Drachen and Canossa 2011, Lankoski and Björk 2015). Measuring game experience is an increasing area of focus among both digital game researchers and developers (Denisova, Nordin and Cairns 2016; Drachen and Canossa 2009; Nacke and Drachen 2011, Johnson, Gardner and Perry 2018).

A vast variety of different methods has been employed by game experience researchers over the years. Most of these methods can be grouped into four main categories: (1) physiological measurements, (2) game metrics, (3) observations, and (4) self-reports. All of these different methods have their strengths and weaknesses. Physiological measurements are clear tools that offer objective analysis of the game experience, but that tend to be costly and hard to interpret (Lankoski and Bjork 2015). Some of the most commonly exploited physiological measurements include electroencephalography (cf. Labonté-Lemoye et al. 2016), heart rate variability (cf. Castellar, Oksanen and Looy 2014), respiration rate, and skin conductivity. Game metrics and in-game behaviour analytics are now an emerging area of focus, but it is still a challenge how to best connect in-game behaviour to the subjective experience. Self-reports, i.e. techniques such as questionnaires, interviews and think-aloud sessions, allow researchers to easily and inexpensively, quantitatively or qualitatively measure introspective feelings and thoughts. These are some of the most popular and effective employed tools to measure experiences, however, they are subjective by nature and mostly done retrospectively, hence sensitive to bias and social desirability (Emmerich and Masuch 2016, 2). Research groups such as Ekman et al. (2012) and Castellar, Oksanene and Looy (2014) also criticize the common practice of relying on self-reports, warning the practitioners of its embedded limitations, such as how the accuracy of the self-reports relies on psychological processes such as memory, attention and introspective ability, which can differ greatly among the subjects.

Nevertheless, the strong influence of HCI on game research made so that various self-report tools have been developed to evaluate gameplay experience (Law, Brühlmann and Mekler 2018). The vast variety of questionnaires employed, often in the absence of clear indications of their relative empirical validity and reliability seems however to represent a problem, in addition to the fact that such a large variety of tools makes it more difficult to compare the results across studies (Johnson, Gardner and Perry 2018). The questionnaire that game experience researchers and practitioners have most widely applied to a broad scope of game genres, user groups, gaming environments and purposes, seems to be the Game Experience Questionnaire (GEQ) (Poels, De Kort and Ijsselsteikn 2007; Law, Brühlmann and Mekler 2018). The GEQ is a multidimensional self-report measure of game experience, meant to probe the multidimensional subjective experiences associated with digital gaming and the relative feelings perceived at the just terminated game session. The tool is composed of three parts: the core questionnaire, the social presence module and the post-game modules. The three-factor scale of the social presence module has already been explained in a previous chapter, and the post-game module is not needed for our purpose. The core GEQ is the part that has been used most widely, and it consists of a seven-factor scale composed of Competence, Sensory and Imaginative Immersion, Flow, Tension/Annoyance, Challenge, Negative affect and Positive affect. The popularity of the GEQ as a self-report tool is mostly due to its multidimensional item structure, its simplicity and its availability, as well as the simple fact that it has been used in a vast multitude of other studies.

The described concerns seem to point to this questionnaire as the right choice for the study, however, even though the authors of the GEQ provided detailed information on the scale development process and explanatory evidence of the criterion of validity (Ijsselsteijn, de Kort and Poels 2007), they never published a complete formal validation of the factorial structure of the questionnaire, except for some overview articles (Cairns et al. 2013). Law, Brühlmann and Mekler (2018) conducted a systematic analysis of the research papers and articles that made use of the GEQ, identified problems in the provision and application of the questionnaire, and through a confirmatory factor analysis followed by an exploratory factor analysis, provided empirical evidence of an instability of the proposed factor structure of the self-report tool. More specifically, they found the Tension, the Challenge and the Negative Affect components to be highly problematic, showing a lot of cross-loading. Johnson, Gardner and Perry (2018) analysed the GEQ and checked its validity through exploratory and confirmatory factor analysis procedures. In the explanatory analysis, the seven-factor structure is not fully supported, suggesting a five-factor solution to best fit the data instead. The authors thus provide an updated questionnaire (GEQ-R), with an updated factor structure for the GEQ in which Negative Affect, Tension and Challenge reflect a single negativity factor, and with eight items presenting cross-loading removed.

To fully capture the multifaceted and complex phenomena that is player experience, or to attempt to do so, it is wiser and more efficient to employ multiple heterogeneous methodologies. The most common choice in the state of the art is to combine physiological measurements with self-reports (Ekman et al. 2012). There are multiple examples in the literature of studies applying this common choice. One example is the experiment by Castellar, Oksanen and Looy (2014), who aimed to assess to what extent in-game behaviour and self-report measures can predict heart rate variability during gameplay. Inspired by other studies that make use of game logs (cf. Wallner and Kriglstein 2015), the authors combine self-reports and in-game behaviour as objective indexes of game performance. To the authors' belief, in-game behaviour recordings provide complementary and supportive interpretation means to psychological data.

Combining self-reports with not only physiological data but also observations is a popular and efficient technique to evaluate game sessions and social interactions, as it allows researchers to register and integrate both verbal and non-verbal behaviour (Emmerich and Masuch 2016, 2). However, while observations such as screen recordings can be useful in offering a tool to analyze in-game behaviour, interpreting the material is a time-consuming task, which some authors (cf. Emmerich and Masuch 2016) disregard in favour of game metrics. Game metrics are numerical data directly logged by the game system during the play session, which are capable of providing objective quantitative measurements of player's interactions with the system. "Assessing and evaluating social play is relevant to reveal relationships between social factors and aspects of [player experience] and to investigate how game design elements can trigger certain behaviours" (Emmerich and Masuch 2016, 1). In their study, Emmerich and Masuch (2016) explore the potential and the applicability of gameplay metrics to access social behaviour and interactions between players. They do so through the conceptualisation, the application and the evaluation of metrics measuring respectively social presence, cooperation and leadership. The authors rely on the nine key types of social behaviour of players that may occur in co-located gaming sessions (as defined by Voida, Carpendale and Greenberg 2010): shared awareness, requesting information, shared history, shared success, shared failure, team optimization, trash talk and self-indulgence. In applying observation methods to game recordings, these behaviour patterns offer a useful basis for assessing social interaction. Emmerich and Masuch picked out the three previously mentioned social play aspects and studied the efficacy of game metrics as analysis tools. The self-report scores and the game metric scores showed high correlations, indicating that both methods asses data underlying the same construct. This supports their hypothesis that some aspects of social play can be investigated using gameplay metrics. Notably, the metrics did not fully correlate with all the subscales in the SPGQ (de Kort, Ijsselsteijin and Poels 2007) as Emmerich and Masuch focused on co-presence, only a subaspect of social presence.

Player Personality

In game studies and game research, user-typologies or gamer-typologies are often produced by scholars and developers in an attempt to better classify and understand the player. A famous example is Bartle’s player typology for Multi-User Dungeons (Bartle 1996), which divides players into Achievers, Explorers, Socializers and Killers. Such classifications are popular in the field of gamification as they allow researchers to better understand which game elements are gonna be more effective in motivating which kind of players. The personalization of gamified systems according to the user’s personality is often praised within the field of gamification and called for in academic literature, as personalizing gameful systems to each user proves to be way more effective than any one-size-fits-all solutions (Tondello et al. 2016). Gameful systems are believed to be more engaging when capable of adapting to personality traits or user types (Dixon 2011; Tondello et al. 2016). The topic of player typologies connects to this thesis experiment because the player’s personality is believed to influence the player’s fall into one of the player types (Nacke, Bateman and Mandryk 2014), which affects how players perceive psychological satisfaction (Johnson and Gardner 2010) and presence (Kober and Neuper 2013) in games. Being social presence a complex and multifaceted subjective perception and state of mind, it has been shown to be affected by individual factors such as the personality of the player. This being the case, it is suitable to add this analytical lens to the study.

Although being a popular game design taxonomy, Bartle’s player typology is old and tailored for Multi-User Dungeons, which are far from being applicable to a more generic version of gameful systems. Marczewski (2015) provides a new option by developing the Gamification User Types Hexad framework, based on motivation, player types and practical design experience. Tondello et al. (2016) decided to then provide a standard assessment protocol for user preferences based on the promising Hexad framework. Tondello et al.’s contribution to the field of gameful design in human-computer interaction is two-folded: firstly, they proposed and validated a survey measure for scoring user’s preferences towards different game design elements according to the Hexad framework, and secondly, they evaluated, resulting in multiple positive correlations, the potential of the Hexad framework as a model to personalize user experience in gameful systems (2016, 1-2).

In Human-Computer Interaction, scholars often base user behaviour research on the principles of Self-Determination Theory (Deci and Ryan 1985; Ryan and Deci 2000;

Ryan, Rigby and Przybylski 2006). As commonly defined by scholars, motivation can be intrinsic or extrinsic, i.e. based on the user's perception of the task itself as enjoyable, or based on some favourable factors or rewards external to the task, respectively. In Self-Determination Theory, intrinsic motivation is given mostly by three components: Competence, the enjoyable feeling of owning the needed skills to complete the task, Autonomy, the degree of success based on the feeling of being in control of the situation, and Relatedness, which is given by the perception to be involved with other beings. The Hexad framework relies on such concepts, with the addition of meaning (purpose) as an intrinsic motivator (Tondello et al. 2016). The Hexad framework, rather than being based on observed behaviour, is based on the personifications of intrinsic and extrinsic behaviour. The six user types and their respective motivations proposed by Marczewski (2015) are Philanthropist, motivated by purpose, Socializer, motivated by relatedness, Free Spirit, motivated by autonomy, Achievers, motivated by competence, Players, motivated by extrinsic rewards, and Disruptors, motivated by the triggering of change. "Although users are likely to display a principal tendency, in most cases they will also be motivated by all the other types to some degree" (Peterson, Park and Seligman 2005; Tondello et al. 2016, 4).

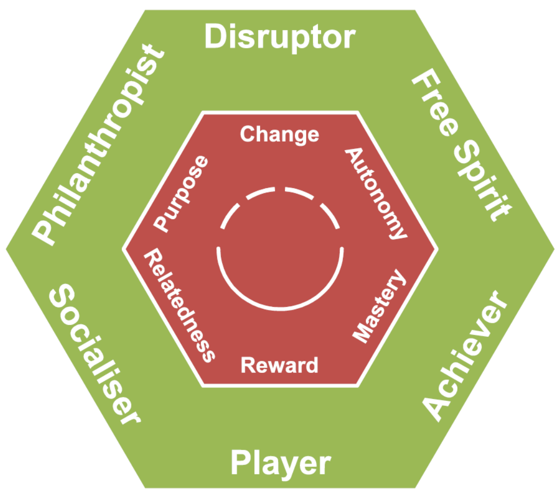


Figure 4: Visual Representation of the Gamification User Type Hexad (Tondello et al. 2016).

Based on Marczewski's framework, Tondello et al. used internal and test-retest reliability analysis, as well as factor analysis to validate and create a new 24-item survey response scale to score users' preferences towards the six different motivations in the Hexad framework. The resulting six-factor scale questionnaire shows considerable reliability in five of the six subscales (the Player subscale being the least reliable, but still very good). The researchers performed also an additional analysis on correlations between the Hexad user types and the personality traits as defined in the five-factor model of personality known as the Big Five, resulting in significant correlations, further validating the Hexad model. Positive correlations were found in Philanthropist and Socializer with extraversion (outgoing personality), Achiever with conscientiousness (thoughtful and organized), and Free Spirit with openness (adventure-seeking and open-to-experience). It would be interesting to find any relationship between players who identify as socializers and philanthropists and the amount of perceived social presence.

Despite the popularity and frequent use that the gamification user type Hexad scale has seen shortly after its creation (Krath et al. 2023), later on, the same authors (Krath et al. 2023) decided to reiterate and develop a shorter version. The motivation for the work was that the length of the original Hexad scale can represent a limiting factor for both academics and practitioners. As explained by the authors, the assessment time needs to be short in order to decrease the dropout rate of participants and reduce the phenomenon of random responding. The duration of the assessment may become too long if participants have to answer multiple questionnaires, and data quality and completion rates tend to decrease significantly with longer questionnaires (Bansak et al. 2018). Finally, a lengthy questionnaire is also not suitable for use in the industry, which prefers rapid and short tools, and is not suitable for mobile, as it may be cumbersome (Krath et al. 2023, 2). For these reasons, the researchers decided to develop a more concise instrument for assessing Hexad user types, which they did by performing an exploratory factor analysis to identify the appropriate items, which they then validated through a confirmatory factor analysis. The final result, the Hexad-12, even outperforms the original survey regarding model fit, reliability, convergent, and discriminant validity. This final instrument is thus the most suitable one for this thesis experiment.

Facial Action Units

In the past years, multiple software tools have been designed and developed to answer the increase of research interest in automatic facial behaviour analysis and understanding. OpenFace is one such tool, developed by Baltrušaitis et al. (2018) for use within the fields of computer vision, machine learning and affective computing research, as well as for the vast array of researchers and developers interested in building interactive applications based on facial behaviour analysis. OpenFace provides a state-of-the-art toolkit capable of facial landmark detection, head-pose estimation, facial action unit recognition, and eye-gaze estimation with available source code for both running and training the models, capable of real-time performance and able to run from a simple webcam without any specialized hardware. The software is not developed for purposes of emotion recognition, but instead to identify the objective interpretative precursor of emotion, i.e. the facial action unit.

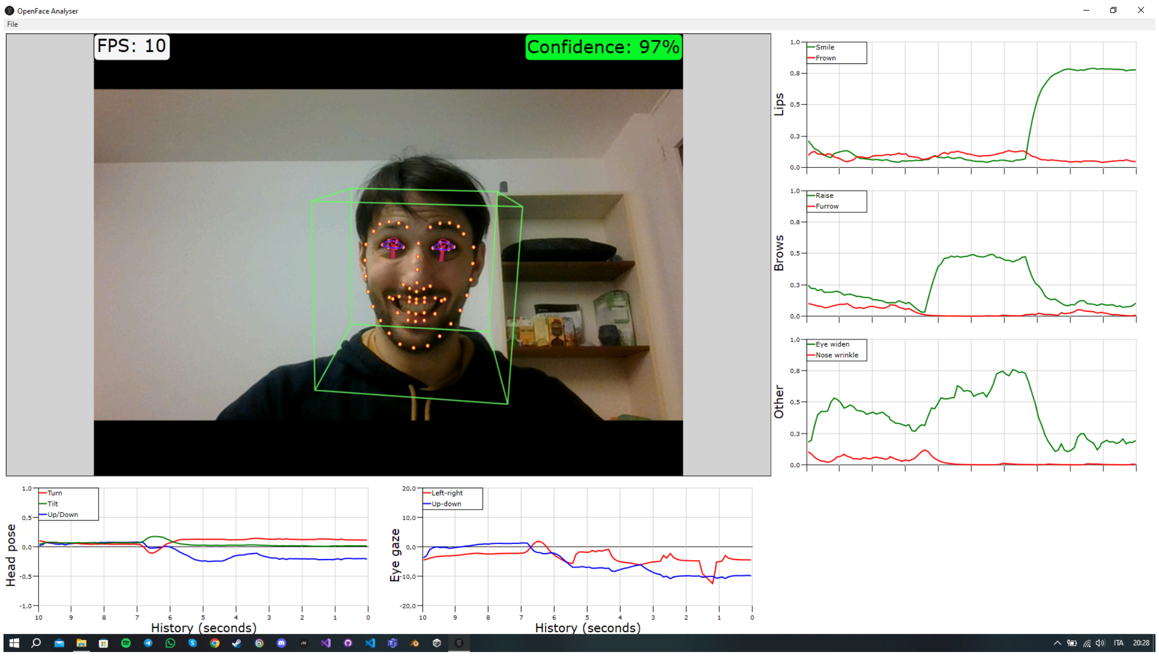


Figure 5: OpenFace Analyzer tool applied to the principal investigator's face for demonstrative purposes. Own screenshot. Software of Baltrušaitis et al. (2018)

The Facial Action Coding System (FACS) is a system to taxonomize human facial movements by their appearance on the face. This system is based on the work of Hjortsjö (1969), adopted and built upon by Ekman and Friesen (1978) and finally updated by Ekman, Friesen and Hager in 2002. The FACS encodes movements of individual facial muscles by capturing subtle, instantaneous changes in facial appearance. Through this system it is possible to code nearly any anatomically possible facial expression, deconstructing it into the specific action units (contractions or relaxations of one or more muscles) and their temporal segments that produced the expression (Freitas-Magalhães 2012). The FACS has become a common standard to objectively describe facial expressions, due to action units being an effective form of measurement independent of any interpretation, but nevertheless very efficient when purposed for emotion recognition. The system consists of an index of all possible facial expressions and the relative muscle contracted or relaxed, excluding the bio-mechanical information about the degree of muscle activation. The main codes of the FACS are:

0	**Neutral face**	
1	Inner brow raiser	frontalis (pars medialis)
2	Outer brow raiser	frontalis (pars lateralis)
4	Brow lowerer	depressor glabellae, depressor supercilii, corrugator supercilii

5	Upper lid raiser	levator palpebrae superioris, superior tarsal muscle
6	Cheek raiser	orbicularis oculi (pars orbitalis)
7	Lid tightener	orbicularis oculi (pars palpebralis)
8	Lips toward each other	orbicularis oris
9	Nose wrinkler	levator labii superioris alaeque nasi
10	Upper lip raiser	levator labii superioris, caput infraorbitalis
11	Nasolabial deepener	zygomaticus minor
12	Lip corner puller	zygomaticus major
13	Sharp lip puller	levator anguli oris (also known as caninus)
14	Dimpler	buccinator
15	Lip corner depressor	depressor anguli oris (also known as triangularis)
16	Lower lip depressor	depressor labii inferioris
17	Chin raiser	mentalis
18	Lip pucker	incisivii labii superioris and incisivii labii inferioris
19	Tongue show	
20	Lip stretcher	risorius with platysma
21	Neck tightener	platysma
22	Lip funneler	orbicularis oris
23	Lip tightener	orbicularis oris
24	Lip pressor	orbicularis oris
25	Lips part	depressor labii inferioris, or relaxation of mentalis or orbicularis oris
26	Jaw drop	masseter; relaxed temporalis and internal pterygoid
27	Mouth stretch	pterygoids, digastric
28	Lip suck	orbicularis oris

Table 1: Main codes of the Facial Action Coding System.

The action unit indexes not included in this table are the ones ranging from 51 to 60 representing head movement codes, from 61 to 69 representing eye movement codes, from 70 to 74 representing visibility codes, and from 98 onwards representing gross behaviour codes. The OpenFace software is able to recognize the following action units:

- **1** - Inner brow raiser
- **2** - Outer brow raiser
- **4** - Brow lowerer
- **5** - Upper lid raiser

6 - Cheek Raiser
7 - Lid tightener
9 - Nose wrinkler
10 - Upper lip raiser
12 - Lip corner puller
14 - Dimpler
15 - Lip corner depressor
17 - Chin raiser
20 - Lip stretcher
23 - Lip tightener
25 - Lips part
26 - Jaw drop
28 - Lip suck
45 - Blink

The OpenFace software identifies the presence of an action unit (if it is visible in the face) and its intensity (from minimal to maximal on a 5-point scale). The software demonstrates proficiency in extracting Action Units from images, image sequences, and videos.

Methods

The Methods section is divided into three subsections describing the decisional process and the technical characteristics of the study. The first subsection analyses in detail the findings of the literature review and applies them to determine the most appropriate methods for the experiment on a theoretical level. The second subsection concretises and approximates the decisions made in the previous section through the tools available, and provides the detailed guidelines to conduct the experiment. Lastly, the third subsection offers a retrospective view of the experiment, with the characteristics of the participants and the difficulties encountered.

Theoretical Model of Experimentation

The goals of this subsection are to extrapolate the relevant findings from the literature and apply them in order to design a rigorous and effective experiment and analysis methodology for this thesis work.

The Ideal Game

The first of the three research questions is "Is it possible to perceive social presence in a multiplayer game experience even with close to no communication?". To design an experiment aimed at answering this research question, it is needed to develop an investigation model where the game played by the participant lacks any form of verbal communication, either in textual, audio or visual form, and where the expected amount of social presence perceived by the player is maximised.

As suggested by the analysis of the academic literature executed by Oh, Bailenson, and Welch (2018), immersion, intended as the medium's technological capacity to generate realistic experiences, emerges as a non-crucial factor in inducing social presence. Social presence is not an inherent quality of the communication medium (Oh, Bailenson, and Welch 2018, 20), although the immersive qualities and the technological features of the medium are expected to have a relevant influence on it (3). Based on this knowledge, designing an experiment where participants interact with the game through head-mounted display and over-hear noise-cancelling headphones would necessarily provide the highest amount of telepresence, and consequently impact positively the amount of perceived social presence. Nonetheless, this positive impact is envisioned to be minimal, and not worth the specificity of such equipment. Making the participants interact with the game through a simple LCD monitor and a gamepad is thus deemed the most fitting solution, even more so considering that such conditions would be closer to the circumstances that inspired the work, i.e. the playtesting of Portal 2. The matter of the impact of immersion on social presence, however, remains a variable to take into consideration, as according to the experiment modalities, the participants need to interact with each other without any form of communication besides the basic mechanics of the game. In the academic literature, communication through computer-mediated communication is well-known for eliciting low levels of social presence when compared to face-to-face communication (Oh, Bailenson, and Welch 2018, 20). While multiple studies provide evidence that interacting through computer-mediated communication for long periods such as months, allows the interactants to adapt to the conditions, thus making the levels of social presence rise, such timeframes would be prohibitive for this experiment and clearly out of the scope of the study. Given the aforementioned conditions, in this master thesis experiment the individuals are envisioned to perceive low levels of social presence, therefore, choosing a correct gameful system capable of boosting this subjective experience seems to be a crucial factor.

Among the features identified by Oh, Bailenson, and Welch (2018) that more strongly impact the perception of social presence in virtual environments, the visual representation of the communication partner stands out for excellence. These results are coherent with the formulation of the three dimensions of social presence: co-presence, psychological involvement and behavioural involvement. Co-presence, originally believed to be the highest reachable stage of social presence intensity, is based on having a strong mental representation of the other person. In a virtual environment, users tend to have avatars, which, apart from providing a visual representation of the person, also comply with the physics and the rules of the system, thus providing a coherent and believable instance of the person in the virtual world. In addition, between photographic, anthropomorphic and behavioural realism, the last one seems to be the approach with the most notable positive effects. Put differently, to maximise the perceived social presence, the avatars don't need to look realistic or human-like, as much as they need to behave like a natural human would, for instance by performing actions like blinking or breathing in their idle animation. With these premises, it was clear that the most suitable game for this study needed to have visual representations of the players in the form of avatars, and that these avatars needed to behave naturally both within the rule systems of the virtual environment and within the natural expectations of the participants. Since photographic and anthropomorphic realism were found to be not relevant factors in inducing social presence, it was deemed safer to avoid the aesthetics altogether, in order to avoid encountering the phenomenon commonly known as Uncanny Valley, which could presumably make both the immersion and the perceived social presence plummet. Therefore the avatars should have stylized or cartoonish aesthetics, anthropomorphic enough to be interpreted as capable of social interaction, but not enough to cause the participant to expect higher levels of realism.

Another feature studied to have a notable impact on the perceived social presence is the perceived agency of the communication partner. The belief that the virtual human is controlled by an actual person rather than by a computer program makes the interactants feel higher levels of social presence, even lowering the unconscious threshold of required realism. Furthermore, as highlighted by multiple studies referenced in this manuscript, playing against a human-controlled avatar rather than a computer-controlled agent elicits higher psychological arousal. In addition, playing against a friend elicits even higher spatial presence, engagement, and self-reported physiological arousal. In the studies on the topic found by Oh, Bailenson, and Welch, the participants were informed of the nature of the opponent, as well as most of the studies found. Put differently, the participants were always primed. Other researchers focused their study on the effects of priming (Weibel et al. 2008), showing that the participants who were informed of being playing against other players were reporting higher levels of immersion, engagement and flow than the participants who were told they were playing against bots, even though both groups were facing players. Similarly, Hudson and Cairns (2014) found players to be sensitive to whom they think they are playing with, attributing more agency to the teammate as well as attempting to communicate more when being told that the virtual human is an avatar. In conclusion, priming has a critical effect on the participant's perceived social presence, but it can possibly be a placebo response. In this thesis experiment, it was chosen to inform the participants of the nature of the playing partner, prioritizing the optimisation of the perceived social presence levels. With the resources of this experiment, it is not possible to have a sample large enough to perform a placebo-controlled trial, but the possibility is nevertheless taken into account.

Some other relevant factors in influencing the perceived social presence are the task typology and the social context. Interactants were found to experience higher levels in conditions where their partner needs to pay close attention to their behaviour, either just as an observation, or with the task to accommodate the said behaviour. These findings point to multiplayer games as a suitable choice, specifically, those games where the player has to pay close attention to the other player, for instance, either in a one-versus-one competitive setting or in a cooperative setting where the task requires high degrees of coordination and synchronization. In addition, it seems that exposing the interactants to social contexts and cues that indicate that they have to interact socially, also leads to heightened levels of social presence. These findings are thus in favour of cooperative games that require coordination, which would lead the participants, even in the absence of traditional communication means, to find a way to communicate, thus perceiving the other as a social being.

Taking into consideration all the reflections made above, all aimed at finding the elements and factors that better heighten the amount of social presence that the participants are supposed to perceive, it is possible to infer the characteristics of the game that is expected to be most suitable for the experiment, as well as the characteristics that the study should have. Starting with the setting:

a. the participants should be put in isolated rooms with no distractions and no possible social interaction outside of the game system,
b. the participants should be informed of when they are playing with a bot and when they are playing with a person, and
c. the participants should know the identity of their playing partner, which possibly should be a person they are in friendly relations with.

As for the game,

1. it doesn't need to be a virtual reality game,
2. it should have avatars that behave and move naturally,
3. it should have a stylized or cartoonish aesthetic,
4. it should be a multiplayer game with a cooperative goal that requires coordination, and
5. it should depict an environment that is commonly considered a social setting.

In addition, some further features are needed in order to facilitate the experiment.

6. The game should be of the arcade genre, which typically consists of short levels, intuitive control schemes, increasing difficulty and the absence of explicit overarching plots. This paradigm makes it so that the game sessions are short and intuitive, and that the players don't focus on anything that is not their and their partner's interaction with the game environment. In this way also all participants are given the same portion of the game to interact with in both sessions.
7. To make sure that the participants are also capable of interacting with the given portion of the game in the same way, the game needs to be entry-level, i.e. it needs to be easy and intuitive enough that interactants with little familiarity with gaming medium don't automatically experience frustration due to the game skill floor being higher than the interactants skill level.
8. Necessarily the multiplayer game will also need to have a mode of the game where players can play in the same conditions of the multiplayer setting, but with bots (computer-controlled agents) instead of other players. Without this feature, the experiment wouldn't be possible at all.
9. Lastly, the game should be affordable. Being this a master thesis experiment, the economic resources are the ones of a master's degree student.

Finding a game that satisfies all these conditions is crucial to better answer the research question "Is it possible to perceive social presence in a multiplayer game experience even with close to no communication?".

A game satisfying all of the above requirements was never found, but few candidates were capable of satisfying all but one. The most interesting candidate was the game

Overcooked! 2 (Team17 and Ghost Town Games 2018). In *Overcooked! 2*, similarly to the other titles of the franchise, players are put in a restaurant setting and given cartoonish avatars to move around the kitchen in order to perform simple tasks such as grabbing ingredients, chopping and/or cooking them, combining them into instructed recipes, and delivering them to the clients. The simple and intuitive gameplay loop makes the game engaging for players of all skill levels, finding its challenges in getting players to cooperate and properly coordinate. In *Overcooked! 2* playing efficiently to the game's goals requires the players to work as an organized team, which is only possible through intense communication and coordination. The approach employed by *Overcooked! 2* to force players to constantly communicate is what makes this candidate so interesting. As elucidated by Brown (2018), in most multiplayer games communication is not truly required. This is due to typical multiplayer structures being symmetrical, i.e. requiring the players to interact symmetrically with the game system, due to the players having the same goals, the same mechanics and the same information. In such a system, players are only needed to interact with the system individually, infrequently calling out for each other assistance. Asymmetrical multiplayer games instead provide the players with different gameplay mechanics or different information, in this way forcing them to cooperate, coordinate and communicate towards the completion of the game goal.

Figure 6: Two players collaborating in a multiplayer gameplay of Overcooked! 2 (Team17 and Ghost Town Games 2018), level 1-3. Own screenshot of own gameplay.

In *Overcooked! 2* the avatars all interact with the game environment with the same mechanics, and all are provided with the same information. Each player is fully capable of performing the full gameplay loop on their own. Given these assumptions, the

Overcooked! 2 multiplayer experience would be symmetrical, with players focusing on different dishes but still performing an almost identical gameplay loop. However, the game's levels are designed with the intention of making it incredibly tedious and inefficient to perform a gameplay loop individually, as navigating the environments to the required points would force the player to traverse multiple times the same portion of the map instead of focusing on the smaller tasks. Coordination between the players however allows to speed up the process considerably, thus encouraging the players to focus on distinct subsections of the game loop, taking control of specific parts of the game map and ending up naturally creating asymmetrical roles. One player gets the ingredients and cuts them, and the other one cooks them and delivers the dishes. This level design creates a surprising need for communication, typically done at the start of the stage, where players immediately split tasks and ponder the most efficient way to prepare the meals. If the design of the game stopped here, *Overcooked! 2* would suffer the same flaw as most asymmetric cooperative multiplayer games: once players have clear roles the gameplay loop falls into predictable patterns where each player familiarizes with their tasks, sticks to them, and communication is no longer required. The game prevents that as multiple events and game mechanics constantly disrupt these comfortable patterns, by forcing players to constantly switch roles and come up with new ones. When the players start getting comfortable with their roles the level design of the map changes entirely, frequently forcing them to change tasks entirely. Consequently, *Overcooked! 2* excels in constantly providing new challenges that make players keep up the communication in order to coordinate better approaches to the game loop.

As discussed extensively in this manuscript, social presence is a subjective experience strongly tied to the perception of the other as a social entity. The perceived necessity of establishing communication with the other being is believed to have a reasonably meaningful impact on social presence. Therefore, a multiplayer cooperative game experience that forces the player to constantly communicate, in conditions where the interactants are not provided with any traditional means of communication, would force them to find a way to understand the playing partner's intentions by their behaviour alone. It could be possible that such conditions would cause the players to individuate communication means other than verbal communication, affecting accordingly the dimensions of psychological involvement and behavioural involvement. Unfavourably, *Overcooked! 2* lacks one crucial requirement listed above, it doesn't have bots. Strenuous and repetitive attempts were made to develop a digital game modification for the game that would add the missing feature, but due to a lack of experience with such practices, it was unsuccessful. *Overcooked! 2* is still believed to be the most interesting candidate, but due to schedule pressures, it was chosen to proceed with the second-best candidate.

The second-best candidate was considered to be the game *Boomerang Fu* (Cranky Watermelon 2020). *Boomerang Fu* is an arcade local multiplayer competitive game disposing of naturally moving and behaving stylised cartoonish avatars, very simple control schemes, simple game mechanics and an identical game structure between the multiplayer mode and the singleplayer with bots mode. The game only has two downsides: it is competitive, thus not the preferred game goal structure to maximise psychological involvement, and it lacks the capability to play the multiplayer mode through internet connection. The latter problem was obviated by utilizing the Steam Remote Play functions.

Figure 7: A player defeating the other in a multiplayer gameplay of Boomerang Fu (Cranky Watermelon 2020). Own screenshot of own gameplay.

Boomerang Fu is a game with a competitive goal structure, thus it is believed to elicit lower levels of psychological involvement compared to games with cooperative goal structures. Nevertheless, *Boomerang Fu* is another excellent candidate for the experiment, as it fulfils excellently all the other previously mentioned requirements. In addition, the gameplay is identical in playing either with players or players and bots, thus minimizing the variables between the two scenarios. Boomerang Fu is therefore chosen as the game for the experiment.

The Ideal Methodologies

The three research questions of this study concern the measurement of physiological or psychological phenomena in the participants. To answer the first research question it is necessary to measure the perceived social presence in the participant, to answer the second one it is needed to measure the physiological emotional behaviour identified as facial action units, and to measure the third it is needed to measure the perceived sense of flow.

Among the vast variety of different methods commonly employed in game experience research, the most operated techniques are forms of physiological measurements, game metrics, observations and self-reports. Physiological measurements provide an objective tool for the analysis but tend to be costly and hard to interpret. Game metrics are directly logged by the game system and provide objective quantitative measurements of the player's interactions with the system, however, it is only possible to utilize this technique when being the owner and developer of the employed software. Observations of in-game behaviour supply complementary material for the qualitative analyses, but their study still doesn't have clear agreed guidelines in the academic literature, and their analysis and interpretation is a non-negligible time-consuming task. Self-reports represent an easy, inexpensive and efficient tool to measure subjective experiences, but their accuracy relies on variable psychological processes such as memory, attention and introspective ability, in addition, they are subjective by nature and mostly done retrospectively, hence sensitive to bias and social desirability. All of these different methods have their strengths and weaknesses. Researchers individuated as best practice the combination of multiple heterogeneous tools in order to cover the weaknesses and better capture the multifaceted and complex phenomena that is game experience. This master thesis experiment aims to do the same, leaving out only the game metrics technique, since the experiment will use a third-party game software.

Physiological measurements do not directly disclose subjective experiences; nevertheless, they can be employed to deduce the underlying processes of such. As a result, they serve as a quantitative indicator for phenomena such as emotion, attention, motivation, attitudes, and unconscious cognitive processes. Among the physiological measurement tools studied in the academic literature for their capability of predicting subjective experiences such as social presence and flow, only electroencephalography emerged successful. Such equipment is out of the scope of the study, as well as out of the capabilities of the researchers. Furthermore, it is one of the goals of this study to assess the relationship between emotional facial behaviour and the previously mentioned subjective experiences. As such the only physiological tool operated will be focused on the measurement of emotional facial behaviour, better conceptualized as facial action units. As previously defined, facial action units are fundamental actions of one specific muscle or a group of specific muscles in the face, identified through indexed code numbers in the taxonomy of human facial movements known as the Facial Action Coding System. The FACS facilitates the encoding of virtually any anatomically plausible facial expression, breaking it down into its specific action units. This system has emerged as a widely adopted standard for the objective depiction of facial expressions due to the clarity inherent in action units as a form of measurement, which is independent of subjective interpretation. While the measurements of facial action units do not represent a form of emotion recognition per se, the action units can be effectively utilized for emotion recognition. Nonetheless, this is not a concern of the study, as the automatic identification of the emotion transpiring on the interactant's face is not the aim of this research. This study aims to find a relationship between the perceived social presence and the frequency and intensity of facial action units during the game session. What emotion is connotated by the engaged muscle groups is out of the scope.

To measure the facial action units, the participants will be recorded through the embedded webcam of the laptop computer devices on which they will interact with the gameful system. The software operated for the recording is the camera application provided by the base Windows 10 operating system functionalities. The recordings of the participants' faces will then be processed through the previously mentioned state-of-the-art software OpenFace, which provides for every frame of the recording two numbers for each facial action unit, one identifying the presence of an action unit and one identifying the intensity. The data will be provided for the following action units:

- Inner brow raiser
- Outer brow raiser
- Brow lowerer
- Upper lid raiser
- Cheek raiser
- Lid tightener
- Nose wrinkler
- Upper lip raiser
- Lip corner puller
- Dimpler
- Lip corner depressor
- Chin raiser
- Lip stretcher
- Lip tightener
- Lips part
- Jaw drop
- Lip suck
- Blink

Statistical analysis will be performed to identify how these physiological measurements vary across the two game sessions.

The observations of in-game behaviour will be performed on the recordings obtained through screen capture of the laptop computer devices on which the participants will interact with the gameful system. The software operated for the recording of the in-game behaviour is the screen capture function provided by the Game Bar overlay, accessible by pressing Windows + G on any device with the Desktop Windows 10 operating system.

In terms of self-reports, a judicious selection is needed. The Game Experience Questionnaire (GEQ) has emerged as the most widely employed instrument in game user research. Its popularity can be attributed to its multidimensional item structure, its simplicity, and its widespread availability. However, critiques have arisen regarding the instability of the proposed seven-factor scale structure in its core part. As a response, other researchers have proposed a more refined five-factor redesign named GEQ-R, optimizing and re-proposing the Competence, Immersion, Flow, Negativity, and Positive Affect scales of the original questionnaire. In the present master's thesis experiment, the Flow and Immersion scales from the GEQ-R will be employed. Addressing the third research question, subjective flow measurements will be gathered to explore the correlation between perceived flow and the sense of social presence among experiment participants. Immersion self-report values will be examined for potential relationships with social presence, to exacerbate the difference between this work and previous studies conducted under similar experimental conditions, in which spatial presence was assessed as a combination of immersion and telepresence. The collection of these data is also believed to enhance comprehension of the social presence phenomenon. Additionally, alongside the core component of the GEQ, this study incorporates its additional module on social presence, the Social Presence in Gaming Questionnaire (SPGQ), with its stable three-component structure: Psychological Involvement (Empathy), Psychological Involvement (Negative Feelings), and Behavioural Involvement. The SPGQ will be administered after each session to analyze how values evolve between the two sessions. Furthermore, considering the literature review's outlined justifications, participants will complete the six-factor Hexad user-type questionnaire. This tool aids researchers in classifying participants within the taxonomy of six user types (Philanthropist, Socializers, Free Spirit, Achievers, Players, and Disruptors), revealing their primary intrinsic motivators, and thus a possible predisposition for social interaction. Specifically, participants will complete the shorter Hexad-12 version before the initial game session, mitigating concerns related to questionnaire length.

Applied Model of Experimentation

In the applied model of experimentation, the participants are to be gathered in a common room while the researchers periodically come to pick up two participants at the time for the experiment. It is to be clearly shown to each participant who is being picked up for the experiment so that each participant knows who they will be playing against in the second session. The schedule of the experimentation must be planned in such a way that as many participants as possible play against individuals with whom they are in friendly relations.

The selected participants are to be seated in two nearly identical rooms, with all sources of distraction or social contact minimized. The participants are to be read the following text:

> Welcome to Michele Fanelli's thesis experiment. Your task as a volunteering participant is to play two digital play sessions.
>
> In the first session, you will play in single-player with a bot, while in the second session, you will play in multiplayer with the person who is now sitting in the other room. There will be no form of textual, audio or face-to-face communication with the person in the other room. During both play sessions, you will be left in the room alone until the completion. At the end of each session, you're supposed to come to the door to notify the end, and I will provide you with a post-game questionnaire to compile.
>
> Your play session will be recorded through screen capture, and your facial behaviour will be recorded through the embedded webcam. The recordings of the facial behaviour will be processed through the OpenFace software, which allows the extrapolation of the numerical values of the Facial Action Units (face movements such as smiling, frowning, raising the eyebrows and narrowing the nostrils). The questionnaire results, the gameplay recordings and the numerical data of the Facial Action Units will be processed and studied in order to answer the following research questions:
>
> 1. Is it possible to perceive Social Presence in a multiplayer game experience even with close to no communication?
> 2. Is there a relationship between perceived Social Presence and a more conspicuous facial emotional behaviour?
> 3. Is it possible for a player to perceive both Social Presence and a sense of Flow?
>
> All of the acquired data will be stored securely on Michele Fanelli's personal computer. The recordings of the in-game behaviours, the self-reports and the numerical data processed out of the facial recordings, will all be anonymized through code number labelling. The anonymized numerical data may be published in the final academic work. The numerical data are such that identification of the person through them is impossible.

Following the description of the experiments, the participants are to be informed of their right to ask any question they desire, as the researchers are required to provide all the requested elucidations. After the short question and answer session, the participants are to be provided with the informed consent form for the use of their personal data, which they are supposed to read thoroughly and carefully and to be reminded of their right to ask any question they desire. The full informed consent form is provided at the appendix of this manuscript (Appendix I).

Following the completion of the introductory phase, participants are to be seated in the pre-established position and the camera for recording facial movements must be adjusted to centre the participant's face. The game application is to be run. The participants are to be provided with the Hexad-12 questionnaire to compile on-site and to be left alone during the filing to avoid answering alterations due to social desirability. The mentioned questionnaire appears as a simple white sheet with the “Questionnaire - Hexad 12” title, a space designated for the filing of the participant's anonymized code, and the twelve questions arranged in a tabular format, employing a double-entry table. The questions are listed in the column, while the potential answers are presented in the row, utilizing a seven-point Likert scale (Strongly Disagree, Disagree, Somewhat Disagree, Neither, Somewhat Agree, Agree, Strongly Agree) that can be selected by marking the corresponding circle with a pen. The questions are:

- It makes me happy if I am able to help others
- The well-being of others is important to me
- I like being part of a team
- I enjoy group activities
- I like mastering difficult tasks
- I enjoy emerging victorious out of difficult circumstances
- Rewards are a great way to motivate me
- If the reward is sufficient, I will put in the effort
- It is important to me to follow my own path
- Being independent is important to me
- I see myself as a rebel
- I dislike following rules

Previous to the experiment corpus the participants are explained the controls of the game. The game session is prepared by the researchers by allowing the participant to choose the preferred character and then, according to the game session, either by selecting the game mode with the bot or by completing the necessary setup for the multiplayer connection. The participant is to be informed of which game session they are playing, this being informed of the nature of the opponent: computer-controlled agent in the first session and player-controlled avatar in the second session. Upon completion of the overall setup, the webcam and the screen capture recording software are both activated, and the participant is instructed to start playing while the researchers leave the room.

Following the notification from the participant that the game session is over, the researchers are to return to the room and administer to the participant the tailored questionnaire measuring immersion, flow and social presence. Similarly to the previous questionnaire, the Immersion, Flow and Social Presence Questionnaire appears as a simple sheet with a title, a space for the participant's anonymized code, and a double-entry table divided on both sides of the sheet, with the immersion and flow scales questions on one side, and the three social presence scales questions on the other. Unlike the previous questionnaire, this one makes use of a five-point Likert scale (Not at all, Slightly, Moderately, Fairly, Extremely), and has a space designated to the specification of which game session is being measured (1 or 2). On top of the table there is written the instruction: “Please indicate how you felt while playing for each of the items. Some of the questions may seem out of place, please answer as best you can.” The questions in order are:

- I was interested in the game’s story
- I was fully occupied with the game
- I thought about other things
- I found it tiresome
- I forgot everything around me
- I lost track of time
- I was deeply concentrated in the game
- It felt like a rich experience
- I lost connection with the outside world
- I empathized with the other
- My actions depended on the other’s actions
- The other's actions were dependent on my actions
- I felt connected to the other
- The other paid close attention to me
- I paid close attention to the other
- I felt jealous about the other
- I found it enjoyable to be with the other
- When I was happy, the other was happy
- When the other was happy, I was happy
- I influenced the mood of the other
- I was influenced by the other moods

- I admired the other
- What the other did affected what I did
- What I did affected what the other did
- I felt revengeful
- I felt schadenfreude (malicious delight)

The steps for the preparation of the first session and the administration of the questionnaire afterwards are repeated very similarly in the second session. While preparing for the second session, the participant is to be reminded who they are playing against. Upon competition, the participant is to be thanked and led back to the common room.

Characteristics and Limitations of the Study

The experiment took place in two different sites. The first was the Nautilusweg Haus B student dormitory in front of the Alpen-Adria-Universität Klagenfurt, the current domicile of the principal investigator, while the second was in Udine, at the residence of the principal investigator. In both settings, the participants were gathered in the kitchen area, where they were offered refreshments and tabletop games to ensure joviality and satisfaction. Every approximately forty-five minutes the main investigator entered the kitchen area to call two participants for the experiment and lead them to the designated two isolated rooms.

In the first setting, the participants were peers of the principal investigation, for the vast majority students of the international master's degree course of Game Studies and Engineering, offered by Alpen-Adria-Universität Klagenfurt. Consequently to the linking ring, the vast majority of participants in the first setting have a considerable amount of experience with digital games, and as such the simple game controls wouldn't represent a challenge or a hindrance to the game experience, thus not affecting the perceived immersion. Fourteen participants were gathered for the occasion, for a total of seven iterations of the experiment distributed over roughly six hours. The mathematical average age of the participants is 24.7, the median is 24 and the variance is 3.6, all rounded to the first significant digit.

In the second setting, the participants were peers of the principal investigator, mostly connected by proximity friendship relations, all living close to the city of Udine. More than half of the participants have a significant amount of experience with digital games, on which consequently apply the same considerations over the game controls not hindering the immersion. Ten participants were gathered for the occasion, for a total of five iterations of the experiment distributed roughly over four hours. The mathematical average age of the participants is 25.7, the median is 26 and the variance is 1.6, all rounded to the first significant digit.

Additionally, two more couples of participants were experimented upon in other two different sittings. Both were in the same physical place and conditions of the first setting. All participants had a sufficient amount of experience with digital games to not find impediments in the simple game controls. For the first couple, the mathematical average age of the participants is 20.5 with a variance of 0.5. For the second couple, the mathematical average age of the participants is 26.5 with a variance of 4.5.

In total, 28 people participated in the experiment, with a mathematical average age of the participants being 24.9, a median is 25 and a variance is 4.3, all rounded to the first significant digit.

In both settings, the experiment rooms were carefully prepared so that no distraction could influence the participant. The laptops and the wired DualShock 4 gamepads were centred on empty tables positioned against walls. In the first setting the walls contained an empty whiteboard except for a simple and minimal diagram remembering the controls of the game in case the player was to forget them (“X to Dash, Square to Slash, Triangle to Throw”). In the second setting, the whiteboards were unavailable, so the diagram was printed on single white A4 sheets on the laptops' right side. Through careful planning, the tables were positioned in a way to optimize the lighting so that the participants’ faces would be uniformly lit. In the first setting the tables were positioned below wide diffused light lamps, in the second setting multiple lamps were positioned facing a white wall so that the lighting would be as uniform as possible. In both settings, the participants were made to sit on the chair and then the laptop inclination would be adjusted so that their faces would be centered in the recording.

As previously mentioned the experiment made use of Steam Remote Play, in order to make accessible through an internet connection a game that otherwise would not provide such functions. In the first setting, the participants were connected to the dormitory Wi-Fi, while in the second setting, a custom system was designed in an attempt to minimize connection-related issues. The principal investigator programmed a subnet of his home Wi-Fi router, to have a portion of the device power dedicated to the experiment and a considerable amount of possible IP addresses available only to the experiment devices. Furthermore, the laptops were given static IP addresses and were connected through 10-meter long and 15-meter long Ethernet cables to a custom ADSL Gateway device which was connected through a 20-meter long Ethernet cable to the Wi-Fi router, which was positioned on a different floor of the building (reason why, without the preparation of such system, the router Wi-Fi connection would not have reached the devices).

As discussed in the Ideal Methodologies subsection, the recordings of the facial emotional behaviour were performed through the use of the Windows 10 embed application Camera. The recordings of the in-game behaviour were acquired through the Windows 10 screen capture function accessible through the Game Bar overlay by pressing the Windows + G combination. In order to facilitate the synchronization of the two recordings during the data preparation phase, a simple manual procedure was enacted right after the initiation of both recording functions and before handing the gamepad to the participants. Within *Boomerang Fu*'s opening screen, the principal investigator and the instructed assistant pronounced out loud “Up, down, up, down” in the precise instant they would press the respective buttons of the DualShock 4 D-pad. During the data preparation phase, each couple of facial behaviour and in-game recordings were uploaded into the video editing software CapCut (ByteDance 2020). Within this software, the two videos were appropriately synchronized using the combination of the visual and auditory signal of the up-down technique and then cut so that both recordings would only display the interval of time of the active gameplay. More specifically, the starting frame was chosen to be the instant the player presses to close the instruction window of *Boomerang Fu*’s controls, and the ending time frame was chosen to be the instant the text of which player won the match touches the top part of the screen, few seconds after one player wins the last match. Out of each couple of recordings, a couple would always be obtained: a recording of the participant’s facial emotional behaviour only within the active gameplay time interval, and a recording of the in-game behaviour within the active gameplay interval with an overlay in the bottom-right corner of the screen of the synchronized facial emotional behaviour. The first extracted video was meant to be studied quantitively through the OpenFace software and the second to be studied qualitatively through the observation techniques.

The OpenFace software was executed on the facial behaviour recordings by Windows PowerShell command, with the appropriate modifiers so that it would process a file in video format and would only extrapolate the facial action units, as instructed on the OpenFace Wiki GitHub page. For each recording, the software processing lasted between 10 and 15 minutes, which considering the the total amount of 52 videos, made so that the software processing phase lasted roughly twelve hours. The output was for each recording a heavily sized CSV file, having 40 columns and tens of thousands of rows. These CSV files were processed through the most basic functions of Python’s pandas library, and the extrapolated data are the focus of the facial action unit results analysis.

Overall, the model of the experimentation was applied to the best of the capabilities, but due to amatorial means of experimentation, multiple difficulties were encountered. First, being the experiment not conducted in a professional research laboratory, the building of choice had few difficulties. The first site, a student dormitory, presented limitations such as compact room sizes, thin walls, and a consistently noisy hallway due to frequent foot traffic. Consequently, participants may have been susceptible to external disturbances, potentially compromising their sense of immersion and social presence by feeling proximal to individuals outside the room. Moreover, the use of Remote Steam Play, an inefficient streaming gaming tool necessitating a robust internet connection, posed a significant constraint. The dormitory's network infrastructure, burdened by a high volume of users, suffered frequently from congestion, resulting in pronounced lag during gameplay sessions. This lag is believed to be the study's major limitation, since it is believed it affected the participants, undermining their immersion and psychological involvement. Even at the second site, in which the dedicated subnet with the custom access point connected through Ethernet connections ended up having occasional instances of lag, albeit negligible. Lastly, due to human error, some of the recordings of two participants were lost, making it necessary to discard the two participants' data in the facial action units analysis. In addition, for two other participants, the facial behaviour recording did not contain the audio, making it more difficult to properly synchronize the recording in the data preparation phase. This difficulty was obviated by synchronizing the visual signal on the in-game behaviour recording to the visual signal in the facial behaviour recording of the thumb pressing the button.

Results

The following section exhibits and analyses the results of the experiment, namely the self-report data, the facial action unit data and the in-game behaviour observations. All the raw data results can be found as CSV files at the project's GitHub Repository, and accessed for demonstrative purposes at the project's Google Colab Notebook.

The Self-Reports Data Results

The following section presents the results of the quantitative analysis conducted through the self-report methodology examining the subjective experiences of player typology, immersion, flow and social presence. Utilising a 7-point Likert scale, the participants provided their introspective relatedness to the different intrinsic and extrinsic motivations of player types. Utilising a 5-point Likert scale, participants provided responses reflecting their personal perceptions elicited by the experiment the subjective experiences of immersion, flow and social presence, the latter composed of the three scales of empathy, negative feelings and behavioural involvement. Following are presented all the statistical variables used in the self-reports analysis. The code labelling of the questionnaire items was chosen to be kept as consistent as possible with the code labels provided by the authors of the Hexad-12 and the GEQ-R, with few variations based on these experiment conditions to improve clarity and avoid ambiguity.

Code label		**Questionnaire item**
p1		It makes me happy if I am able to help others
p2		The well-being of others is important to me
s1		I like being part of a team
s2		I enjoy group activities
a1		I like mastering difficult tasks
a2		I enjoy emerging victorious out of difficult circumstances
r1		Rewards are a great way to motivate me
r2		If the reward is sufficient, I will put in the effort
m1		It is important to me to follow my own path
m2		Being independent is important to me
d1		I see myself as a rebel
d2		I dislike following rules

i1b	i1f	I was interested in the game’s story
f1b	f1f	I was fully occupied with the game
i2b	i2f	I thought about other things
i3b	i3f	I found it tiresome
f2b	f2f	I forgot everything around me
f3b	f3f	I lost track of time
f4b	f4f	I was deeply concentrated in the game
i4b	i4f	It felt like a rich experience
f5b	f5f	I lost connection with the outside world
e1b	e1f	I empathized with the other
b1b	b1f	My actions depended on the other’s actions
b2b	b2f	The other's actions were dependent on my actions
e2b	e2f	I felt connected to the other
b3b	b3f	The other paid close attention to me
b4b	b4f	I paid close attention to the other
n1b	n1f	I felt jealous about the other
e3b	e3f	I found it enjoyable to be with the other
e4b	e4f	When I was happy, the other was happy
e5b	e5f	When the other was happy, I was happy
n2b	n2f	I influenced the mood of the other
n3b	n3f	I was influenced by the other moods
e6b	e6f	I admired the other
b5b	b5f	What the other did affected what I did
b6b	b6f	What I did affected what the other did
n4b	n4f	I felt revengeful
n5b	n5f	I felt schadenfreude (malicious delight)

Table 2: Full list of items composing the self-report tools in line with the respective identification code label.

In the code labelling of the Hexad-12 questionnaire items, the letter “p” groups the items of the Philatropist scale, the letter “s” groups the items of the Socializer scale, the letter “a” groupes the items of the Achievers scale, the letter “r” groupes the items of the “Player” scale, the letter “m” groupes the items of the Free spirit scale, and the letter “d” groupes the items of the Disruptor scale. The Free spirit is the only scale on which the indicator letter has been changed from the original one, done to avoid ambiguity with the factor label of the Flow scale. In the code labelling of the Revised Game Experience Questionnaire items, the letter “i” groupes the items of the Immersion scale, the leftmost letter “f” groups the items of the Flow scale, the letter “e” groups the items of the Psychological Involvement - Empathy scale, the letter “n” groups the items of the Psychological Involvement - Negative Feelings, and the leftmost letter “b” groups the items of the Behavioural Involvement. Additionally, to differentiate between the identical instances of the GEQ-R scales administrated to the participants in the two distinct game sessions, extra letters were added on the label’s rightmost position: “b” for a game session played with a bot (computer-controlled virtual agent) and “f” for a game session played with a friend (human-controlled avatar). The following subsections encompass a comprehensive statistical analysis aimed at elucidating patterns, relationships and significance of the collected data. Statistical analysis including descriptive statistics, reliability analysis, correlation analysis, and regression analysis are employed to explore the central tendencies, internal consistency, associations, and predictive relationships within the measurements. The complete unmodified collected data of the 28 participants is displayed in the appendix of this manuscript (Appendix II).

Descriptive Statistics

The first basic step is to describe comprehensively the collected data with a smaller set of values, employing measurements of central tendency and dispersion. Following are the calculated values using Google Sheet embed functions AVERAGE, MEDIAN, MODE, MAX, MIN, QUARTILE, VARA, and STDEV. Note that in case of an even number of values, MEDIAN interpolates between the two centre values. Note that with multiple candidates being the most occurring numbers in the set, MODE returns the lower one. Note that VARA calculates the unbiased sample variance, thus dividing by n-1 instead of n.

Label	Mean	Median	Mode	Range	IQR	Variance	Deviation
p1	6.3	6.5	7.0	2.0	1.0	0.7	0.8
p2	6.2	6.0	6.0	2.0	1.0	0.5	0.7
s1	5.3	5.0	5.0	4.0	1.0	1.1	1.0
s2	5.1	5.0	5.0	5.0	1.0	1.3	1.1
a1	5.6	6.0	6.0	5.0	1.0	1.4	1.2
a2	6.1	6.0	7.0	3.0	1.0	0.8	0.9
r1	5.2	5.0	5.0	5.0	1.0	2.0	1.4
r2	5.1	5.5	6.0	4.0	2.0	1.8	1.4
m1	5.2	5.0	4.0	5.0	2.0	2.0	1.4
m2	5.6	6.0	6.0	5.0	2.0	2.0	1.4

d1	3.1	3.0	2.0	4.0	2.0	1.5	1.2
d2	2.9	2.5	2.0	5.0	2.0	1.8	1.4
i1b	2.0	2.0	1.0	4.0	1.3	1.5	1.2
f1b	4.5	5.0	5.0	2.0	1.0	0.4	0.6
i2b	1.9	1.0	1.0	4.0	1.3	1.4	1.2
i3b	1.5	1.0	1.0	3.0	1.0	0.8	0.9
f2b	3.1	3.0	2.0	3.0	2.0	1.2	1.1
f3b	3.1	3.0	2.0	4.0	2.0	1.9	1.4
f4b	4.2	4.5	5.0	3.0	1.0	0.9	1.0
i4b	3.0	3.0	3.0	4.0	2.0	1.3	1.2
f5b	2.8	3.0	3.0	4.0	2.0	1.3	1.2
e1b	1.3	1.0	1.0	1.0	0.3	0.2	0.4
b1b	3.6	4.0	5.0	4.0	2.0	1.9	1.4
b2b	2.8	3.0	3.0	4.0	1.3	1.3	1.1
e2b	1.5	1.0	1.0	3.0	1.0	0.6	0.8
b3b	2.1	2.0	1.0	4.0	2.0	1.4	1.2
b4b	4.0	4.0	4.0	4.0	1.3	1.0	1.0
n1b	2.0	1.0	1.0	4.0	2.0	1.6	1.3
e3b	2.5	2.0	1.0	4.0	3.0	1.8	1.3
e4b	1.1	1.0	1.0	3.0	0.0	0.3	0.6
e5b	1.2	1.0	1.0	3.0	0.0	0.4	0.6
n2b	1.5	1.0	1.0	4.0	0.3	1.1	1.0
n3b	1.7	1.0	1.0	4.0	1.0	1.5	1.2
e6b	1.8	1.0	1.0	3.0	1.3	1.1	1.1
b5b	3.4	4.0	4.0	4.0	2.0	1.5	1.2
b6b	2.5	2.0	2.0	4.0	1.3	1.5	1.2
n4b	3.6	4.0	4.0	4.0	1.0	1.4	1.2
n5b	3.6	4.0	5.0	4.0	2.0	1.9	1.4
i1f	2.1	2.0	1.0	4.0	2.0	1.7	1.3
f1f	4.8	5.0	5.0	1.0	0.0	0.2	0.4
i2f	1.7	1.0	1.0	4.0	1.0	1.0	1.0
i3f	1.6	1.0	1.0	4.0	0.3	1.4	1.2
f2f	3.8	4.0	4.0	3.0	2.0	1.3	1.1

f3f	3.6	4.0	4.0	4.0	1.3	1.5	1.2
f4f	4.6	5.0	5.0	2.0	1.0	0.3	0.6
i4f	3.8	4.0	4.0	4.0	1.3	1.2	1.1
f5f	3.4	4.0	4.0	4.0	1.3	1.4	1.2
e1f	3.6	4.0	4.0	4.0	1.0	1.1	1.1
b1f	4.0	4.5	5.0	4.0	1.3	1.6	1.3
b2f	3.7	4.0	4.0	4.0	1.3	1.1	1.0
e2f	3.8	4.0	4.0	4.0	0.3	0.8	0.9
b3f	3.9	4.0	4.0	3.0	0.5	0.9	1.0
b4f	4.3	4.0	4.0	4.0	1.0	0.8	0.9
n1f	2.0	1.5	1.0	4.0	2.0	1.6	1.2
e3f	4.2	4.0	4.0	4.0	1.0	0.8	0.9
e4f	2.4	2.5	1.0	3.0	2.0	1.3	1.1
e5f	2.7	3.0	3.0	4.0	2.3	1.6	1.3
n2f	3.4	3.5	4.0	4.0	1.0	1.1	1.1
n3f	2.9	3.0	2.0	4.0	2.0	1.5	1.2
e6f	3.1	3.5	4.0	4.0	2.0	1.6	1.3
b5f	4.1	4.0	4.0	4.0	1.0	1.1	1.0
b6f	3.9	4.0	4.0	3.0	1.3	1.0	1.0
n4f	3.0	3.0	4.0	4.0	2.0	2.0	1.4
n5f	3.6	3.5	5.0	4.0	2.0	1.6	1.3

Table 3: The computed descriptive statistics (Mean, Median, Mode, Range, Interquartile Range, Variance and Deviation) of the self-report data results.

While the presented data is useful for calculations later on, it may be more convenient for the readers to compress it into a more accessible format. According to the guidelines provided by the authors of the two utilized self-report tools, it is possible to obtain the value of the scale scores by calculating the arithmetic mean of the items within the scale. The following table (*Table 4*) presents thus the so computed scale scores. Note that scores of items i2 and i3 were reversed before calculating the mean as the items are formulated as the negation on the examined latent variable.

Scale	Mean	Median	Mode	Range	IQR	Vara	Dev
Philanthropist	6.3	6.3	7.0	2.0	1.5	0.5	0.7
Socializer	5.2	5.3	5.5	4.0	1.5	1.0	1.0

Achiever	5.8	5.5	5.5	3.0	1.0	0.7	0.8
Player	5.2	5.5	6.0	4.5	2.0	1.6	1.2
Free Spirit	5.4	5.5	6.0	5.0	1.6	1.7	1.3
Disruptor	3.0	2.8	2.0	4.5	2.0	1.3	1.1
Immersion with Bot	3.4	3.5	4.0	2.8	1.3	0.5	0.7
Flow with Bot	3.5	3.7	3.0	3.0	1.1	0.6	0.8
Empathy with Bot	1.6	1.5	1.3	1.3	0.7	0.2	0.4
Negative Feelings with Bot	2.5	2.5	2.6	3.4	0.9	0.5	0.7
Behavioural Involvement with Bot	3.1	2.9	2.8	4.0	0.9	0.8	0.9
Immersion with Friend	3.6	3.8	4.3	3.3	1.0	0.7	0.8
Flow with Friend	4.0	4.1	3.4	2.4	0.9	0.4	0.6
Empathy with Friend	3.3	3.3	3.0	3.7	0.9	0.6	0.8
Negative Feelings with Friend	3.0	3.1	3.4	2.8	1.1	0.5	0.7
Behavioural Involvement with Friend	4.0	4.0	4.0	2.5	1.1	0.5	0.7

Table 4: The computed descriptive statistics (Mean, Median, Mode, Range, Interquartile Range, Variance and Deviation) of the scales of the self-report data results, calculated as the average of the items composing each scale.

Glancing at *Table 4* emerges how the arithmetic mean and the median are very close values for all the scales, which is a preliminary indicator of the absence of notable outliers within the sample. The measurements of distribution seem also to point to a moderate distribution at most.

It seems that most participants consider themselves closer to the philanthropist player type (i.e. motivated by purpose), as it seems to be the scale with the overall higher scores, and with a notably small measured dispersion. Contrary, very few participants consider themselves close to the disruptor player type (i.e. motivated by the triggering of change), even though there is a medium dispersion and a notable range. The other four player types seem to cluster between “Somewhat Agree” and “Agree”, with different degrees of dispersion, as observable in the table.

On average, participants in both game sessions seem to have felt between "Moderately" and "Fairly" immersed in the experience and in the flow state, with fairly low dispersion but notable outliers. It seems like that hardly any participant felt empathy for the computer-controlled agent, almost unanimously. The empathy is expected notably higher in the session with the friend, with very small dispersion but some outliers. The negative feelings and the behavioural involvement are both one point higher in the game session with the human-controlled avatar, with fewer outliers compared to the session with the bot, and comparable dispersion.

A preliminary glance at *Table 4* also shows that all the scale scores increase in the second session, showcasing as expected a generally higher psychological and emotional arousal when facing a human-controlled avatar rather than a computer-controlled agent. The Psychological Involvement - Empathy seems to be the most affected scale, with the Behavioural Involvement scale directly following and Psychological Involvement - Negative Feelings changing only slightly. Correlation analysis and simple hypothesis testing will follow to determine if and how the scores are related. Before calculating correlation coefficients, it is however most cautious to perform a reliability analysis in order to confirm the validity of the utilised tool.

Reliability Analysis

Reliability analysis is performed in order to validate existing scales within a self-report tool, and to allow researchers to evaluate the consistency and stability of the scale's performance across diverse settings, while also assessing the precision of the measurement intervals. The term reliability refers to the extent to which each scale produces consistent or stable results in various applications of the measurement instrument, as well as providing results with interval consistency. Moreover, reliability analyses serve to scrutinize the properties of scales and individual items, aiming to ascertain whether they capture a shared underlying construct representing the same latent variable.

The items composing a scale should match strongly, showcased in a high inter-item correlation. If the answers to scale items demonstrate significant correlation, the items are to be considered reliable and contributing to a coherent scale. Cronbach's alpha stands as the predominant method for assessing the internal consistency of a scale and measuring the reliability of self-report tools. Functioning as a measurement of the internal consistency of a scale, Cronbach's alpha quantifies the degree of interrelatedness among the items, thereby furnishing an estimate of the overall accuracy. However, it is important to note that Cronbach's alpha does not provide any insights into whether individual items are influenced by singular or multiple latent variables. Nevertheless, in instances where all items capture a singular latent variable, Cronbach's alpha assesses the extent to which the items effectively encapsulate that underlying construct. The formula for calculating Cronbach’s alpha is the following:

$$\alpha = \left(\frac{K}{K-1}\right)\left(1 - \frac{\Sigma s_j^2}{s_y^2}\right)$$

Equation 1: Cronbach’s Alpha

where K is the cardinality of the items, s_y^2 is the variance of the sum of the item scores and s_j^2 is the individual variance of each item score. Cronbach’s Alpha ranges between 0 and 1, with higher values indicating that the survey or questionnaire is more reliable. George and Mallery (2003) provide a table (*Table 5*) for interpreting the Cronbach’s alpha value.

Cronbach’s Alpha	**Interpretation**
$0,9 \leq \alpha$	Excellent
$0.8 \leq \alpha < 0.9$	Good
$0.7 \leq \alpha < 0.8$	Appropriate
$0.6 \leq \alpha < 0.7$	Questionable
$0.5 \leq \alpha < 0.6$	Poor
$\alpha < 0.5$	Unacceptable

Table 5: Cronbach’s Alpha interpretation table for determining the degree to which extent the items within a scale measure the same latent variable.

All the necessary calculations to determine Cronbach’s Alpha were performed on Google Sheets using the above-mentioned function and the data collected from the present experiment. The results were also double-checked and confirmed using the alternative formula:

$$\alpha = \frac{N\bar{c}}{\bar{v}+(N-1)\cdot\bar{c}}$$

Equation 2: Alternative formula for Cronbach’s Alpha

where $\bar{c}$ is the average inter-item covariance, $\bar{v}$ is the average variance and N is the cardinality of the items. The results of the second formula are highly consistent with the displayed ones. The results are the following:

Scales	Cronbach's Alpha
Philanthropist Scale	0.73
Socializer Scale	0.82
Achiever Scale	0.29
Player Scale	0.76
Free Spirit Scale	0.83
Disruptor Scale	0.72
Immersion with Bot Scale	0.45
Flow with Bot Scale	0.79
Empathy with Bot Scale	0.43
Negative Feelings with Bot Scale	0.52
Behavioural Involvement with Bot Scale	0.84
Immersion with Friend Scale	0.71
Flow with Friend Scale	0.59
Empathy with Friend Scale	0.82
Negative Feelings with Friend Scale	0.51
Behavioural Involvement with Friend Scale	0.73

Table 6: The computed Cronbach's Alpha for each scale of the self-report data results.

As images clearly from *Table 6*, the used scales have varying reliability and varying extents to which their items effectively encapsulate that latent variable. The scales of the Hexad-12 questionnaire offer several scales with appropriate and good internal consistency, with only exception of the Achiever scale. According to the calculations performed out of the data collected during the experiment, the Achiever scale of the Hexad-12 Questionnaire seems to have an internal consistency of only 0.29, thus unacceptable according to the interpretation table. As for the other self-report tool, the alpha values showcased by the GEQ-R scales seem to vary notably between each other, even between applications of the same tools in different settings. While the Immersion scale is inaccurate in the computer-controlled agent setting, it seems to become appropriate in the human-controlled avatar setting. The Flow scale varies from appropriate to poor, the Empathy scale increases from unacceptable to good, and the Behavioural scale lowers from good to appropriate. The only immutable one is the Negative Feelings scale which seems to be equally poor in both instances. These variations seem to point to either a severe lack of consistency between the scales in different experiment settings, or to a more complex phenomenon concerning the underlying latent variable that these scales try to measure. For instance, the Empathy scale has an understandable change: questions such as “I empathized with the other”, “I felt connected to the other”, and “When I was happy, the other was happy” seem to measure different constructs according to the opponent type. In a narrative digital game, the computer-controlled characters can be written in such a way that the player feels connected to them and empathizes with them, however, the players would reasonably hardly believe that their happiness could make the virtual agents happy. While these three items may reliably measure the same set of latent variables when playing with a human opponent, their accuracy plummets when applied to a computer-controlled character, probably due to the underlying set of latent variables diverging from each other according to the setting.

While it has to be considered that the dataset utilized for this analysis only disposes of 28 entries, the results are hardly convincing. With the experimental knowledge of the scales of the GEQ-R are hardly appropriate, the extrapolated findings must be examined with a careful approach.

Comparative Analysis

Assessed the reliability of the data, the next step is to individuate and establish relationships among the variables and calculate whether these associations are statistically significant through a comparative analysis and the respective hypothesis testing. Comparative analysis concerns the methodical comparison of two or more entities to determine the presence or absence of shared characteristics or divergences. Hypothesis testing, instead, is a fundamental concept in statistical analysis and empirical research, consisting of a set of systematic methods to evaluate the validity of hypotheses and draw conclusions about the relationships between variables. Through hypothesis testing it is possible to assess the likelihood that observed differences or associations in the data are not due to random chance but rather reflect true relationships in the underlying population. Such analysis is conducted by making statistical inferences about a population parameter based on sample data, typically through the proposal of a null hypothesis, stating that there is no effect or no difference between groups or variables, and an alternative hypothesis, stating the opposite. The goal of hypothesis testing techniques is to calculate evidence against the null hypothesis.

The data collected by the self-reports consists of one dichotomous nominal variable indicating the test setting (0: computer-controlled agent, 1: human-controlled avatar), 12 categorical ordinal variables on a 7-point Likert scale from the Hexad-12 questionnaire, and 26 pairs of categorical ordinal variables on a 5-point Likert scale from the GEQ-R questionnaire. It is important to note that this data, not being continuous and metric in nature, would not be fit for statistical techniques that require continuous data, however, the ordinal variables on Likert scales of at least five categories have been regarded acceptable in the academic research for use as continuous without any harm to the statistical analysis (Johnson and Creech 1983; Sullivan and Artino 2013). Nevertheless, these ordinal variables are not normally distributed, and as such it is not possible to use hypothesis testing techniques that would require the data to be normally distributed such as t-testing or analysis of variance. This study will thus favour the non-parametric counterparts of the most common techniques. Another characteristic worth noticing is that this data is extrapolated on the same group of individuals in different settings, as such the sample is to be considered dependent and paired. Considering all of the above it is chosen to perform as first analysis a Wilcoxon signed-rank test on the pairs of the categorical ordinal variables of the Immersion, Flow and Social Presence scales.

The Wilcoxon signed-rank test serves as the non-parametric counterpart to the t-test when analyzing dependent samples, i.e. samples where measurements are paired, often representing repeated assessments on the same subject. While the t-test measures the difference between observed values and their mean, the Wilcoxon signed-rank test measures the difference between the ranks of the paired values and their mean rank. This statistical approach applies to paired data with at least ordinal scale characteristics, where assessments are made on two related random samples. The null hypothesis is

H_0: *In the population, the central tendencies of the two dependent samples are the same.*

and the alternative hypothesis is

H_a: *In the population, the central tendencies of the two dependent samples are unequal.*

In other words, the alternative hypothesis states that the different settings have an impact on the central tendency of the data. The procedure to perform the Wilcoxon signed-rank test consists of computing the difference between the paired values, ranking the unsigned calculated difference, and summing according to the sign the ranked values into a positive rank sum T^+ and a negative rank sum T^-. Afterwards, a test statistics W is calculated by taking the lower of the two values, and an expected value μ_W calculated as

$$\mu_W = \frac{n \cdot (n+1)}{4}$$

Equation 3: Wilcoxon Signed-Rank expected value

where *n* is the cardinality of the sample. Then the standard deviation is calculated with the following formula

$$\sigma_W = \sqrt{\frac{n\cdot(n+1)\cdot(2n+1) - \sum_{i=1}^{k}\frac{t^3-t}{2}}{24}}$$

Equation 4: Wilcoxon Signed-Rank Standard Deviation

where k is the number of tied ranks and t is the frequency of those tied ranks. Considering a sample size greater than 25, the Wilcoxon signed-rank test can be considered to fairly approximate a normal distribution, and consequently, it is possible to calculate a p-value of the approximated t-score. The t-score is calculated as

$$t = \frac{W-\mu_W}{\sigma_W}.$$

Equation 5: Wilcoxon Signed-Rank t-score

Utilizing the t-score and the sample size it is possible to interpolate the p-value, i.e. the probability of the data to be collected given the premises on the null hypothesis. Upon comparing the p-value to the significance level α it is possible to determine whether the scrutinized association is statistically significant or not. For ethical reasons, it is recommended to choose the significance level α before calculating the p-value. The most common significance levels used in hypothesis testing are 0.05, 0.01, and sometimes 0.10. These levels are considered conventional choices and are widely accepted in many fields of research. With a sample size as small as 28 this study is characterized by a lower statistical power to detect true correlations, which makes it less likely to detect smaller effects even if existing in the population. The choice of the significance level α must balance the risk between Type I (false positives) and Type II (false negatives) errors. Choosing a lower significance level such as 0.01 may reduce the risks of Type I errors, but may increase accordingly the risk of Type II errors due to reduced statistical power. Since no insight is provided by the authors of the utilized self-report tools, in this master thesis study it is chosen to use the balanced significance level of 0.05.

Following are the results of the application of the described procedure within Google Sheets for each item of the Immersion, Flow, and Social Presence scales:

Item		**t-score**	**p-value**
i1	I was interested in the game’s story	-0.28	0.79
f1	I was fully occupied with the game	-2.11	0.07
i2	I thought about other things	-1.08	0.30
i3	I found it tiresome	-0.55	0.59

f2	I forgot everything around me	-2.48	0.02
f3	I lost track of time	-2.69	0.02
f4	I was deeply concentrated in the game	-2.23	0.05
i4	It felt like a rich experience	-2.62	0.02
f5	I lost connection with the outside world	-2.56	0.02
e1	I empathized with the other	-4.57	$1.15 \cdot 10^{-4}$
b1	My actions depended on the other’s actions	-1.53	0.15
b2	The other's actions were dependent on my actions	-3.32	$1.12 \cdot 10^{-3}$
e2	I felt connected to the other	-4.52	$1.31 \cdot 10^{-4}$
b3	The other paid close attention to me	-4.43	$1.76 \cdot 10^{-4}$
b4	I paid close attention to the other	-1.71	0.12
n1	I felt jealous about the other	-0.10	0.93
e3	I found it enjoyable to be with the other	-4.05	$6.35 \cdot 10^{-4}$
e4	When I was happy, the other was happy	-3.67	$1.62 \cdot 10^{-3}$
e5	When the other was happy, I was happy	-4.08	$5.84 \cdot 10^{-4}$
n2	I influenced the mood of the other	-3.97	$5.71 \cdot 10^{-4}$
n3	I was influenced by the other moods	-3.19	$4.62 \cdot 10^{-3}$
e6	I admired the other	-3.89	$9.13 \cdot 10^{-4}$
b5	What the other did affected what I did	-2.48	0.02
b6	What I did affected what the other did	-3.89	$9.10 \cdot 10^{-4}$
n4	I felt revengeful	-1.56	0.13
n5	I felt schadenfreude (malicious delight)	-0.29	0.78

Table 7: The t-score and p-value computed using the Wilcoxon Signed-Rank test for all items of the self-report data results.

By comparing these items’ p-values to the significance level α of 0.05 it is possible to determine which associations are statistically significant. In more detail, for the items of which the probability of the data resulting as it did (i.e. the p-value) is lower than the threshold given by the significance levels, the null hypothesis must be rejected in favour of the alternative hypothesis. Thus, for the items f2, f3, f4, i4, f5, e1, b2, e2, b3, e3, e4, e5, n2, n3, e6, b5 and b6, there is a statistically significant association to the experiment dichotomous variable. For the same reasons of readability elucidated previously, the same analysis is performed also on the scale scores, i.e. the arithmetic mean of the items within a scale.

Scale	t-score	p-value
Immersion Scale	-1.59	0.13
Flow Scale	-3.91	$7.00 \cdot 10^{-4}$
Empathy Scale	-4.54	$1.00 \cdot 10^{-4}$
Negative Feelings Scale	-2.83	$9.50 \cdot 10^{-3}$
Behavioural Involvement Scale	-4.46	$2.00 \cdot 10^{-4}$

Table 8: The t-score and p-value computed using the Wilcoxon Signed-Rank test for the scales of the self-report data results.

It seems that the null hypothesis, stating that there is no variation between the central tendencies of the data collected in the two different experiment settings, has to be rejected for the Flow, Empathy, Negative Feelings and Behavioural Involvement scales. The results of the Wilcoxon signed-rank test point to a statistically significant association between the test setting (computer-controlled agent and human-controlled avatar) and all of the scales but Immersion. Notably, the p-values of the Flow, Negative Feelings and Behavioural Involvement scales are also considerably low, especially if compared to the p-values of the corresponding items. To provide further insight into the phenomenon, it is chosen to calculate the effect size of the utilized method, i.e. the measurement of the strength of the association. For the Wilcoxon signed-rank test, the effect size is calculated as

$$r = \frac{t}{\sqrt{n}} .$$

Equation 6: Wilcoxon Signed-Rank effect size

Cohen (1988) provides the following table (*Table 9*) as a guideline for the interpretation of this effect size:

$\lvert r \rvert < 0.1$	no effect / very small effect
$\lvert r \rvert > 0.1$	small effect
$\lvert r \rvert > 0.3$	medium effect
$\lvert r \rvert > 0.5$	large effect

Table 9: The effect size interpretation table for determining the degree of variation between the central tendencies of the two dependent samples.

Taking into account all the aforementioned formulas and the interpretation guidelines just outlined, the calculations are once again computed using Google Sheets, providing the following results:

Item/Scale		**Effect Size**
i1	I was interested in the game's story	-0.09
f1	I was fully occupied with the game	-0.75
i2	I thought about other things	-0.31
i3	I found it tiresome	-0.17
f2	I forgot everything around me	-0.55
f3	I lost track of time	-0.67
f4	I was deeply concentrated in the game	-0.67
i4	It felt like a rich experience	-0.62
f5	I lost connection with the outside world	-0.60
e1	I empathized with the other	-0.90
b1	My actions depended on the other's actions	-0.40
b2	The other's actions were dependent on my actions	-0.69
e2	I felt connected to the other	-0.89
b3	The other paid close attention to me	-0.89
b4	I paid close attention to the other	-0.51
n1	I felt jealous about the other	-0.03
e3	I found it enjoyable to be with the other	-0.88
e4	When I was happy, the other was happy	-0.82
e5	When the other was happy, I was happy	-0.89
n2	I influenced the mood of the other	-0.79
n3	I was influenced by the other moods	-0.70
e6	I admired the other	-0.85
b5	What the other did affected what I did	-0.57
b6	What I did affected what the other did	-0.85
n4	I felt revengeful	-0.35
n5	I felt schadenfreude (malicious delight)	-0.07
i	Immersion Scale	-0.32

f	Flow Scale	-0.80
e	Empathy Scale	-0.87
n	Negative Feelings Scale	-0.58
b	Behavioural Involvement Scale	-0.88

Table 10: The computed Wilcoxon Signed-Rank test effect sizes for all the items and scales of the self-report data results.

As emerges clearly from *Table 10*, the effect sizes seem to be consistent with the p-values previously calculated. Notably, the Flow, Empathy and Behavioural Involvement scales seem to have the stronger association with the experiment setting, with Immersion having barely medium effect and Negative feelings having a large effect.

In an attempt to increase the scientific rigour and in the aim to provide more confirmation of this detected association, further analysis is conducted on the same data using the Chi-square analysis. The chi-square analysis is a hypothesis test analysis commonly utilized to determine the presence of a relationship between two categorical variables, i.e. variables characterized by discrete and finite values, typically nominal. The chi-square test can be used with different analysis scopes, such as to assess the independence between two categorical variables (independence test), to examine the concordance between observed and expected values (distribution test) and to determine the similarity of multiple samples' origins (homogeneity test). This master's thesis goal is to investigate the relationships between subjective experiences, thereby discerning their independence or interdependence and potential influence on each other. In the procedure to perform an independence test through the calculation of the chi-square value, it is necessary to dispose of both observed and expected frequencies of the data, where the expected frequencies consist of the hypothetical results with the premises of full independence between the variables. The expected frequencies under the premises of full independence can be calculated with the formula

$$E(i, j) \;=\; \frac{RowSum(i) \cdot RowSum(j)}{N}$$

Equation 7: Chi-Squared expected frequencies

where i and j are the rows and columns of the table respectively. The main purpose of the chi-squared method is to offer a tool for significance testing, thus performing calculations that lead to evidence rejecting the null hypothesis. This study maintains the same significance level α of 0.05 for all the hypothesis testing tools utilized. In the case of the chi-squared hypothesis testing, the null hypothesis is

H_0: *there is no relationship between the item scores and the experiment setting* and the alternative hypothesis is

H_a: *there is a relationship between the item scores and the experiment setting.*

While it is possible to calculate a p-value to compare to the significance level α, which is also the procedure when utilizing the chi-squared test for distribution testing, chi-squared independence testing is performed through the consultation of a chi-squared distribution table. The chi-squared distribution is a statistical probability distribution that emerges from the summation of n independent and identically distributed random variables, each following a normal distribution, where n denotes the degrees of freedom. When conducting an independence hypothesis test utilizing the chi-squared test, it is necessary to compare the computed chi-squared statistic, acquired from the processing of the collected data, to the chi-squared critical value, acquired through consultation of the chi-squared distribution. It is possible to extrapolate the chi-square critical value from the chi-square distribution table by interpolating between the degrees of freedom and the chosen significance level. If the computed chi-squared statistic value falls below the critical value, the null hypothesis can be retained, if it is greater, the null hypothesis must be rejected. Considering this significance level α of 0.05, and the calculated degrees of freedom which will be shown in a following table (*Table 11*), it is possible to extrapolate the following critical values:

df	2	3	4
Critical value	5.991	7.815	9.488

Table 11: The critical values from the chi-squared distribution table given the specified degree of freedom.

Prepared the data for the comparison, it is time to compute the chi-squared statistic. The formula to calculate the chi-square is

$$\chi^2 = \sum_{k=1}^{n} \frac{(O_k - E_k)^2}{E_k}$$

Equation 8: Chi-Squared formula

where *O* is the observed value and *E* is the respective expected value. Additionally, the chi-square test enables the assessment of the strength of the relationship between variables through effect size estimation. There are multiple tools to calculate the measure of the chi-squared independence test effect size, such as Phi φ, Odds Ratio and Cramér's V. Only the last one is suitable for this study's data structure as the other two methods are limited to 2x2 contingency tables. Cramer's V can be calculated through the following formula

$$V = \sqrt{\frac{\chi^2}{n \cdot df}}$$ *Equation 9: Cramer's V formula*

where n is the total amount of observations and df is the degree of freedom calculated using the following formula

$$df = (number\ or\ rows - 1) \cdot (number\ of\ colums) - 1)$$
Equation 10: Chi-Squared degree of freedom formula

Cramer's V can be interpreted with the following table:

df*	small	medium	large
1	.10	.30	.50
2	.07	.21	.35
3	.06	.17	.29
4	.05	.15	.25
5	.04	.13	.22

Table 12: Cramer's V threshold table for determining measure the degree of association between nominal variables.

Having prepared all the comparative tools for the analysis, the study continues calculating the aforementioned values for each item of the Flow, Immersion and Social Presence self-report. The p-value is calculated using the Google Sheets CHISQ.DIST.RT function, and double-checked with the CHITEST function. Following are the results for each item:

	χ2	**df**	**p-value**	***V***
i1	0.72	4	0.95	0.05
f1	3.24	2	0.20	0.31
i2	3.00	4	0.56	0.20
i3	6.73	4	0.15	0.45
f2	5.72	3	0.13	0.44

f3	4.85	4	0.30	0.32
f4	4.36	3	0.23	0.34
i4	7.24	4	0.12	0.48
f5	4.81	4	0.31	0.32
e1	45.20	4	$3.62 \cdot 10^{-9}$	3.02
b1	3.69	4	0.45	0.25
b2	9.24	4	0.06	0.62
e2	39.28	4	$6.10 \cdot 10^{-8}$	2.62
b3	26.94	4	$2.02 \cdot 10^{-5}$	1.80
b4	2.63	4	0.62	0.18
n1	0.61	4	0.96	0.04
e3	23.12	4	$1.20 \cdot 10^{-4}$	1.54
e4	22.59	3	$4.92 \cdot 10^{-5}$	1.74
e5	24.98	4	$5.07 \cdot 10^{-5}$	1.67
n2	30.39	4	$4.07 \cdot 10^{-6}$	2.03
n3	16.83	4	$2.08 \cdot 10^{-3}$	1.12
e6	13.38	4	0.01	0.89
b5	6.55	4	0.16	0.44
b6	16.41	4	$2.52 \cdot 10^{-3}$	1.10
n4	5.38	4	0.25	0.36
n5	1.00	4	0.91	0.07

Table 13: The computed Chi-Squared vaues, degree of freedom, p-values and Cramer's V for all the items of the self-report data results.

From these computations emerges already that the probabilities given the premises of the null hypothesis are incredibly small for multiple items of self-reports. The items of the Empathy and Behavioural Involvement scales present remarkably small p-values, and many other items show notably strong effect sizes. To improve readability and to facilitate the identification of patterns, the same calculations are performed also on the scale values:

	χ2	**df**	**p-value**	**V**

Immersion	5.17	3	0.16	0.40
Flow	5.93	3	0.11	0.46
Empathy	45.23	4	$3.57 \cdot 10^{-9}$	3.02
Negative Feelings	3.60	3	0.31	0.28
Behavioural Involvement	12.31	4	0.02	0.82

Table 14: The computed Chi-Squared vaues, degree of freedom, p-values and Cramer's V for the scales of the self-report data results.

As described previously, whenever the computed chi-square statistics is smaller than the chi-squared critical value, it is considered to have an absence of statistically significant difference, and the null hypothesis is retained. Conversely, if the computed chi-square statistics is greater than the chi-squared critical value, the null hypothesis must be rejected in favour of the alternative hypothesis. Considering that the critical value for a significance level α of 0.5 and 4 degrees of freedom is 9.488, both the Empathy scale and the Behavioural Involvement scale show a statistically significant difference between the observed data and the expected data, thus accepting the alternative hypothesis stating that there is a relationship between the scales and the experiment setting. The remaining three scales, Immersion, Flow and Negative Feelings, seem to have a computed chi-squared statistic lower than the critical value of 7.815, thus leading to the retention of the null hypothesis claiming that there is no relationship between the scale and the experiment setting.

Having confirmed the presence of a relationship for at least two scales through two comparative analyses, an attempt is made to identify the characteristics of this relationship through a correlation analysis.

Correlation Analysis

Correlation analysis constitutes a fundamental statistical set of techniques utilized to explore and assess the relationships between two or more variables within a dataset. Its primary objective is to identify relationships between variables and individuate the presence of significant correlations. This approach serves as a central tool in understanding how variations in one variable can correspond to changes in another, identifying the relationship showcased by these changes, its strength and its properties. If the correlation analysis determines that two variables have a relationship, a potential predictive relationship can be subsequently investigated in regression analysis. Two of the most prevalent techniques employed in correlation analysis are the Pearson correlation coefficient and the Spearman correlation coefficient. The Pearson correlation analysis is utilized when variables exhibit a linear relationship, while the Spearman correlation analysis is employed in scenarios where variables are ranked or ordinal in nature, or when the assumption of linearity is violated.

The Pearson correlation analysis is specifically designed to quantify linear relationships between variables and is widely utilized across various disciplines to uncover patterns of dependency. The Pearson correlation coefficient *r* gives information on both the strength and direction of the observed correlation, thereby furnishing researchers with comprehensive insights into the strength and nature of the relationship. The calculation of the Pearson correlation coefficient is based on sample data, yet its applicability extends to hypothesis testing about population parameters. This process involves assessing whether the observed correlation is statistically significant within the population context, which is determined by examining whether the correlation coefficient is statistically significantly different from zero, thereby signifying the presence of a meaningful relationship. It is necessary however to ensure that the examined variables are normally distributed, as otherwise the validity of inferential statistics such as the t-test and associated p-values may be compromised and be interpreted unreliably. The Pearson correlation coefficient can be calculated with the following formula:

$$r = \frac{\Sigma(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\Sigma(x_i - \overline{x})^2 \Sigma(y_i - \overline{y})^2}}$$

Equation 11: Pearson correlation coefficient formula

where $\overline{x}$ and $\overline{y}$ are the sample mean values. The direction of the correlation is thus either negative or positive, as exemplified by the correlation coefficient sign. Furthermore, according to the correlation coefficient value, it is possible to determine the strength of the relationship. Following is the interpreting table by Kuckartz et al (2013).

r	**Strength of the correlation**
$0.0 \leq \lvert r \rvert < 0.1$	No correlation
$0.1 \leq \lvert r \rvert < 0.3$	Little correlation
$0.3 \leq \lvert r \rvert < 0.5$	Medium correlation
$0.5 \leq \lvert r \rvert < 0.7$	High correlation
$0.7 \leq \lvert r \rvert$	Very high correlation

Table 15: The Point-Biserial correlation coefficient interpretation table for determining the strength of correlation between variables.

A special case of the Pearson correlation is the point-biserial correlation, which examines the relationship between a dichotomous variable (i.e. a binary, jointly exhaustive and mutually exclusive) and a metric continuous variable. The Person correlation and the point-biserial, are mathematically equivalent and it is possible to calculate the latter using the same formula as the former, simply by having the dichotomous variable being expressed as two distinct numerical, values such as 0 and 1. As mentioned in the previous section, this study wants to evaluate the relationship between a dichotomous variable (0: computer-controlled agent, 1: human-controlled avatar) and the ordinal categorical data given by the 5-point Likert scales of the self-report. It is important to remark that categorical data is not continuous, but ordinal data from Likert scales of at least 5 points is considered to be a good approximation and fit for continuous statistical analysis. The ordinal data is however still not normally distributed, and as such it is not possible to do a correlation hypothesis testing utilizing the Pearson or the point-biserial correlation, as the t-test and the p-value would result unreliable. The processing is still performed because even though it can not be used to identify a statistical significance, it can still be useful as a tool for identifying patterns in the relationships.

Following are the results of the correlation calculated with the Google Sheets CORREL function, having on the abscissa the concatenation of the item scores for the two different settings (first the “with bot” and then the “with friend” item scores), and on the ordinate an array of 28 “0” and 28 “1” symbolising the two different settings (0: computer-controlled agent, 1: human-controlled avatar).

Label	**Item**	***r***
i1	I was interested in the game’s story	0.04
f1	I was fully occupied with the game	0.23
i2	I thought about other things	-0.10
i3	I found it tiresome	0.04
f2	I forgot everything around me	0.31
f3	I lost track of time	0.22
f4	I was deeply concentrated in the game	0.25
i4	It felt like a rich experience	0.32
f5	I lost connection with the outside world	0.24
e1	I empathized with the other	0.83
b1	My actions depended on the other’s actions	0.18

b2	The other's actions were dependent on my actions	0.40
e2	I felt connected to the other	0.82
b3	The other paid close attention to me	0.63
b4	I paid close attention to the other	0.13
n1	I felt jealous about the other	0.00
e3	I found it enjoyable to be with the other	0.61
e4	When I was happy, the other was happy	0.57
e5	When the other was happy, I was happy	0.61
n2	I influenced the mood of the other	0.68
n3	I was influenced by the other moods	0.43
e6	I admired the other	0.49
b5	What the other did affected what I did	0.31
b6	What I did affected what the other did	0.53
n4	I felt revengeful	-0.20
n5	I felt schadenfreude (malicious delight)	-0.03

Table 16: The computed Point-Biserial correlation coefficient for all the items of the self-report data results.

As emerges clearly in *Table 16*, the items of the Immersion scale show barely any correlation while the items of the Flow scale show medium to little correlation, showing this consistency with the results of the comparative analysis. The items of the Empathy scale range from high to very high correlation, while the items of the Negative Feeling and Behavioural Involvement scales show a more nuanced strength. This analysis allows to see the direction of the relationship, which seems to be positive for the majority of the items, except for a few items on the Negative Feelings Scale. The following table provides the same analysis applied to the scale scores:

Scales:	***r***
Immersion Scale	0.16
Flow Scale	0.34
Empathy Scale	0.81
Negative Feelings Scale	0.23
Behavioural Involvement Scale	0.50

Table 17: The computed Point-Biserial correlation coefficient for the scales of the self-report data results.

Consistently with the previous findings, the Empathy scale and the Behavioural Involvement scale have the stronger relationships with the experiment settings, while both the Immersion and Negative Feelings scales present little correlation, and the Flow scale presents a medium correlation. All the computed relationships are positive in the direction, suggesting that all of these subjective experiences are aroused more greatly when facing a human opponent rather than a computer one, consistently with the findings in the literature.

Similarly to the comparative analysis, in an attempt to increase the scientific rigour and in the aim to find a statistically significant correlation, further correlation analysis is conducted on this data using non-parametric counterparts. In situations characterized by non-normally distributed data, researchers commonly turn to non-parametric variants of correlation methods, such as the Spearman correlation analysis. Similar to its Pearson or point-biserial counterpart, Spearman correlation analysis examines the association between two variables. However, unlike the Pearson method, Spearman utilizes ranked data rather than raw values. These rankings represent discrete ordinal data and do not necessitate adherence to normal distribution assumptions. The methodological distinction lies in the transformation of raw data into ranks before applying the Pearson correlation coefficient formula in the Spearman version. While the Pearson correlation test requires continuous data, Spearman measures the strength of association through a monotonic relationship between ranked variables. However, the Spearman correlation analysis is also known to perform poorly in data sets with a lot of rank ties, which is the case for this Likert scale dataset. The rank-biserial correlation, the non-parametric counterpart for the point-biserial correlation also is not suitable for paired samples as it is based on the difference of the mean of the ranks, which in paired samples is always zero. For this reason, this study will resort to Kendall’s Tau correlation analysis as it is commonly deemed more appropriate and suitable for such a situation. Like the other mentioned correlation analyses, Kendall’s Tau evaluates the relationship between variables and similarly to the Spearman correlation method it operates as a non-parametric test, bypassing the need for normally distributed data, but only requiring an ordinal scale. The reason that makes Kendall’s Tau the right choice for this analysis is that this method outperforms Spearman’s rank correlation when confronted with minimal data exhibiting numerous rank ties, as observed in the present experiment. The formula to calculate Kendall’s Tau is:

$$\tau = \frac{n_C - n_D}{\binom{n}{2}}$$

Equation 12: Kendall’s Tau formula

or

$$\tau = \frac{n_C - n_D}{n_C + n_D}$$

Equation 13: Kendall's Tau alternative formula

where n_C is the number of concordant pairs, n_D is the number of discordant pairs and n is the cardinality of the sample. In the case of many rank ties it is suggested to use the formula

$$\tau = \frac{n_C - n_D}{\sqrt{\left(\binom{n}{2} - n_x\right) \cdot \left(\binom{n}{2} - n_y\right)}}$$

Equation 14: Kendall's Tau formula when many rank ties

where n_x and n_y are of ties in the variable *x* and in the variable *y*. To calculate the pairs it is first necessary to rank all the data, then to order the second sample according to the ranking order of the first sample. In this second sample, the adjacent values are compared and if the relationship is increasing then it is a concordant pair, if it is decreasing then it is a discordant pair, and if it is stationary then it is a tie. Identifying and quantifying concordant and discordant pairs is essential in calculating Kendall's Tau coefficient, which measures the degree of association or correlation between variables based on the proportion of concordant and discordant pairs relative to the total number of pairs. A higher proportion of concordant pairs indicates a stronger positive association, while a higher proportion of discordant pairs suggests a stronger negative association.

Analogously to the Pearson correlation coefficient *r*, Kendall's tau varies between 1 and -1, thus allowing to similarly evaluate both the direction and the strength of the relationship. The same interpreting table applies.

τ	**Strength of the correlation**
$0.0 \le \lvert\tau\rvert < 0.1$	No correlation
$0.1 \le \lvert\tau\rvert < 0.3$	Little correlation
$0.3 \le \lvert\tau\rvert < 0.5$	Medium correlation
$0.5 \le \lvert\tau\rvert < 0.7$	High correlation
$0.7 \le \lvert\tau\rvert$	Very high correlation

Table 18: The Kendall' rank correlation coefficient interpretation table for determining the strength of correlation between variables.

Following are the results of the correlation calculated with the above-mentioned procedure on each item of the self-report. The computations were performed with the last described formulas and double-checked with the first proposed tau formula.

Label	Item	τ
i1	I was interested in the game's story	-0.01
f1	I was fully occupied with the game	0.08
i2	I thought about other things	-0.11
i3	I found it tiresome	-0.13
f2	I forgot everything around me	0.26
f3	I lost track of time	0.06
f4	I was deeply concentrated in the game	0.12
i4	It felt like a rich experience	0.15
f5	I lost connection with the outside world	0.09
e1	I empathized with the other	0.58
b1	My actions depended on the other's actions	0.16
b2	The other's actions were dependent on my actions	0.29
e2	I felt connected to the other	0.73
b3	The other paid close attention to me	0.54
b4	I paid close attention to the other	0.07
n1	I felt jealous about the other	-0.02
e3	I found it enjoyable to be with the other	0.37
e4	When I was happy, the other was happy	0.47
e5	When the other was happy, I was happy	0.54
n2	I influenced the mood of the other	0.66
n3	I was influenced by the other moods	0.37
e6	I admired the other	0.39
b5	What the other did affected what I did	0.22
b6	What I did affected what the other did	0.44
n4	I felt revengeful	-0.12
n5	I felt schadenfreude (malicious delight)	-0.09

Table 19: The computed Kendall rank correlation coefficient for all the items of the self-report data results.

According to this results, there is a very high positive correlation between the item "I felt connected to the other" of the Empathy scale and the experiment setting. Multiple other items of the Empathy and Behavior Involvement scales show a high or medium correlation positive correlation, and two items of the Negative Feelings scale show medium and high positive correlation while the rest exhibit little negative correlation. The correlation between the experiment setting and the Immersion and Flow scales seems to be hardly relevant. The same calculations are also performed for the scale scores.

Scale	$\boldsymbol{\tau}$
Immersion	0.12
Flow	0.14
Empathy	0.53
Negative Feelings	0.27
Behavioural Involvement	0.39

Table 20: The computed Kendall rank correlation coefficient for the scales of the self-report data results.

From these results emerges a high correlation between the Empathy scale and the experiment settings, a medium correlation for the Behavioural involvement scale, and little correlation for the remaining scales.

The study proceeds thus with the hypothesis testing of the computed correlation coefficients. The null hypothesis would be

H_0: *there is no correlation between the variables, thus the correlation coefficient* $\boldsymbol{\tau}$ *is 0*

and the alternative hypothesis would be

H_a: *there is a correlation between the variables, thus the correlation coefficient* $\boldsymbol{\tau}$ *is different from 0.*

To check if the correlation coefficient is significantly different from zero in the sample collected, it is needed to use a z-distribution as an approximation. It is important to note that it is recommended to have a sample of at least 40 entries to calculate the z-value from Kendall's tau. In this study, the dataset only disposes of 28 entries, so the following computations have limited reliability. The formula to calculate the z-value, i.e. the number of standard deviations a data point is above or below the mean of the distribution, is

$$Z = \frac{3\tau\sqrt{n(n-1)}}{\sqrt{2(2n+5)}}$$

Equation 15: Kendall's Tau z-score formula

The following table (*Table 21*) presents the application of the just described formula on the data acquired so far, with the addition of the computation of the relative p-value, i.e. the likeability of the observed data to be collected under the premises of no correlation.

Item/Scale		**z-score**	**p-value**
i1	I was interested in the game's story	-0.14	1.11
f1	I was fully occupied with the game	0.87	0.38
i2	I thought about other things	-1.20	1.77
i3	I found it tiresome	-1.45	1.85
f2	I forgot everything around me	2.88	$4.00 \cdot 10^{-3}$
f3	I lost track of time	0.62	0.54
f4	I was deeply concentrated in the game	1.31	0.19
i4	It felt like a rich experience	1.63	0.10
f5	I lost connection with the outside world	0.97	0.33
e1	I empathized with the other	6.35	$2.18 \cdot 10^{-10}$
b1	My actions depended on the other's actions	1.74	0.08
b2	The other's actions were dependent on my actions	3.12	$1.81 \cdot 10^{-3}$
e2	I felt connected to the other	7.99	$1.33 \cdot 10^{-15}$
b3	The other paid close attention to me	5.89	$3.75 \cdot 10^{-9}$
b4	I paid close attention to the other	0.72	0.47
n1	I felt jealous about the other	-0.21	1.16
e3	I found it enjoyable to be with the other	3.98	$7.01 \cdot 10^{-5}$
e4	When I was happy, the other was happy	5.10	$3.39 \cdot 10^{-7}$
e5	When the other was happy, I was happy	5.84	$5.08 \cdot 10^{-9}$
n2	I influenced the mood of the other	7.14	$9.19 \cdot 10^{-13}$
n3	I was influenced by the other moods	4.07	$4.63 \cdot 10^{-5}$

e6	I admired the other	4.28	$1.84 \cdot 10^{-5}$
b5	What the other did affected what I did	2.45	0.01
b6	What I did affected what the other did	4.74	$2.14 \cdot 10^{-6}$
n4	I felt revengeful	-1.35	1.82
n5	I felt schadenfreude (malicious delight)	-0.99	1.68
i	Immersion Scale	1.335	0.18
f	Flow Scale	1.516	0.13
e	Empathy Scale	5.755	$8.68 \cdot 10^{-9}$
n	Negative Feelings Scale	2.924	$3.46 \cdot 10^{-3}$
b	Behavioural Involvement Scale	4.197	$2.70 \cdot 10^{-5}$

Table 21: The computed Kendall rank correlation z-scores and p-values for all the items and the scales of the self-report data results.

Multiple items of the social presence scales present a statistically significant correlation between their values and the experiment settings, as do the three social presence scales, having p-values notably below the significance level α of 0.05.

An extra correlation analysis is conducted in order to find possible correlations between the social presence dimensions and the hexad player typologies. Following are the results of a point-biserial correlation calculated on the arithmetic means of each personality scale, and on the difference between the social presence dimension measured within the two settings.

Scales	**Psychological Involvement (Empathy)**	**Psychological Involvement (Negative Feelings)**	**Behavioural Involvement**
Philanthropist Scale	-0.16	-0.12	0.00
Socializer Scale	0.24	-0.37	0.27
Achiever Scale	-0.05	0.10	-0.39
Player Scale	0.02	0.17	-0.36
Free Spirit Scale	0.07	0.28	-0.08

Disruptor Scale	-0.25	0.24	-0.12

Table 22: The Point-Biserial correlation coefficients computed between the Hexad-12 Scales and the SPGQ scales of the self-report data results.

Observing these results it is clear that for most combinations the correlation coefficient is very low. The Socializer scale presents a negative medium correlation with the Psychological Involvement (Negative Feelings Scale), suggesting that the more the individual considers themselves motivate by relatedness, the less they develop negative feelings when interacting with the other entity. Both the Achiever and the Player scales present a negative medium correlation with the Behavioural Involvement scale, suggesting that the players that consider themselves motivated by competence and extrinsic rewards, find their behaviour less influenced by the other entity's behaviour. Since it is not possible to calculate significance on this data, Kendall's tau is calculated as well.

Scales	**Psychological Involvement (Empathy)**	**Psychological Involvement (Negative Feelings)**	**Behavioural Involvement**
Philanthropist Scale	-0.07	0.04	0.09
Socializer Scale	0.11	-0.25	0.22
Achiever Scale	-0.04	0.18	-0.25
Player Scale	0.05	0.18	-0.15
Free Spirit Scale	0.08	0.35	-0.07
Disruptor Scale	-0.12	0.20	-0.11

Table 23: The Kendall rank correlation coefficients computed between the Hexad-12 Scales and the SPGQ scales of the self-report data results.

Through this second correlation analysis the coefficients seem to be even smaller than before, having all the combinations presenting low correlation, except for the combination of the Free Spirit scale and the Psychological Involvement (Negative Feelings) scale. Further calculations were performed to obtain the p-values, which resulted in non-significance for all the values.

Facial Action Units Data Results

The following section presents the results of the quantitative analysis conducted through the physiological measurement methodology examining the subtle movements of individual facial muscles codified in the Facial Action Coding System. For each participant, two recordings were made, one for each game session. Unfortunately, two recordings of two different participants were lost due to human error, leading to the discard of the whole recorded pair of each of the two participants, thus four recordings in total were discarded. The analysis will consequently be performed on the remaining 52 videos. The recordings of the players' facial behaviour were processed through the OpenFace software which outputted tables of data in CSV files, each with 40 columns and a row for each frame of the recording, thus constituting a row number proportional to the length of the video (typically between 7 and 15 minutes) and the framerate (30fps). The first 5 columns contain technical values such as the frame index and the timestamp. The 17 central columns contain the continuous metric value of the intensity of the 17 measured facial action units on a 5-point scale, with 0 denoting the absence of the relative facial muscle activation, 1 denoting the presence at the minimum intensity and 5 the presence at the maximum intensity. The last 18 columns contain the binary measurement of the presence or absence of facial muscle activation relative to the 18 present facial action units (one more than the previous set). In this master thesis study, only the central 17 columns denoting the facial action unit intensity were needed, the rest was pruned. In addition, in order to make it possible to work on the acquired data, it was chosen to not work on the entirety of the measurements, but only on their descriptive statistics. This choice comes from the fact that the total amount of physiological data consists of 52 tables of on average 10 thousand rows each multiplied by 17 facial action units. Considering the tools at disposal, the data amount needs to be reduced with a sensible decision.

It was thus chosen to extrapolate for each table only three descriptive statistics for each facial action unit: the arithmetic mean, the maximum value and the standard deviation. These three key indicators were chosen because of how the research question is formulated: “Is there a relationship between perceived social presence and a more visible facial emotional behaviour?”. To define “more visible facial emotional behaviour” three measures were thought necessary: a measurement of the central tendency, to identify if a face activates more muscle groups in general (the arithmetic mean was preferred over the median to take into consideration the outliers); a measurement of the distribution, to identify if there is more variation in the facial behaviour; and the maximum, because a “more visible facial emotional behaviour” is more appropriately showcased by clear recognizable facial expressions rather then the more subtle face movements. The described data selection resulted in six double-entry tables: the arithmetic mean of each facial action unit for each participant in the setting with the computer-controlled agent, the arithmetic mean of each facial action unit for each participant in the setting with the human-controlled avatar, the maximum value of each facial action unit for each participant in the setting with the computer-controlled agent, the maximum value of each facial action unit for each participant in the setting with the human-controlled avatar, the standard deviation of each facial action unit for each participant in the setting with the computer-controlled agent, the standard deviation of each facial action unit for each participant in the setting with the human-controlled avatar. The integral six tables are provided in the appendix of this manuscript (Appendix III). For future reference, the muscle groups related to each measured facial action unit are

AU	**Description**
01	Inner Brow Raiser
02	Outer Brow Raiser
04	Brow Lowerer
05	Upper Lid Raiser
06	Cheek Raiser
07	Lid Tightener
09	Nose Wrinkler
10	Upper Lip Raiser
12	Lip Corner Puller
14	Dimpler
15	Lip Corner Depressor
17	Chin Raiser
20	Lip Stretcher
23	Lip Tightener
25	Lips part
26	Jaw Drop
45	Blink

Table 24: The Facial Action Units studied in the experiment and their code.

Descriptive Statistics

Following is a comprehensive and condensed description of the extrapolated data in the form of descriptive statistics for each table. The values are calculated through Google Sheets embed functions AVERAGE, MEDIAN, MAX, MIN, VARA, and STDEV. At the end of each table will be provided an Overall Facial Action value calculated as the average of all the Facial Action Units in the original tables, computed in an attempt to generalize the measurements into a wider facial activation measurement.

Arithmetic Mean Values										
	With Bot					**With Friend**				
AU	**Mean**	**Median**	**Range**	**Vara**	**Dev**	**Mean**	**Median**	**Range**	**Vara**	**Dev**
01	0.21	0.19	0.27	0.01	0.08	0.22	0.19	0.35	0.01	0.09
02	0.12	0.11	0.19	0.00	0.04	0.12	0.10	0.17	0.00	0.05
04	0.52	0.29	1.86	0.34	0.58	0.48	0.29	1.88	0.29	0.54
05	0.07	0.06	0.13	0.00	0.03	0.06	0.06	0.07	0.00	0.02
06	0.25	0.10	1.09	0.08	0.29	0.52	0.31	1.60	0.24	0.48
07	0.32	0.18	1.43	0.14	0.38	0.43	0.29	1.91	0.21	0.46
09	0.07	0.07	0.13	0.00	0.03	0.08	0.08	0.11	0.00	0.03
10	0.66	0.53	2.04	0.38	0.61	0.87	0.89	2.12	0.50	0.71
12	0.40	0.35	0.97	0.09	0.30	0.79	0.94	1.45	0.25	0.50
14	0.81	0.73	1.89	0.31	0.55	1.07	1.21	2.41	0.43	0.65
15	0.16	0.15	0.17	0.00	0.04	0.18	0.16	0.28	0.00	0.06
17	0.42	0.38	0.58	0.02	0.14	0.47	0.39	0.69	0.04	0.21
20	0.12	0.11	0.13	0.00	0.03	0.12	0.11	0.13	0.00	0.04
23	0.15	0.15	0.18	0.00	0.05	0.15	0.14	0.20	0.00	0.05
25	0.47	0.41	0.85	0.05	0.21	0.60	0.53	1.08	0.08	0.29
26	0.38	0.37	0.38	0.01	0.11	0.40	0.36	0.54	0.02	0.14
45	0.24	0.23	0.23	0.00	0.07	0.24	0.23	0.26	0.01	0.07
OFA	0.31	0.31	0.47	0.02	0.12	0.40	0.42	0.59	0.03	0.18

Table 25: The computed descriptive statistics (Mean, Median, Range, Variance and Deviation) of the arithmetic means of the facial action units data results.

Table 25 presents the descriptive statistics of the data table containing the arithmetic means of the measurements for each frame of a recording. As such the Arithmetic Mean Values table provides a measurement of the central tendency of the comprehensive facial behaviour of each participant, thus offering one perspective over the overall facial muscle activations during the game session. *Table 25* shows that the average of the means tends to be between 0 and 1 which, considering that the software writes 1 as minimum intensity with detected activation and 0 as no detected activation, means that for the majority of time during the game session, the examined muscle groups were not activated. Comparing the means between the two game sessions it is clear that the distinction is minimal, tendentially increasing but not for all action units. Comparing the means to the medians reveals how the two values often do not match, therefore indicating a non-uniform distribution around the centre, probably unbalanced by the presence of outliers. The measurements of distribution seem to vary greatly between the different action units, with some pointing to a more dynamic behaviour and others pointing to mostly immobile muscle groups. Glancing at the overall facial action variable, the data seem to describe an averagely low facial activation, albeit increasing in the second game session setting. The action units will be examined individually in the Discussion session of this manuscript.

Maximum Value Values										
	With Bot					**With Friend**				
AU	**Mean**	**Median**	**Range**	**Vara**	**Dev**	**Mean**	**Median**	**Range**	**Vara**	**Dev**
01	4.21	4.37	2.54	0.61	0.78	4.17	4.47	2.95	0.81	0.90
02	3.62	3.54	3.04	0.91	0.95	3.55	3.59	3.38	0.88	0.94
04	2.55	2.51	3.82	1.10	1.05	2.65	2.60	3.81	1.12	1.06
05	2.34	2.22	3.17	0.80	0.89	2.23	2.25	2.73	0.60	0.77
06	2.30	2.45	3.24	0.72	0.85	2.90	3.02	3.05	0.71	0.84
07	2.73	2.81	4.20	1.15	1.07	3.00	2.99	3.37	0.78	0.89
09	2.37	2.44	4.14	0.95	0.98	2.46	2.19	4.07	1.07	1.04
10	2.59	2.58	3.81	0.78	0.88	3.09	3.13	3.46	0.73	0.85
12	2.75	2.73	2.18	0.39	0.62	3.57	3.73	3.34	0.43	0.65
14	3.07	3.20	3.44	0.67	0.82	3.47	3.51	3.05	0.83	0.91
15	3.55	3.49	3.48	1.51	1.23	3.64	3.72	3.55	1.19	1.09
17	3.18	3.26	2.90	0.67	0.82	3.33	3.18	3.19	0.46	0.68
20	2.26	2.16	3.46	0.68	0.83	2.41	2.28	3.53	0.61	0.78
23	2.80	2.46	3.97	1.27	1.13	3.08	2.73	3.67	1.32	1.15
25	3.97	4.00	3.15	0.65	0.80	4.42	4.51	1.85	0.36	0.60
26	3.35	3.33	2.81	0.60	0.78	3.57	3.47	2.40	0.47	0.69
45	3.50	3.40	2.30	0.30	0.55	3.57	3.67	2.09	0.30	0.55
OFA	3.01	3.13	1.63	0.17	0.41	3.24	3.25	1.85	0.18	0.42

Table 26: The computed descriptive statistics (Mean, Median, Range, Variance and Deviation) of the maximum values of the facial action units data results.

While the Arithmetic Mean table provides expectedly moderate results, the Maximum Value table should instead offer an understanding of the peak measurements of the different facial muscle groups. Some muscle groups seem to have exhibited a peak intensity activation for most of the participants in the experiment, while most of the others seem to have stayed mostly low, albeit showcasing a considerably large distribution. The overall measure is positioned at the middle of the intensity measurement scale, with distribution parameters covering almost the totality of the measurement range, but pointing to a concentration in the middle. Similarly to the previous table (*Table 25*), it seems that for most facial action units, the peak intensity values are higher in the human opponent setting.

Standard Deviation Values						
	With Bot			**With Friend**		
AU	**Mean**	**Median**	**Range**	**Mean**	**Median**	**Range**
01	0.44	0.44	0.56	0.42	0.37	0.58
02	0.32	0.32	0.53	0.31	0.29	0.47
04	0.24	0.24	0.59	0.25	0.21	0.63
05	0.20	0.18	0.46	0.18	0.15	0.24
06	0.31	0.27	0.86	0.47	0.44	1.00
07	0.30	0.33	0.63	0.40	0.36	0.98
09	0.19	0.17	0.36	0.20	0.18	0.32
10	0.32	0.34	0.70	0.41	0.39	0.92
12	0.41	0.40	0.82	0.63	0.65	0.89
14	0.44	0.42	0.77	0.51	0.52	0.73
15	0.29	0.27	0.46	0.32	0.29	0.61
17	0.40	0.36	0.52	0.42	0.36	0.66
20	0.23	0.23	0.29	0.23	0.21	0.31
23	0.31	0.32	0.37	0.31	0.29	0.34
25	0.59	0.53	0.89	0.76	0.73	0.94
26	0.44	0.43	0.39	0.44	0.41	0.50
45	0.48	0.47	0.56	0.46	0.45	0.51
OFA	0.35	0.34	0.30	0.40	0.39	0.34

Table 27: The computed descriptive statistics (Mean, Median, Range, Variance and Deviation) of the standard deviation values of the facial action units data results.

Similarly to the previous tables (*Table 25* and *Table 26*), the Standard Deviation table presents higher distribution measurements for the same action units that present higher values in the previous datasets, thus suggesting the facial muscle groups that were more activated during the experiment, were not activated uniformly during the length experiment. For example, the muscle group responsible for raising the inner brows seems to have reached peak intensity values for the majority of the participants, but much lower mean values and moderately high standard deviation. These numbers indicate that the inner brows were raised to the maximum extent by most participants, but were not kept in that position for long, or were not brought to that position often. Conversely, the brow-lowering muscle group presents higher mean values than the inner bro raiser, but notably lower peak intensity values and even lower distribution parameters, indicating possibly that the muscle group was moderately engaged for more time than the previous one, but possibly never brought to its maximum extent. Furthermore, albeit minimal, there seem to be increases in the standard deviation value in the human opponent setting compared to the computer opponent one.

An additional descriptive analysis that is necessary to conduct on these datasets, is the assessment of the normality of the data distribution. In contrast to the self-report results situation, where the data exhibited ordinal characteristics, and thus naturally deviated from a normal distribution, the present context poses uncertainty due to the continuous nature of the data, and its distribution within a specific range featuring a discontinuity between the values 0 and 1. To evaluate the normality, the Jarque-Bera test is employed, serving as a goodness-of-fit assessment that discerns whether the sample data's skewness and kurtosis align with those expected in a normal distribution. The computation of both skewness and kurtosis is conducted through the embedded functions in Google Sheets, and the formula

$$JB = n\left(\frac{S^2}{6} + \frac{(K-3)^2}{24}\right)$$

Equation 16: Jarque-Bera test formula

where S is the skewness, K is the kurtosis and n is the sample size. The calculations were performed on tables and as a result, it seems that none of the action unit samples are normally distributed. Considering this, the following comparative analysis will have to utilize non-parametric analysis methods.

Comparative Analysis

To individuate and establish the presence of relationships between the action units and the experiment setting it is necessary to conduct another comparative analysis and its respective hypothesis testing. As described above, the current data is continuous and metric in nature, but not normally distributed, as such it is necessary to employ non-parametric methods. Analogously to the self-report results, the processed facial action unit data is extrapolated on the same group of individuals in two different settings and as such the sample is to be considered dependent and paired. The most suitable analysis to perform is again the Wilcoxon signed-rank test, which will be conducted on the arithmetic mean data, the maximum value data and the standard deviation data. For a description of the procedure of the analysis refer to the explanation given in the Comparative Analysis part of the self-reports data results analysis subsection. The null hypothesis is thus

H_0: *In the population, the central tendencies of the two dependent samples are the same.*

and the alternative hypothesis is

H_a: *In the population, the central tendencies of the two dependent samples are unequal.*

The significance level α will be maintained at 0.05. The following table provides the computed results:

	Arithmetic Mean		**Maximum Value**		**Standard Deviation**	
AU	**p-value**	**Effect Size**	**p-value**	**Effect Size**	**p-value**	**Effect Size**
01	0.81	-0.05	0.95	-0.01	0.89	0.03
02	0.87	-0.03	0.51	-0.13	0.89	-0.03
04	0.10	-0.33	0.31	-0.20	0.61	-0.10
05	0.66	-0.09	0.98	0.00	0.51	-0.13
06	$6.43 \cdot 10^{-4}$	-0.76	$9.48 \cdot 10^{-4}$	-0.73	$4.64 \cdot 10^{-4}$	-0.79
07	0.02	-0.51	0.03	-0.47	0.02	-0.47
09	0.22	-0.25	0.10	-0.33	0.34	-0.19
10	$8.89 \cdot 10^{-4}$	-0.74	$1.23 \cdot 10^{-3}$	-0.71	$1.80 \cdot 10^{-3}$	-0.68
12	$4.07 \cdot 10^{-4}$	-0.80	$5.83 \cdot 10^{-4}$	-0.77	$4.07 \cdot 10^{-4}$	-0.80
14	$4.64 \cdot 10^{-4}$	-0.79	$4.64 \cdot 10^{-4}$	-0.79	$5.64 \cdot 10^{-4}$	-0.77
15	0.02	-0.48	0.21	-0.25	0.09	-0.35
17	0.15	-0.29	0.82	0.04	0.45	-0.15
20	0.52	-0.13	0.33	-0.20	0.35	-0.19
23	0.54	-0.12	0.49	-0.14	0.99	0.00
25	$2.47 \cdot 10^{-3}$	-0.66	0.01	-0.54	0.01	-0.59
26	0.52	-0.13	0.51	-0.13	0.95	-0.01
45	0.81	-0.05	0.86	-0.03	0.16	-0.29

OFA	$6.86 \cdot 10^{-4}$	-0.76	$2.98 \cdot 10^{-3}$	-0.65	$1.39 \cdot 10^{-3}$	-0.70

Table 28: The p-values and effect sizes computed using the Wilcoxon Signed-Rank test on the facial action units data results.

It seems that the statistical computations give very consistent results for thee three key indicators. The action units 6, 7, 10, 12, 14, 15 (only for the mean data), 25 and the overall facial action, all have p-value lower than the significance level, leading thus to the rejection of the null hypothesis. There is a statistically significant relationship between the experiment setting and the action units identifying the facial muscle groups having the following functions: Cheek Raiser, Lid Tightener, Upper Lip Raiser, Lip Corner Puller, Dimpler, Lip Corner Depressor and Lips part. Furthermore, there is a statistically significant relationship between the experiment setting and the overall facial action activation, calculated as the arithmetic mean of the activation of the 17 measured facial action units. The relative calculated effect sizes describe the strength of this relationship.

Correlation Analysis

To explore and assess the relationships between the action units and the experiment setting it is necessary to conduct another correlation analysis. Despite the data being non-parametric it was chosen to perform a Pearson/point-biserial correlation analysis to individuate some possible trends and patterns. In addition, Kendall's Tau has been calculated to overcome the barrier of the lack of normal. Analogously to the previous correlation analysis, it was not possible to calculate a rank-biserial or a Spearman rank correlation. The following table provides the computed results:

	Arithmetic Mean		**Maximum Value**		**Standard Deviation**	
AU	***r***	***τ***	***r***	***τ***	***r***	***τ***
01	0.03	-0.11	-0.03	0.01	-0.07	-0.22
02	0.01	-0.14	-0.04	-0.05	-0.05	-0.20
04	-0.04	-0.17	0.05	-0.13	0.02	-0.19
05	-0.12	-0.21	-0.07	-0.16	-0.14	-0.25
06	0.33	0.14	0.34	0.20	0.35	0.16
07	0.13	0.16	0.14	0.08	0.21	0.16
09	0.14	0.16	0.05	0.07	0.08	0.15
10	0.16	-0.04	0.28	0.10	0.24	0.07
12	0.44	0.13	0.55	0.30	0.42	0.22
14	0.21	0.11	0.23	0.13	0.21	-0.02
15	0.23	0.13	0.04	0.08	0.13	0.10

17	0.16	-0.10	0.10	0.11	0.08	-0.09
20	-0.03	-0.26	0.09	0.01	-0.06	-0.25
23	0.09	-0.03	0.12	-0.01	0.03	-0.06
25	0.27	0.21	0.30	0.20	0.34	0.25
26	0.08	-0.13	0.15	0.06	0.02	-0.18
45	-0.03	-0.20	0.07	-0.01	-0.09	-0.25
OFA	0.27	0.05	0.27	0.14	0.25	0.01

Table 29: The computed Pearson correlation coefficients and Kendall rank correlation coefficients of the facial action units data results.

Clearly the Pearson correlation coefficient is not suitable to measure the strength of the found relationship, since, albeit showing results overall consistent with the previous findings, the magnitude is too low in comparison. This is most probably an indicator that the relationship is non-linear. However, it is unclear why also the Kendall correlation coefficient τ, which should not be bound to linear relationships, fails as well to match the effect size computed previously.

In-Game Behaviour Recordings Observations Results

The following section presents the results of the qualitative analysis conducted in the form of observations on the recordings composed of both the gameplay and the facial behaviour. From the careful observational analysis performed on the recordings, 11 facial behaviour patterns emerged.

1. It was noted that the players tighten their lips when in situations of pressure such as when within the game they are in an unsafe condition that could determine the outcome of the game round.
2. It was noted that the players part their lips and open their mouths to express mainly four feelings: disbelief, when failing due to something unexpected or not yet processed; discontent, when unhappy with the outcome of the interaction; internalised fear, such as when being chased by the opponent; and internalised pain, when the avatar is killed and the player emits a sound such as if their actual body just received a physical injury. It was not possible to discern reliably the four classifications, mostly because the four identified feelings and emotions were often presented in a merged and nuanced expression, thus the decision to merge the four into one clearly recognizable face movement. In multiple occurrences with different players, the pattern was also exhibiting instances of spoken language such as the player saying “no!” when parting the lips.
3. It was noted that the players wrinkle their noses to express contempt, anger or disgust, when unsatisfied with their in-game performance, usually resulting in the

death of the playing character. While the expressed emotions are somehow similar to the previous pattern, the two facial movements are clearly utilized in different situations and with different approaches to the failure, the first being more submissive and accepting and the second more judgemental and annoyed.

4. It was noted that the players widen their mouths creating dimples in the cheeks when acknowledging and accepting the outcome of an interaction. For example, this behaviour has often been observed following the disbelief lips parting, indicating that the player has processed what happened within the game. Very often the cheek dimpling was accompanied by a positive nodding movement, and in other individual examples, the dimpling muscles were engaged together with the lips corners raiser, lips corners depressor or lips stretcher muscles.
5. It was noted that the players pucker their lips in an “o” shape when excited for something that has just occurred, or in admiration for a non-trivial successful performance of the opponent. There have been occurrences in individual players of this muscle group activation in conjunction or in alternation with the lips parting pattern previously described.

Furthermore, the players were observed smiling, i.e. the raising of the cheeks and the lips corners, in three different contexts that were thus divided into three distinct patterns as it was possible to identify different feelings and intentions behind the muscle group activation.

6. It was noted that the players can often smile when their character dies within the game. The identified underlying emotions seem to be amusement at the hilarity of the circumstances, such as when dying in unexpected humoristic ways; out of admiration and sportsmanship, as when recognising the skills or decision-making of the opponent in causing the player’s death; or when acknowledging one own mistake.
7. It was noted that players can smile when the opponent’s avatar dies. The identified underlying emotions seem to be amusement at the hilarity of the circumstances, schadenfreude, release of tension or satisfaction.
8. The last observed smile typology doesn’t occur in any of the previously mentioned circumstances, as it occurs only when both players’ avatars are still alive and the round outcome is not at stake. This pattern has mostly been observed in situations where the two players are not attempting to defeat each other according to the game’s goals, but are instead pursuing, even momentarily, some humorous goals, such as intentionally inefficiently chasing each other in a circle, running on the place, or interacting in other non-aggressive means. The underlying feeling has been interpreted as enjoyment of diversion.

9. It was noted that players raise their inner brow in surprise when reacting to an unexpected outcome of some interaction, or in admiration when finding impressive the opponent's performance. This pattern has been observed occurring in conjunction with the lips puckering pattern, easily interpretable as a conjunction of surprise and excitement.
10. It was noted that players lower their brows and corrugate the forehead when confused at the outcome of an interaction, not being able yet to process the causing factors, or when concerned about a situation, such as worrying about being close to failure. This pattern has been observed performed in conjunction with the nose wrinkling pattern showcasing a conjunction of contempt and concern.
11. Lastly, it was noted that almost all players perform facial expressions of any kind in the vast majority of the moments where the tension is released, such as in between rounds, when the fight is momentarily paused. Since the majority of facial movements occur in those moments of tension release, during the tense moments, i.e. the fighting, players exhibit a resting immobile face. Interestingly, this resting face has been shown to change within the computer-controlled agent setting and the human-controlled avatar setting. In the first setting, 24 players exhibited an unexpressive rest face, with no facial muscle group activation, while only 2 players exhibited a resting face in which the lips corner raiser muscles were activated for the first minutes of the game session, thus having a sliming resting face. In the second setting, only 12 players exhibited a resting face without any muscle activation, while the other 14 players engaged the lips corner raiser muscles while in resting face for the first minutes of the game session.

Many other patterns were noted in the participants. It was chosen to omit these other behaviours because occurring only in a few or even singular players, thus more likely individual behaviours, and not common patterns occurring in the wider population. These individual patterns include nodding, closing the eyes upon dying, chin raising, turning the head on one side, shaking the head, lid tightening, mouth tilting, protruding the lips, tilting the head forward, talking, moving the head in rhythm with the game music, covering the mouth with the hand, blinking in rapid succession, chewing, sticking the tongue out, and gesturing. Following are presented the descriptive statistics of the registered occurrences of the observed patterns per individual in the computer-controlled agent setting and the human-controlled avatar setting.

	Bot				**Friend**			
Pattern	**Mean**	**Med**	**Range**	**Dev**	**Mean**	**Med**	**Range**	**Dev**
Lips Tightning (Pressure)	2.0	2	6	1.8	3.0	2	15	3.6

Lips Parting (Disbelief/Pain/Fear)	4.5	4	13	3.8	6.2	6	19	4.1
Nose Wrinkling (Contempt/Disgust)	0.7	0	4	1.1	0.8	0	5	1.2
Dimpling (Acceptance)	1.5	1	7	2.0	1.7	1	7	1.8
Lips Puckering (Excitement/Admiration)	0.8	0	4	1.2	0.6	0	3	1.0
Smiling (Enjoyment of Diversion)	0.3	0	3	0.7	3.3	3	10	2.4
Smiling (Schadenfreude/Satisfaction)	1.6	1	5	1.5	4.0	3	16	3.3
Smiling (Hilarity/Sportsmanship)	1.8	1	7	2.0	4.0	4	9	2.5
Inner Brow Raising (Surprise/Admiration)	2.1	1	9	2.0	2.6	2	8	2.4
Brow Lowering (Confusion/Concern)	1.2	1	5	1.4	1.8	1	11	2.6
Overall	1.7	1	4	0.8	2.8	3	5	1.1

Table 30: The computed descriptive statistics (Mean, Median, Rande and Standard Deviation) of the observed emotional facial behaviours.

From *Table 30* emerges that there is an increase in the occurrences of all patterns between the two experiment settings, however no hypothesis testing will be attempted on this data as the acquired data is not to be considered reliable. This qualitative analysis was conducted by the principal investigator alone, based on their judgement and interpretation. This kind of analysis is subjective by nature, based on the principal investigator's capability to discern and interpret emotions. Arbitrary decisions were needed over the interpretation of the patterns and over the judgement of what was to be considered an occurrence and what was not. For instance, while some participants would exhibit clear and marked emotional facial behaviours, others would be more subtle and more difficult to notice. In all those cases where, for example, the lips corners are raised only slightly, it is a personal decision to consider it or not an occurrence of a smile. Another similar complication was given by the interpretation of what is a prolonged pattern, thus measured as one occurrence, and a repeated pattern in a short time frame, which for consistency was also measured as one. Lastly, measuring facial behaviour through pattern occurrence is not a proper choice considering that the recordings are not of the same length. It is fair to assume that in longer play sessions, participants have more chances to exhibit more patterns. Given all of the above, it is reminded that these qualitative analysis results are intended to be used only as complementary information to the more rigorous results of the facial action units.

Discussion

The following section discusses the results obtained from the analyses, elucidating their implication and addressing the research questions posed at the start of the study. Through an attentive examination of the findings, it should be possible to discern patterns and draw meaningful conclusions, as well as provide a new understanding of the phenomena under investigation.

Subjective Experiences Discussion

In this master thesis experiment, as described in the methods section, the participants were placed in isolated rooms, devoid of human presence except for the participant. The participants were asked to play a game that did not have any form of textual, video or audio communication, and they were asked to play it twice, once with the information of being playing against a computer-controlled agent, and once with the information of being playing against a human-controlled avatar. Through tailored questionnaires administered after each session, the three subjective experiences of immersion, flow and social presence were measured. Following is a simple bar graph providing a visual representation of the average scale values, afterwards the section will enter in detail into each scale.

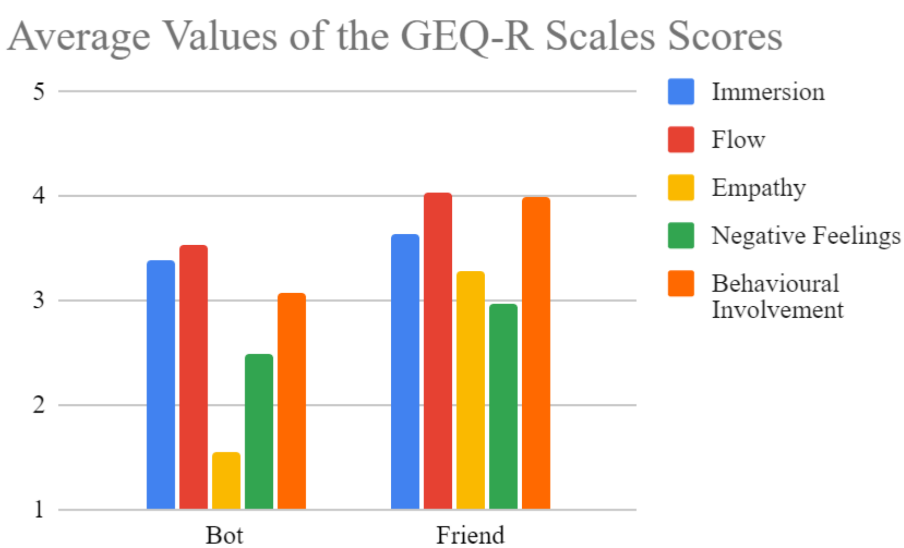


Figure 8: Bar plot visualizing the average values of the five GEQ-R scales.

The participant's perception of immersion, defined as the degree to which the player feels removed from physical reality, was measured through the questionnaire items "I was interested in the game's story", "I thought about other things", "I found it tiresome" and "It felt like a rich experience".

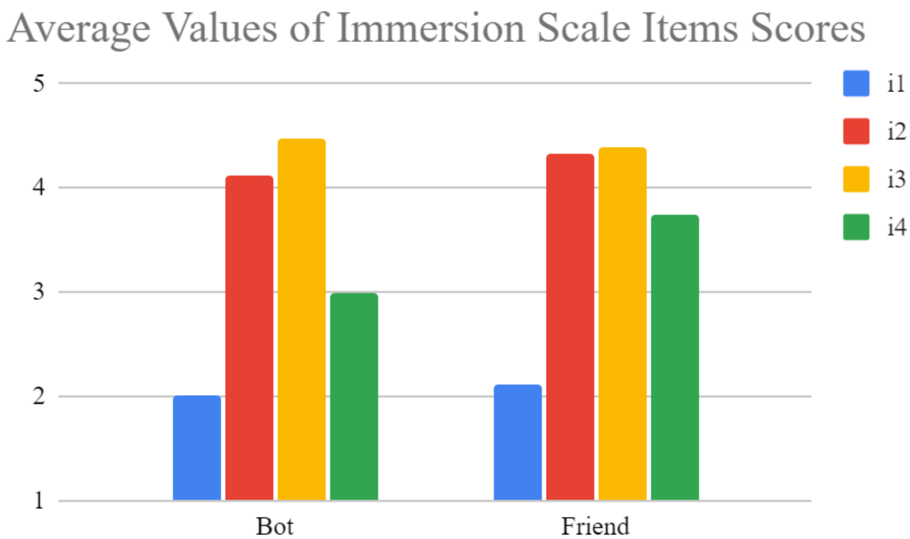


Figure 9: Bar plot visualizing the average values of the items composing the GEQ-R Immersion scale. In order i1 "I was interested in the game's story", i2 "I thought about other things", i3 "I found it tiresome" and i4 "It felt like a rich experience".

The scores of the first item are fairly low, with results interpreted to be affected by game characteristics. Boomerang Fu is a game that doesn't exhibit any explicit narrative, which was interpreted by the participants as lacking a "game's story". The second and third item scores are presented reversed for visual consistency, as the formulation of the questions points to a lack of immersion rather than a presence of the subjective experience. For example, the value of the second item in the computer-controlled setting is actually 1.9, but in the graph, it is given the reserved value of 4.1 representing the hypothetical reversed question "I did not think about other things". The fourth item is the one that presents the greater difference between the two game sessions, a difference that seems to be statistically significant. As discussed previously, within this scale the range is wide for almost all items, as there are outliers in both directions, however, the standard deviation is small (between 0.9 and 1.3), and the median is very close to the mean, indicating that the sample data is mostly clustered around the centre of the distribution. As shown in the scales plot, the average value of the Immersion scale, calculated as the arithmetic mean of the first and fourth item, and of the reversed second and third item, does increase from the first to the second game session. In the conducted Wilcoxon Signed-Ranked, attempting to find a statistically significant inequality in the central tendencies between the two game session settings, the four items resulted in order in a p-value of 0.79, 0.30, 0.59 and 0.02, thus suggesting a statistically significant inequality for the fourth item "It felt like a rich experience". For that one item, the effect size was also large, thus pointing to a strong relationship between the item and the effect size. The chi-squared analysis, searching for a statistically significant relationship between the scores and the experiment settings, provided fairly consistent results, respectively 0.95, 0.56, 0.14 and 0.12, even though none of the items' p-values are statistically significant in this case. The Immersion scale does not reach significance in either of the two tests. Analogously the correlation values are very low when calculated with the point-biserial test, except for the fourth item reaching a medium level of correlation, and even lower in Kendall's tau analysis, never reaching significance, albeit getting closer in the case of the fourth item. Consistently the Immersion scale results in little correlation in both analyses. In conclusion, the evidence collected suggests that there is no significant relationship or correlation between the Immersion scale and the information of playing with a computer-controlled agent or a human-controlled avatar. These results bring further insight into the findings of Ravaja et al. (2006) which connected the opponent type to a subjective experience composed of both social presence and immersion. This study's results show that immersion is not a relevant part of the subjective experience when evaluating the psychological effects of the opponent type. "It felt like a rich experience" is the only question that seems to have connected significantly in one analysis to the experiment settings. Notably, the reliability analysis conducted on the Immersion scales showed that the reliability was unacceptable in the first setting and questionable in the second. It is very possible that the item does not measure the same latent variable as the others, which is instead very probable considering the given definition of immersion. The formulated question seems almost to connect more to another question of the Empathy scale that will be addressed later in this subsection.

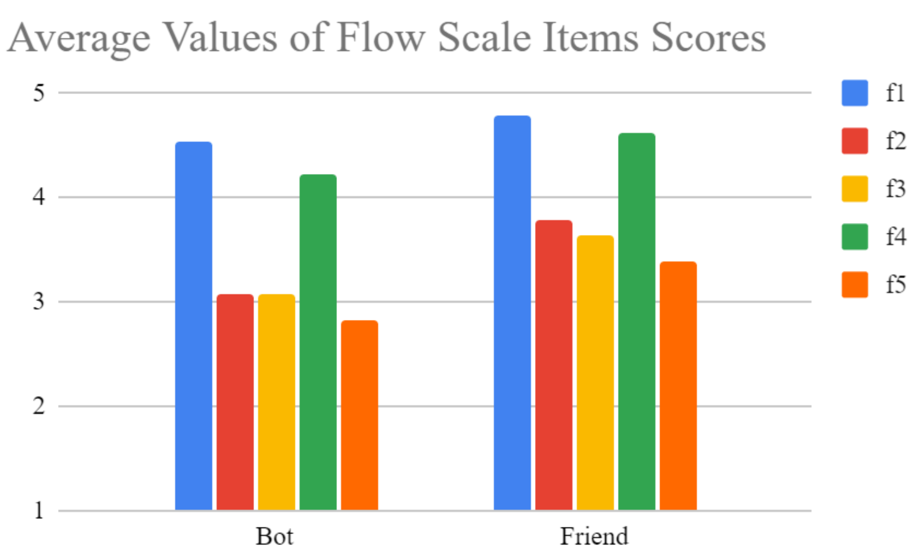


Figure 10: Bar plot visualizing the average values of the items composing the GEQ-R Flow scale. In order f1 "I was fully occupied with the game", f2 "I forgot everything around me", f3 "I lost track of time", f4 "I was deeply concentrated in the game" and f5 "I lost connection with the outside world".

The scores of the Flow scale items seem to range between a medium and a high score, always increasing from the computer-controlled agent setting to the human-controlled avatar one, consistently with the Flow scale average value. Similarly to the Immersion scale, even though the range of answers covers the full 5-point Likert scale width, the measurements of distribution suggest that most of the scores are clustered closer to the centre of the distribution. In the conducted Wilcoxon Signed-Ranked, attempting to find a statistically significant inequality in the central tendencies between the two game session settings, the five items resulted in order in a p-value of 0.07, 0.02, 0.02, 0.049 and 0.02, suggesting the statistical significance for the questions “I forgot everything around me”, “I lost track of time”, “I was deeply concentrated in the game”, “I lost connection with the outside world”, but not for the question “I was fully occupied with the game”. All of these statistically significant inequalities in the central tendency resulted in large effect sizes. The chi-squared analysis, searching for a statistically significant relationship between the scores and the experiment settings, provided fairly consistent results, respectively 0.20, 0.13, 0.30, 0.23 and 0.31, none statistically significant, yet all with large effect sizes. Consistently, the Flow scale resulted in having a statistically significant variation in the central tendency between the two settings, but not a statistically significant relationship with it. Concerning the correlation, both analyses resulted in small correlation values for all the items, yet for the third item the correlation coefficient was small but statistically significant. The correlation values for the Flow scale are low as well with 0.34 in the first case and 0.14 in the second, and no statistical significance was found. In conclusion, the scores of the Flow scales seem to statistically significantly variate slightly between the two settings, but not enough for a statistically significant relationship to emerge, except for the third item “I lost track of time”. It is important to note that the reliability of the Flow scale varied from appropriate to questionable, so it is unclear whether the five items properly measure the same underlying construct, but it is safe to assume that they mostly do.

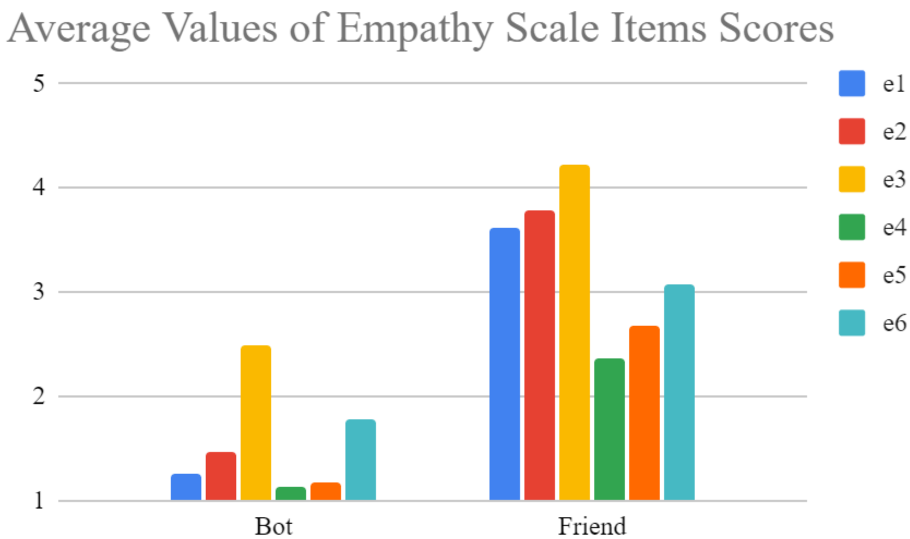


Figure 11: Bar plot visualizing the average values of the items composing the GEQ-R Psychological Involvement (Empathy) scale. In order e1 "I empathized with the other", e2 "I felt connected to the other", e3 "I found it enjoyable to be with the other", e4 "When I was happy, the other was happy", e5 "When the other was happy, I was happy" and e6 "I admired the other"

The scores of the Psychological Involvement (Empathy) scale, one of the three dimensions of social presence, is by far the scale that was most affected by the experiment setting. Of the questions "I empathized with the other", "I felt connected to the other", "I found it enjoyable to be with the other", "When I was happy, the other was happy", "When the other was happy, I was happy" and "I admired the other", all reached statistical significance in the Wilcoxon test and in the Chi-squared test, always with very large effect sizes. All items reached very high correlation in the linear analysis and at least medium correlation in the non-parametric counterpart, all statistically significant. The Empathy scale has thus proved to have a statistically significance relationship with the experiment settings and to be statistically significantly positively strongly affected by it. Notably though, while the scale presents good reliability in the human-controlled avatar setting, it seems to be unacceptable in the computer-controlled agent one. As it has already been discussed this most probably means that the questions measure multiple latent variables that align with each other in social settings, but diverge meaningfully when another social entity is not present. "I empathized with the other", "When I was happy, the other was happy", and "When the other was happy, I was happy" are very difficultly applicable to a non-sentient being, as they do require that the object of interest has the faculty to feel, which seem to be understandably assumed to be lacking in simple game AI processes. Notably the question "I found it enjoyable to be with the other" denotes a different feeling, since, as players have reported, sentience is not a requirement to find the experience with the other entity enjoyable.

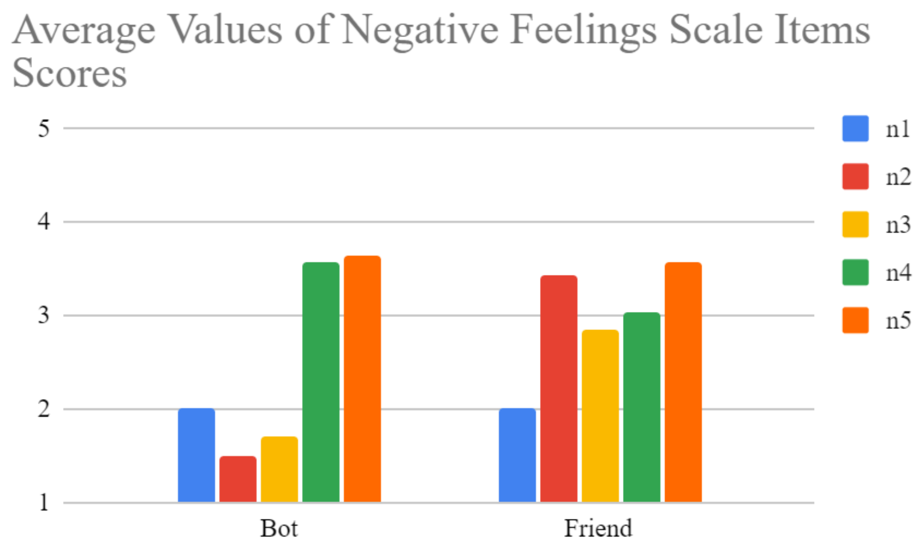


Figure 12: Bar plot visualizing the average values of the items composing the GEQ-R Psychological Involvement (Negative Feelings) scale. In order n1 "I felt jealous about the other", n2 "I influenced the mood of the other", n3 "I was influenced by the other moods", n4 "I felt revengeful" and n5 "I felt schadenfreude (malicious delight)".

Among the items of the Psychological Involvement (Negative Feelings), there is one item that decreases instead of increasing as all the other scales. The decreasing item is "I felt revengeful" which lowers on average from 3.6 to 3, having a wide range of choices among the participants and a medium clustering towards the centre of the distribution. The increasing items are "I influenced the mood of the other" and "I was influenced by the other moods" on which seem to apply the same concluding argument of the last paragraph, i.e. that these questions require a perceived sentience in the other entity. The two stationary items are "I felt jealous about the other" and "I felt schadenfreude (malicious delight)". All the items present a wide range but a clustering towards the centre of the distribution. Interestingly, the decreasing and the stationary items are the ones denoting negative feelings, while the two increasing ones seem more connected to a form of empathy. The reliability of the scales is however very consistent between the two settings, poor in both. In the conducted Wilcoxon Signed-Ranked, attempting to find a statistically significant inequality in the central tendencies between the two game session settings, the five items resulted in order in a p-value of 0.92, 0.0006, 0.005, 0.13 and 0,78, thus having only the two increasing items statistically significant, furthermore very large effect sizes. The chi-squared analysis, searching for a statistically significant relationship between the scores and the experiment settings, provided consistent results, having the same two items statistically significant and no one else. The Negative Feelings scale is statistically significant in the Wilcoxon Signed-Ranked test, thus having a significant difference in the central tendency between the two experiment settings, but not in the chi-squared test, thus not exhibiting a significant relationship. As expected, the two stationary items show no correlation with the setting, the decreasing ones present a little negative correlation, and the increasing ones exhibit a high and medium positive correlation, in both correlation analyses. Both the positive correlations are also significant, making the overall scale present a little positive statistically significant correlation. Interestingly, moving the two items “I influenced the mood of the other” and “I was influenced by the other moods” from the Negative Feelings scale to the Empathy scale and recalculating the reliability coefficient on the new Empathy scale produces a Cronbach’s Alpha of 0.64 (questionable) on the computer-controlled agent setting, higher than both the original scales, and of 0.81 (good) on the human-controlled agent setting, almost equal to the Empathy scale and considerably higher than Negative Feelings scale. It could be that these items measure one of the same latent variables measured in the Empathy scale. An exploratory analysis could shine more light on the underlying structure of the scale but it is outside the scope of this study.

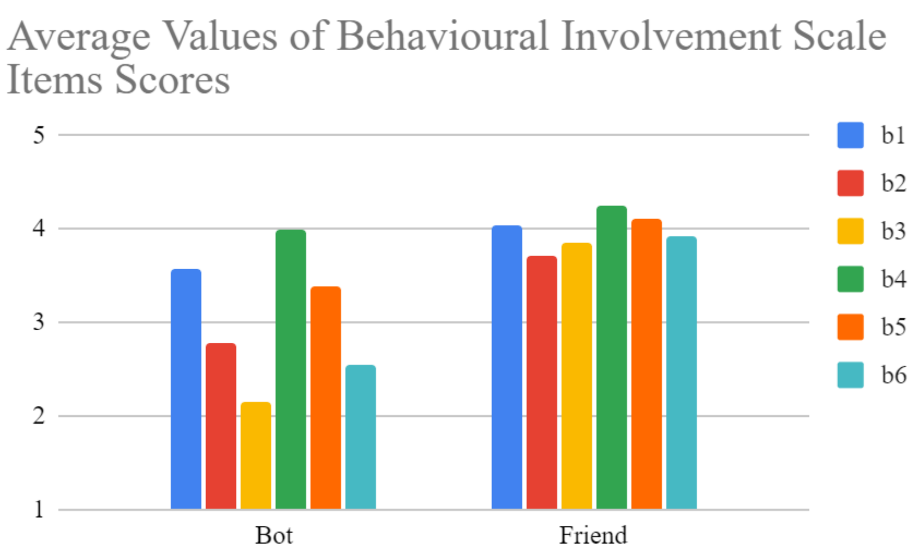


Figure 13: Bar plot visualizing the average values of the items composing the GEQ-R Behavioural Involvement scale. In order b1 “My actions depended on the other’s actions”, b2 “The other's actions were dependent on my actions”, b3 “The other paid close attention to me”, b4 “I paid close attention to the other”, b5 “What the other did affected what I did” and b6 “What I did affected what the other did”.

Finally, the Behavioural Involvement scale seems to have resulted in the most significant results after the Empathy scale. Few items such as b1 “My actions depended on the other’s actions” and b4 “I paid close attention to the other” only slightly increased, while all the others, “The other's actions were dependent on my actions”, “The other paid close attention to me”, “What the other did affected what I did”, and “What I did affected what the other did”, increased moderately from the computer-controlled agent setting to the human-controlled avatar one. Similarly to all the other items in the module, the range is wide but the dispersion is moderate, having the majority of data points within one standard deviation from the centre of the distribution. In the conducted Wilcoxon Signed-Ranked, attempting to a statistically significant inequality in the central tendencies between the two game session settings, the six items resulted in order in a p-value of 0.15, 0.003, 0.0002, 0.119, 0.02 and 0.0009, thus making statistically significant all the increasing items. Among these four, b2 “The other's actions were dependent on my actions”, b3 “The other paid close attention to me” and b6 “What I did affected what the other did”, resulted also in statistical significance in the chi-square test for independence, thus showcasing a statistically significant relationship between the variables and the experiment settings. In both tests, the Behavioural Involvement scale resulted in statistical significance. The three last-mentioned items presented a high correlation with the experiment setting in the point-biserial analysis, while the other items exhibited little to medium correlation. The Kendall’s tau was lower yet consistent, and b2 “The other's actions were dependent on my actions”, b3 “The other paid close attention to me”, b5 “What the other did affected what I did” and b6 “What I did affected what the other did” all showed a statistically significant correlation. There seems to be a statistically significant medium correlation between the experiment settings and the Behavioural Involvement scale. It is interesting how the same reflections on sentience can be made on this scale as well as the previous two, since the most affected items are the ones that require the assumption of interacting with a sentient entity.

In conclusion, according to all the described results of the analysis of the self-report scores, all social presence dimensions seem to have a significant positive correlation with the experiment settings and a significant variation in the central tendency. In addition, two of the three social dimensions also present a significant relationship with the settings. Considering the above, the experiment conditions and the high results in two of the three dimensions in the human-controlled avatar setting, it seems correct to answer the first research question by stating that it is possible to perceive social presence in a multiplayer game experience even with close to no communication. Furthermore, according to the self-report scores, it is possible to experience both the social presence subjective experience and the sense of flow, as both seem to be high and consistently affected by the believed presence of another social entity. No significant correlation was between the perceived social presence dimensions and the Hexad user types. found Finally, an additional final chi-square independence testing was calculated between the differences between the human-controlled avatar setting and the computer-controlled agent one for the Psychological Involvement (Empathy) scale and the Flow scale, resulting in a statistical chi-square value of 14.135 and a critical chi-square value of 12.592, thus confirming the presence of a statistically significant relationship between the two scales. The Pearson correlation of the two differences is only 0.07 while Kendall's Tau is 0.18, with a p-value of 0.18. In conclusion, there is a statistically significant relationship between the variation in the Psychological Involvement (Empathy) values of the participant and the variation of their Flow state between the two settings, but a statistically significant correlation was not found.

Facial Behaviour Discussion

In this master thesis experiment, the participants' facial behaviour and in-game behaviour were recorded during the two play sessions and then analysed quantitively through facial action unit recognition software and analyzed qualitatively through the observation technique. Following are the bar plots providing a visual representation of the average values of each action unit for each chosen indicator, divided into two sets according to whether the action unit resulted having a statistically significant variation in the central tendency or not. Afterwards, it is given a bar plot for the average values of the observed pattern occurrences.

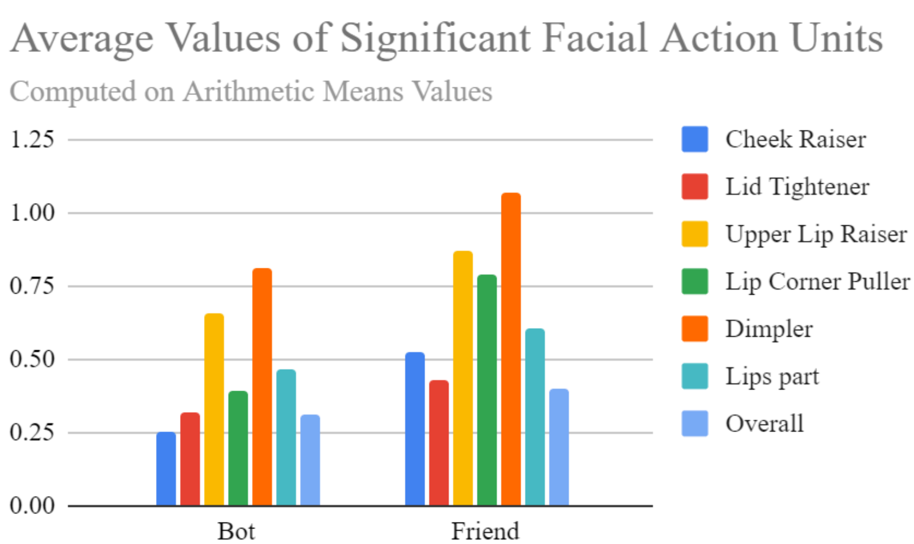


Figure 14: Bar plot visualizing the average values of the arithmetic means of the Facial Action Units that resulted having a significant variation in central tendency.

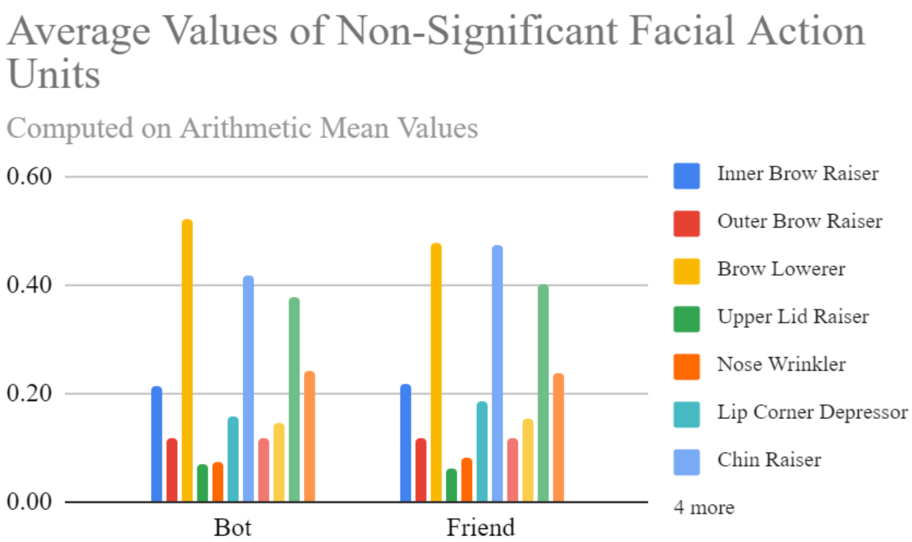


Figure 15: Bar plot visualizing the average values of the arithmetic means of the Facial Action Units that resulted not having a significant variation in central tendency.

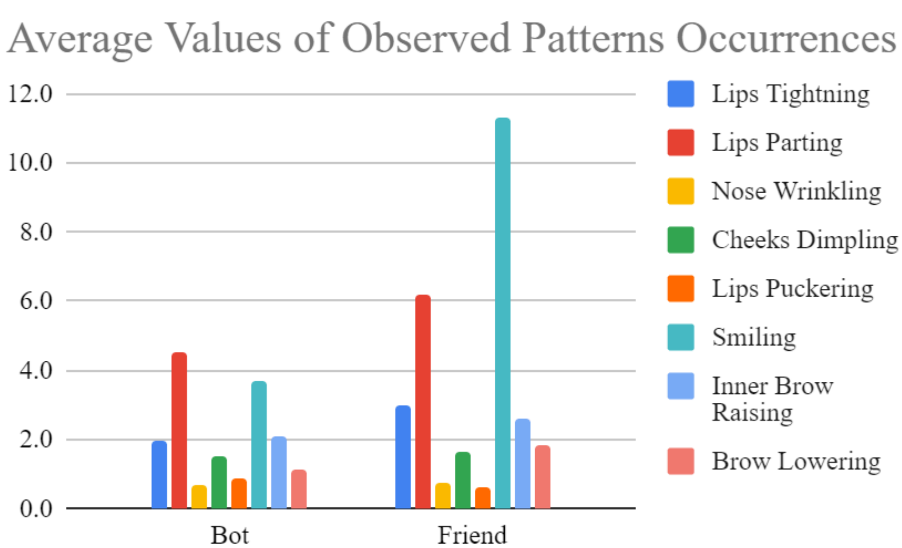


Figure 16: Bar plot visualizing the frequency of the observed emotional facial behaviour patterns.

The action unit 6 and the action unit 12, labelling the activation of the muscle groups orbicularis oculi, pars orbitalis (AU06 Cheek Raiser) and zygomaticus major (AU12 Lip Corner Puller), are muscle groups typically associated with the genuine smile, and considered in the Emotional Facial Action Coding System (EMFACS, Friesen and Ekman 1983) to be indicators of the feeling of happiness. For both these action units, the arithmetic means averagely double from the computer-controlled agent setting to the human-controlled avatar setting, yet the maximum values increase only moderately, and the standard deviation remains fairly low. These results suggest that the amount of smiling was mostly consistent during the game session, and that the smile was more pronounced in the second session. Both the action units proved to have a statistically significant variation of central tendency between the two experiment settings for all three of the chosen indicators, thus confirming that playing with a human opponent leads the participant to smile more when compared to playing with computer-controlled characters. The observations suggest that smiling was more frequent in the second setting and divided into three distinct patterns: smiling at one own's death, smiling at the opponent's death and smiling for the enjoyment of diversion.

Action units 1, 4 and 15, labelling Frontalis, pars medialis (AU01 Inner Brow Raiser), Corrugator supercilii, Depressor supercilii (AU04 Brow Lowerer) and Depressor anguli oris (AU15 Lip Corner Depressor) are the muscle groups typically associated with frowning, thus denoting the sadness feeling in the EMFACS. None of the three showed a notable difference between the two settings in either of the key indicators. It seems that the experiment setting did not affect participants' frowning. In the qualitative analysis, frowning was not a frequent pattern. The lip corner depressor muscles tended to be engaged only together with the cheek dimpler muscles or, for very brief moments together with lip protruding in a few individual behaviours

The just mentioned muscle groups tend to be engaged also when experiencing other typically negative feelings, such as fear (AU01 Inner Brow Raiser, AU02 Outer Brow Raiser, AU04 Brow Lowerer, AU05 Upper Lid Raiser, AU07 Lid Tightener, AU20 Lip Stretcher, and AU26 Jaw Drop), anger (AU04 Brow Lowerer, AU05 Upper Lid Raiser, AU07 Lid Tightner, and AU23 Lip Tightner) and disgust (AU09 Nose Wrinkler, AU15 Lip Corner Depressor and AU17 Chin Raiser). Among these, the only muscle groups that showed any notable difference and any statistically significant variation in central tendency are the orbicularis oculi and pars palpebralis (AU07 Lid Tightener). Occurring only on its own it does not denote the presence of the previously mentioned emotions. It seems that the experiment setting did not affect the participants' facial behaviour concerning negative emotions such as fear, anger and disgust. In the observation analysis, multiple patterns associated with negative feelings emerged. Consistently with the quantitative analysis, the difference in central tendency between the two settings is minimal. Considering the significance found in the lid tightener muscle, the activation has been noticed in some individuals and it was even added as a possible pattern. The pattern was however later discarded because not noticeable in the majority of the participants. The contradiction between the qualitative and quantitative analysis suggests that the movement is subtle, and more easily measurable through software than noticeable through the human eye.

Other muscle groups that seem to have been affected by the experiment setting are the Levator labii superioris (AU10 Upper Lip Raiser), the buccinator (AU14 Dimpler) and the Depressor labii inferioris or the relaxation of Mentalis or Orbicularis oris muscle groups (AU25 Lips Part). All of these show a statistically significant difference in central tendencies between the computer-controlled agent setting and the human-controlled avatar one. The buccinator muscle (AU14 Dimpler) together with zygomaticus major (AU12 Lip Corner Puller) could be an indicator of contempt (Farnsworth 2023) while the levator labii superioris (AU10) is often measured together with the depressor labii inferioris (AU25), for which no specific emotion is given. In the qualitative analysis, all three muscle groups were observed consistently and each was the principal indicator of a pattern. The raising of the upper lips occurred in conjunction with the nose wrinkling pattern, to express contempt and disgust. The dimpling of the cheek muscles occurred in the participants as an expression of acceptance and acknowledgement of the opponent's good performance. The parting of the lips was also noticed as a highly common reaction of internalized pain, internalised fear or disbelief.

Overall, by calculating an arithmetic mean of all the measured facial muscle group activation, it is possible to obtain a generic indicator for the overall facial muscles activation. This final statistic exhibits an increment in the descriptive statistics from the computer-controlled agent setting to the human-controlled avatar one, presenting a statistically significant variation in the central tendency and a small positive correlation. Given the above information, it seems correct to state that there is a relationship between the believed presence of another social entity and a more conspicuous facial emotional behaviour.

Conclusions

This master's thesis was inspired by the observation made by Chet Falizek regarding the playtesting of *Portal 2*, wherein playtesters reportedly expressed appreciation for the comedic aspects of the single-player mode, yet exhibited no facial expression during gameplay, while conversely, they smiled and laughed during the playtesting of the multiplayer game mode, despite the absence of any direct form of communication with other players. These claims sparked curiosity and prompted an inquiry into the physiological effects of flow and social presence, and their reciprocal influence on each other.

Flow is a highly desirable mental state characterized by cognitive efficiency, effortless focus, intrinsic motivation and entailing complete absorption in the activity, with the exclusion of all irrelevant thoughts and emotions. The evidence linking flow to talent development and creative accomplishments has made this state of mind a general academic focus, as well as a key requirement in shaping the notion of challenge within game design. During the state of flow, it is not uncommon to present an immobile face, as usually within activities devoid of social interaction, the mental faculties typically allocated to facial expressions are subdued, as cognitive resources are redirected towards the task at hand. The variance in the presence of facial expressions within the playtesting of the two game modes has been interpreted as an effect of social play.

Social play, happening whenever more than one player interacts with a game at once, has been studied extensively for how it affects game experiences, by facilitating and enlivening social interaction, as well as serving as a major motivator for individuals to engage in digital gaming due to its innate ability to fulfil a fundamental human need for social connection. The last two decades of game research have highlighted the contribution of social interactions in shaping the game experience, providing evidence that social play leads to greater enjoyment and immersion, elicits higher emotional levels, spurs cooperation, lowers hostility and turns failure into enjoyable experiences. Such findings prompted researchers to emphasise the interplay between game content and social interactions and to advocate for recognizing digital gaming technology as social presence technology.

Social presence is the subjective experience of feeling in the presence of another social entity and perceiving their thoughts and emotions, a concept that has long been a primary goal of networked communication systems as it has a reliable augmentator of immersion and enjoyment. Of the three dimensions that compose social presence, digital games seem to be natural vectors of two, copresence and behavioural involvement, and facilitators of one, psychological involvement. Results of games research showcase the impact of social presence on the experience, as it has been observed to increase engagement and positive emotional responses, as well as being in a predictive relationship with empathy, emotional expression and the general synchronization of participants' physiological responses. Studies investigating the influence of winning and losing on perceived social presence among teammates found that, while the game outcome doesn't affect social presence within competitive goal structures, there is a significant relationship between the two in cooperative games, suggesting either that winning enhances cooperative social presence or that low social presence undermines team coordination. As research continues to explore the multifaceted effects of social presence, its significance in shaping game experiences becomes increasingly clear, as well as the importance of further academic inquiry. The Social Presence in Gaming Questionnaire (SPGQ) has been developed for such purpose, to measure social presence in game experience research, capturing the latent variables of psychological involvement (empathy and negative feelings) and behavioural involvement.

Faliszek's claims were thus formalized and restructured into three research questions: "Is it possible to perceive social presence in a multiplayer game experience even with close to no communication?", "Is there a relationship between perceived social presence and a more visible facial emotional behaviour?" and "Is it possible for a player to perceive both social presence and the state of flow?" To attempt to answer these inquiries, an investigation was conducted in which players were isolated in separate rooms with no means of communication and asked to engage in the same digital game initially against a computer-controlled opponent and subsequently against the other isolated players. The experiment was meticulously crafted to optimize the perceivable social presence by incorporating factors documented in research to have a beneficial impact. This included selecting a digital game deemed effective in fostering social presence and explicitly informing participants about the identity of their playing partner (whether computer-controlled or human). Through self-reports, physiological metrics, and observational data, the psychophysiological responses of participants were analyzed in order to reach three goals: (1) identifying a statistically significant difference in the self-reported social presence scores between the computer-controlled agent setting and the human-controlled avatar setting, (2) identifying a statistically significant relationship between the self-reported social presence scores and the correspondent facial action units activation values, and (3) identifying a statistically significant relationship between the self-reported social presence and the flow scores.

The study employed a few scales from the Game Experience Questionnaire (GEQ), the Social Presence in Gaming Questionnaire (SPGQ) and the Hexad-12 user types questionnaire for the measurement of the subjective experiences, which allowed to identify a significant variation in the central tendency of two of the three social presence scales between the computer-controlled agent setting and the human-controlled avatar setting, and a significant positive relationship between the variation of the psychological involvement dimensions and the flow dimension. The participants' facial behaviour has been studied quantitatively and qualitatively through facial action unit extraction software and observation, which allowed the identification of a significant variation of central tendency in the overall facial behaviour, and more specifically with higher values for the muscle groups associated with smiling and with reacting in acceptance or disbelief.

Although the findings are indeed promising, it is important to note that the investigation encountered certain difficulties during both the preparation and the conduction of the experiment. It is hypothesized that conducting the study within a dedicated testing facility and utilizing the digital game described in the methods, i.e. a digital game characterized by a cooperative game goal structure and a gameplay loop necessitating ongoing asymmetric communication, would give even more compelling results. Moreover, it is imperative to remark that participants were informed of the nature of their playing partner, which is a non-negligible strong form of priming. It is unclear whether the observed outcomes stem from the actual presence of another sentient social entity or from participants' belief in interacting with such an entity. Therefore, repeating the experiment with control groups and placebos is essential for elucidating these aspects further.

This study has shed some light on the multifaceted relationship between social presence, facial emotional behaviour and the state of flow in multiplayer gaming experiences. While the findings provide valuable insights into how interaction with another sentient social entity affects subjective experiences and physiological responses, much remains unexplored, as more open questions emerge concerning the role of priming and of the game goal structure and mechanics. More research on these topics could enrich our understanding of the nuanced interplay between social play and game mechanics, thus advancing the academic and practitioner knowledge of game research and game design. In the presentation that motivated this work, Faliszek joked "[i]f you're going to do a comedy game, make it co-op, because otherwise you're gonna have years of misery and self-doubt as you watch it playtested" (2016, 31:25). The analysed results seem to support these claims as the idea of being playing with another human entity leads to smiling more, as well as feeling more engaged, affected and empathetically connected, even devoid of any form of communication. Consequently, a game attempting to tap into social play's potential, may not need to establish traditional means of communication, but could instead endeavour creatively to find uncommon means. Additionally, were the role of priming studied further, the game designing focus on the multiplayer paradigm could shift to making the player feel like they are interacting with other human entities, rather than just entities. Several possible interpretations can be given to the results of this study.

This manuscript concludes with the anecdote of an unexpected behaviour observed among participants: following the termination of the individual experiment session, the vast majority of the participants reached for the playing partner's place of testing, waiting for them to finish filing the questionnaire, to offer a handshake and compliment them about the just played game. Some other participants instead asked for a rematch.

Figure 17: After the conclusion of the experimentations at the second site, the participants asked for one of the utilized laptops to be brought to the gathering area, so that they could play rematches against the experimentation partners, and then against the winners of the rematches.

References

Aitken Ken and Colwyn Trevarthen. 1997. "Self/other organization in human psychological development," *Dev Psychopathol* 9 (4): 653-77. doi: 10.1017/s0954579497001387. PMID: 9449000.

Appel, J., von der Pütten, A., Krämer, N. C., and Gratch, J. 2012. "Does humanity matter? analyzing the importance of social cues and perceived agency of a computer system for the emergence of social reactions during human-computer interaction." *Advanced Human-Computer Interaction* vol. 2012. https://doi.org/10.1155/2012/324694.

Axelsson, A. S., Abelin, Å., Heldal, I., Schroeder, R., and Wideström, J. 2001. "Cubes in the cube: a comparison of a puzzle-solving task in a virtual and a real environment." *CyberPsychology & Behaviour* 4 (2): 279–286. https://doi.org/10.1089/109493101300117956

Aymerich-Franch, L., Karutz, C., and Bailenson, J. N. 2012. "Effects of facial and voice similarity on presence in a public speaking virtual environment," in Proceedings of the International Society for Presence Research Annual Conference (Philadelphia, PA).

Babiloni, Fabio and Laura Astolfi. 2014. "Social neuroscience and hyperscanning techniques: Past, present and future." *Neuroscience & Biobehavioral Reviews* 44: 76-93. https://doi.org/10.1016/j.neubiorev.2012.07.006.

Bansak, Kirk, Jens Hainmueller, Daniel J. Hopkins, and Teppei Yamamoto. 2018. "The Number of Choice Tasks and Survey Satisficing in Conjoint Experiments." *Political Analysis* 26 (1): 112–19. https://doi.org/10.1017/pan.2017.40.

Bartle, Richard. 1996. "Hearts, Clubs, Diamonds, Spades: Players Who Suit MUDs." *Journal of MUD Research* 1 (19).

Bastarache-Roberge, Marie-Christine, Pierre-Majorique Léger, François Courtemanche, Sylvain Sénécal and Marc Fredette. 2015. "Measuring Flow Using Psychophysiological Data in a Multiplayer Gaming Context." In: *Information Systems and Neuroscience. Lecture Notes in Information Systems and Organisation*, eds. Davis, F., Riedl, R., vom Brocke, J., Léger, PM., Randolph, A. vol 10. Springer, Cham. https://doi.org/10.1007/978-3-319-18702-0_25.

Bente, G., Rüggenberg, S., Krämer, N. C., and Eschenburg, F. 2008. "Avatar-mediated networking: increasing social presence and interpersonal trust in net-based collaborations." *Human Communication Research* 34 (2): 287–318. https://doi.org/10.1111/j.1468-2958.2008.00322.x.

Bernhaupt, Regina. 2010. *Evaluating user experience in games: concepts and methods.* New York: Springer.

Biocca, Frank. 1997. "The Cyborg's Dilemma: Progressive Embodiment in Virtual Environments [1]." Journal of Computer-Mediated Communication 3 (2): 0. https://doi.org/10.1111/j.1083-6101.1997.tb00070.x.

Biocca, Frank, and Chad Harms. 2002. "Defining and measuring social presence: Contribution to the networked minds theory and measure." In *Proceedings of the 10th Annual InternationalWorkshop on Presence* (Porto).

Biocca, Frank, Chad Harms, and Judee K. Burgoon. 2003. "Toward a More Robust Theory and Measure of Social Presence: Review and Suggested Criteria." Presence: Teleoperators & Virtual Environments 12 (5): 456–80. https://doi.org/10.1162/105474603322761270.

Biocca, Frank, Chad Harms and Jennifer L. Gregg. 2001. "The Networked Minds Measure of Social Presence: Pilot Test of the Factor Structure and Concurrent Validity." Media Interface and Network Design (M.I.N.D.) Labs.

Blascovich, J., Loomis, J., Beall, A. C., Swinth, K. R., Hoyt, C. L., and Bailenson, J. N. 2002. "Immersive virtual environment technology as a methodological tool for social psychology." *Psychological Inquiry* 13(2): 103–124. https://doi.org/10.1207/S15327965PLI1302_01.

Brown, Emily, and Paul Cairns. 2004. "A grounded investigation of game immersion." In *Extended Abstracts on Human Factors in Computing Systems*, 1297-1300. https://doi.org/10.1145/985921.986048.

Brown, Mark. "How Overcooked's Kitchens Force You to Communicate," YouTube Video, 8:15, Sep 21, 2018, https://www.youtube.com/watch?v=C3M8BvWcJQY&ab_channel=GameMaker%27sToolkit

Bryce, Jo, and Jason Rutter. 2003. "The Gendering of Computer Gaming: Experience and Space." In *Leisure Cultures: Investigations in Sport, Media and Technology*, eds. S. Fleming, and I. Jones, 3-22. Leisure Studies Association.

ByteDance. *CapCut*. V 3.2.0. ByteDance. Windows 10. May 2019

Cairns, Cox, A. L. Day, H. M. Martin, and T. Perryman. 2013. "Who but not where: the effect of social play on immersion in digital games." *International Journal of Human Computer Studies* 71 (11): 1069-1077.

Carr, D., Schott, G., Burn, A., & Buckingham, D. 2004. "Doing game studies: A multi-method approach to the study of textuality, interactivity, and narrative space." *Media International Australia incorporating Culture and Policy* 110: 19-30.

Castellar, Elena Nunez, Kimmo Oksanen, and Jan Van Looy. 2014. "Assessing game experience: Heart rate variability, in-game behavior and self-report measures." Sixth International Workshop on Quality of Multimedia Experience (QoMEX), Singapore, 18-20 September 2014, 292-296. https://doi.org/10.1109/QoMEX.2014.6982334.

Chanel, Guillaume, J. Matias Kivikangas, and Niklas Ravaja. 2012. "Physiological Compliance for Social Gaming Analysis: Cooperative versus Competitive Play." Interacting With Computers 24 (4): 306–16. https://doi.org/10.1016/j.intcom.2012.04.012.

Choi, J. J., and Kwak, S. S. 2017. "Who is this? identity and presence in robot-mediated communication." *Cognitive Systems Research* 43 174–189. https://doi.org/10.1016/j.cogsys.2016.07.006.

Cohen, Jacob. 1988. *Statistical Power Analysis for the Behavioral Sciences (2nd ed.).* Hillsdale, NJ: Lawrence Erlbaum Associates.

Cortese, J., and Seo, M. 2012. "The role of social presence in opinion expression during FtF and CMC discussions." *Communication Research Reports* 29 (1): 44–53. https://doi.org/10.1080/08824096.2011.639913.

Cranky Watermelon. 2020. *Boomerang Fu.* Cranky Watermelon.

Croes, E. A., Antheunis, M. L., Schouten, A. P., and Krahmer, E. J. 2016. "Teasing apart the effect of visibility and physical co-presence to examine the effect of CMC on interpersonal attraction." *Computers in Human Behaviour* 55(A): 468–476. https://doi.org/10.1016/j.chb.2015.09.037.

Csikszentmihalyi, Mihaly. 1975. "Play and intrinsic rewards." *Journal of Humanistic Psychology* 15 (3): 41–63. https://doi.org/10.1177/002216787501500306.

Csikszentmihalyi, Mihaly 1991. *Flow: The Psychology of Optimal Experience: Steps toward Enhancing the Quality of Life.* New York: Harper Collins Publishers.

Csikszentmihalyi, Mihaly. 1996. *Creativity, flow and the psychology of discovery and invention.* New York: Harper Collins.

Csikszentmihalyi, Mihaly 2008. "Mihaly Csikszentmihalyi: Flow, the secret to happiness," YouTube Video, 18:55, Oct 25, 2008, https://www.youtube.com/watch?v=fXIeFJCqsPs&ab_channel=TED.

Csikszentmihalyi, Mihaly, and Barbara Schneider. 2000. *Becoming adult: How teenagers prepare for the world of work.* Basic Books.

Csikszentmihalyi, Mihaly, Kevin Rathunde, Raymond Whalen and Samuel Wong, Maria. 1993. *Talented teenagers: The roots of success and failure.* New York, NY: Cambridge University Press

Cummings, J. J., and Bailenson, J. N. 2016. "How immersive is enough? A meta-analysis of the effect of immersive technology on user presence." *Media Psychology* 19 (2): 272–309. https://doi.org/10.1080/15213269.2015.1015740.

Cummings, J. J., Bailenson, J. N., and Fidler, M. J. 2012. "How immersive is enough? A foundation for a meta-analysis of the effect of immersive technology on measured presence," in Proceedings of the International Society for Presence Research Annual Conference (Philadelphia, PA).

Daft, R. L., and Lengel, R. H. 1986. "Organizational information requirements, media richness and structural design." *Management Science* 32 (5), 554–571. https://doi.org/10.1287/mnsc.32.5.554.

Deci, Edward L. and Richard M. Ryan. 1985. *Intrinsic Motivation and Self-Determination in Human Behavior*. New York: Springer.

Denisova, Alena, A. Imran Nordin, and Paul Cairns. 2016. "The Convergence of Player Experience Questionnaires." In Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '16). Association for Computing Machinery, New York, NY, USA, 33–37. https://doi.org/10.1145/2967934.2968095.

Dixon, Dan. 2011. "Player Types and Gamification." In *CHI 2011 Workshop Gamification Using Game Design Elements in NonGame Contexts*, 12–15.

Drachen, Anders and Alessandro Canossa. 2009. "Analyzing user behavior via gameplay metrics." In Proceedings of the 2009 Conference on Future Play. Association for Computing Machinery, New York, NY, USA, 19–20. https://doi.org/10.1145/1639601.1639613.

Drachen, Anders and Alessandro Canossa. 2011. "Evaluating motion: spatial user behaviour in virtual environments" *International Journal of Arts and Technology* 4 (3). DOI:10.1504/IJART.2011.041483

de Greef, H. Paul, and Wijnand A. IJsselsteijn. 2000. "Social Presence in the PhotoShare Tele-Application." Teleoperators and Virtual Environments - Presence 27-28.

de Kort, Yvonne A.W., Wijnand A. IJsselsteijn, and Karolien Poels. 2007. "Digital Games as Social Presence Technology: Development of the Social Presence in Gaming Questionnaire (SPGQ)." *In 10th Annual International Workshop on Presence (PRESENCE 2007)*, Barcelona, October 25-27 2007, 195-203. Barcelona: StarLab.

de Kort, Y.A.W., and Ijsselsteijn, W. A. 2008. "People, places, and play: a research framework for digital game experience in a socio-spatial context." *ACM Comput. Entertain* 6 (2) http://doi.acm.org/10.1145/1371216.1371221.

Ekman, Inger, Guillaume Chanel, Simo Järvelä, J. Matias Kivikangas, Mikko Salminen, and Niklas Ravaja. 2012. "Social Interaction in Games: Measuring Physiological Linkage and Social Presence" *Simulation & Gaming* 43 (3): 321–38. https://doi.org/10.1177/1046878111422121.

Ekman Paul, and Wallace V. Friesen. 1978. *Facial Action Coding System: A Technique for the Measurement of Facial Movement.* Palo Alto: Consulting Psychologists Press.

Ekman Paul, Wallace V. Friesen and Joseph C. Hager. 2002. *Facial Action Coding System: The Manual on CD ROM*. Salt Lake City: A Human Face.

Emmerich, Katharina, and Maic Masuch. 2016. Game Metrics for Evaluating Social In-Game Behavior and Interaction in Multiplayer Games. https://doi.org/10.1145/3001773.3001793.

Ermi, Laura, and Frans Mäyrä. 2005. “Fundamental Components of the Gameplay Experience: Analyzing Immersion.” In *Changing Views: Worlds in Play, Selected Papers of the 2005 Digital Games Research Association's Second International Conference*, eds. Castell, and Jenson, 15-27. Vancouver: Digital Games Research Association & Simon Fraser University.

Farnsworth, Bryn. 2022. “Facial Action Coding System (FACS) – A Visual Guidebook” *imotions.com*. Last edited September 21 2023. https://imotions.com/blog/learning/research-fundamentals/facial-action-coding-system/

Francescato, D., Porcelli, R., Mebane, M., Cuddetta, M., Klobas, J., and Renzi, P. 2006. “Evaluation of the efficacy of collaborative learning in face-to-face and computer-supported university contexts.” *Computers in Human Behaviour* 22: 163–176. https://doi.org/10.1016/j.chb.2005.03.001.

Freitas-Magalhães, A. 2013. “The science of reading human faces: History, theory, research and assessment.” In *Handbook on facial expression of emotion*, eds. A. Freitas-Magalhães, C. Bluhm and M. Davis, 19–65. Edições Universidade Fernando Pessoa.

Frostling-Henningsson, Maria. 2009. “First-Person Shooter Games as a Way of Connecting to People: “Brothers in Blood”.” *CyberPsychology & Behavior* 12 (5): 557-562. https://doi.org/10.1089/cpb.2008.0345.

Gajadhar, Brian J., Yvonne A. W. de Kort,and Wijnand A. IJsselsteijn. 2008. “Shared Fun Is Doubled Fun: Player Enjoyment as a Function of Social Setting.” In *Fun and Games 2008. Lecture Notes in Computer Science*, ed. Markopoulos, de Ruyter, IJsselsteijn, and W., Rowland. 106-117 Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-88322-7_11.

George, D., and Mallery, P. 2003. *SPSS for Windows step by step: A simple guide and reference 11.0 update (4th ed.).* Boston: Allyn & Bacon

Harris, H., Bailenson, J. N., Nielsen, A., and Yee, N. 2009. “The Evolution of Social Behavior over Time in Second Life.” *Presence: Teleoperators and Virtual Environments* 18 (6): 434–448. https://doi.org/10.1162/pres.18.6.434.

Heldal, I., Schroeder, R., Steed, A., Axelsson, A. S., Spant, M., and Wideström, J. 2005. “Immersiveness and symmetry in copresent scenarios.” in Proceedings of IEEE VR. New York, NY: IEEE.

Herrewijn, L., and Poels, K. 2015. “The impact of social setting on the recall and recognition of in-game advertising”. *Computers in Human Behaviour* 53: 544-555. https://doi.org/10.1016/j.chb.2014.06.012.

Hjortsjö Carl-Herman. 1969. *Man's face and mimic language.* Lund: Studentlitteratur

Hsu, Chin, and Hsi Lu. 2004. "Why Do People Play On-line Games? An Extended TAM with Social Influences and Flow Experience." *Information & Management* 41(7): 853-868. Accessed January 23, 2024. https://doi.org/10.1016/j.im.2003.08.014.

Hudson, Matthew and Paul Cairns. 2014. “Interrogating social presence in games with experiential vignettes.” *Entertainment Computing* 5 (2): 101-114. https://doi.org/10.1016/j.entcom.2014.01.001.

Hudson, Matthew and Paul Cairns. 2016. “The effects of winning and losing on social presence in team-based digital games.” *Computers in Human Behavior* 60: 1-12. https://doi.org/10.1016/j.chb.2016.02.001.

IJsselsteijn, Wijnand A., Yvonne A.W de Kort, Karolien Poels, Audrius Jurgelionis, and Francesco Bellotti. 2007. “Characterising and measuring user experiences in digital games.” In *International conference on advances in computer entertainment technology* 2 (27).

IJsselsteijn, Wijnand A., W.M. Hoogen, van den, C. Klimmt, Yvonne A.W Kort, de, C. Lindley, K. Mathiak, K. Poels, N. Ravaja, M. Turpeinen, P. Vorderer 2008. “Measuring the experience of digital game enjoyment.” In *Proceedings of Measuring Behavior 2008,* Maastricht, 88-89. Noldus.

IJsselsteijn, Wijnand A., Yvonne A.W. de Kort and Karolien Poels. 2013. “The Game Experience Questionnaire.” Technische Universiteit Eindhoven.

Jansz, Jeroen and Martin Tanis. 2007. “Appeal of Playing Online First Person Shooter Games.” *CyberPsychology & Behavior* 10 (1): 133-136. https://doi.org/10.1089/cpb.2006.9981.

Johnson, Daniel and John Gardner. 2010. “Personality, motivation and video games.” In *Proceedings of the 22nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction (OZCHI '10). Association for Computing Machinery*, 276–279. New York, NY, USA. https://doi.org/10.1145/1952222.1952281.

Johnson, Daniel, M. John Gardner, Ryan Perry. 2018. “Validation of two game experience scales: The Player Experience of Need Satisfaction (PENS) and Game Experience Questionnaire (GEQ).” *International Journal of Human-Computer Studies* 118: 38-46. https://doi.org/10.1016/j.ijhcs.2018.05.003.

Johnson, David R., and James C. Creech. 1983. “Ordinal measures in multiple indicator models: A simulation study of categorization error.” *American Sociological Review* 48 (3): 398–407. https://doi.org/10.2307/2095231

Kang, S. H., and Gratch, J. 2014. “Exploring users’ social responses to computer counseling interviewers’ behavior.” *Computers in Human Behaviour* 34: 120–130. https://doi.org/10.1016/j.chb.2014.01.006.

Kiili, Kristian, Sara de Freitas, Sylvester Arnab, Timo Lainema. 2012. “The Design Principles for Flow Experience in Educational Games” *Procedia Computer Science* 15: 78-91. https://doi.org/10.1016/j.procs.2012.10.060.

Kim, D., Frank, M. G., and Kim, S. T. 2014. “Emotional display behavior in different forms of computer mediated communication.” *Computers in Human Behavior* 30: 222–229. https://doi.org/10.1016/j.chb.2013.09.001.

Kivikangas, Mathias J and Niklas Ravaja. 2013 "Emotional Responses to Victory and Defeat as a Function of Opponent," In *IEEE Transactions on Affective Computing* 4 (2): 173-182. doi:10.1109/T-AFFC.2013.12.

Kober, Silvia Erika and Christa Neuper. 2013. “Personality and presence in virtual reality: Does their relationship depend on the used presence measure?” *International Journal of Human-Computer Interaction* 29 (1): 13–25. https://doi.org/10.1080/10447318.2012.668131.

Kohonen-Aho, Laura and Pauli Alin. 2015. “Introducing a Video-Based Strategy for Theorizing Social Presence Emergence in 3D Virtual Environments.” *Presence: Teleoperators and Virtual Environments* 24: 113-31.

Krath, Jeanine, Maximilian Altmeyer, Gustavo F. Tondello, and Lennart E. Nacke. 2023. “Hexad-12: Developing and Validating a Short Version of the Gamification User Types Hexad Scale.” In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA. https://doi.org/10.1145/3544548.3580968.

Kuckartz, Udo, Stefan Rädiker, Thomas Ebert, and Julia Schehl. Statistics. *A Comprehensible Introduction*. VS Verlag.

Labonté-LeMoyne, Élise, Pierre-Majorique Léger, Beverly Resseguier, Marie-Christine Bastarache-Roberge, Marc Fredette, Sylvain Sénécal, and François Courtemanche. 2016. “Are We in Flow Neurophysiological Correlates of Flow States in a Collaborative Game.” *CHI Conference Extended Abstracts on Human Factors in Computing Systems*, *San Jose, 2016,* 1980-1988 https://doi.org/10.1145/2851581.2892356.

Lankoski, Petri and Staffan Björk. 2015. “Formal Analysis of Gameplay.” In *Game Research Methods: An Overview, eds.* Petri Lankoski and Staffan Björk, 23-35. ETC Press.

Law, Effie L.C., Florian Brühlmann, and Elisa D. Mekler. 2018. “Systematic Review and Validation of the Game Experience Questionnaire (GEQ) - Implications for Citation and Reporting Practice.” In Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '18). Association for Computing Machinery, New York, NY, USA, 257–270. https://doi.org/10.1145/3242671.3242683.

Lee, K. M. 2004. “Presence, explicated.” *Communication Theory* 14 (1), 27–50. https://doi.org/10.1111/j.1468-2885.2004.tb00302.x.

Lee, K. M., and Nass, C. 2004. “The multiple source effect and synthesized speech.” *Human Communication Research* 30 (2): 182–207. https://doi.org/10.1111/j.1468-2958.2004.tb00730.x.

Lee, K. M., Park, N., and Song, H. 2005. "Can a robot be perceived as a developing creature? Effects of a robot's long-term cognitive developments on its social presence and people's social responses toward it." *Human Communication Research* 31 (4), 538–563. https://doi.org/10.1111/j.1468-2958.2005.tb00882.x.

Lee, K. M., Jung, Y., Kim, J., and Kim, S. R. 2006. "Are physically embodied social agents better than disembodied social agents? The effects of physical embodiment, tactile interaction, and people's loneliness in human–robot interaction". *International Journal of Human Computer Studies* 64 (10), 962–973. https://doi.org/10.1016/j.ijhcs.2006.05.002.

Léger, Damien, François Beck, Jean-Baptiste Richard, Fabien Sauvet and Brice Faraut. 2014. "The Risks of Sleeping "Too Much". Survey of a National Representative Sample of 24671 Adults (INPES Health Barometer)." *PLOS ONE* 9 (9). https://doi.org/10.1371/journal.pone.0106950.

Lim, Sohye, and Byron Reeves. 2010. "Computer agents versus avatars: Responses to interactive game characters controlled by a computer or other player." *International Journal of Human-Computer Studies* 68 (1-2): 57-68. https://doi.org/10.1016/j.ijhcs.2009.09.008.

Lin, J. 2013. The science behind shaping player behavior in online games. Presented at GDC 2013 http://gdcvault.com/play/1017940/The-Science-Behind-Shaping-Player

Marczewski, Andrzej. 2015. "Hexad: A player-type framework for gamification design." *Gamified UK.*

Moneta, Giovanni B., & Csikszentmihalyi, Mihaly 1999. "Models of concentration in natural environments: A comparative approach based on streams of experiential data". *Social Behavior and Personality: An International Journal,* 27 (6), 603–638. https://doi.org/10.2224/sbp.1999.27.6.603.

Moreno, R., and Mayer, R. E. 2004. "Personalized messages that promote science learning in virtual environments." *Journal of Educational Psychology* 96 (1): 165–173. https://doi.org/10.1037/0022-0663.96.1.165.

Nacke, Lennart and Anders Drachen. 2011. "Towards a Framework of Player Experience Research" (Pre-print).

Nacke, Lennart E., Sophie Stellmach, and Craig A. Lindley. 2010. "Electroencephalographic Assessment of Player Experience: A Pilot Study in Affective Ludology." *Simulation and Gaming* 42 (5). https://doi.org/10.1177/1046878110378140.

Nacke, Lennart E., Chris Bateman and Regan L. Mandryk. 2014. "BrainHex: A neurobiological gamer typology survey." *Entertainment Computing* 5 (1): 55-62. https://doi.org/10.1016/j.entcom.2013.06.002.

Nakamura, Jeanne, & Csikszentmihalyi, Mihaly 2002. “The concept of flow.” In *Handbook of positive psychology*, eds. C. R. Snyder and S. J. Lopez, 89–105. Oxford University Press.

Oh, Catherine S., Jeremy N. Bailenson, and Gregory F. Welch. 2018. "A Systematic Review of Social Presence: Definition, Antecedents, and Implications." *Frontiers in Robotics and AI 5*. https://doi.org/10.3389/frobt.2018.00114.

Pan, X., Gillies, M., and Slater, M. 2008. “The impact of avatar blushing on the duration of interaction between a real and virtual person,” in Proceedings of the 11th Annual InternationalWorkshop on Presence October 26–28 (Edinburgh).

Peng, Wei and Gary Hsieh. 2012. “The influence of competition, cooperation, and player relationship in a motor performance centered computer game.” *Computers in Human Behaviour* 28 (6): 2100-2106. https://doi.org/10.1016/j.chb.2012.06.014.

Perttula, Arttu, Kristian Kiili, Antero Lindstedt, and Pauliina Tuomi. 2017. “Flow Experience in Game Based Learning – a Systematic Literature Review”. *International Journal of Serious Games* 4 (1). https://doi.org/10.17083/ijsg.v4i1.151.

Poels, Karolien, Yvonne De Kort, and Wijnand IJsselsteijn. 2007. “Game Experience Questionnaire: development of a self-report measure to assess the psychological impact of digital games.” Eindhoven: Technische Universiteit Eindhoven.

Ratan, R. A., and Hasler, B. 2009. “Self-presence standardized: Introducing the self-presence questionnaire (SPQ),” in Proceedings of the 12th Annual International Workshop on Presence (Los Angeles, CA).

Ravaja, Niklas, Timo Saari, Marko Turpeinen, Jari Laarni, Matti Salminen, and Matias Kivikangas. 2006. “Spatial Presence and Emotions During Video Game Playing: Does It Matter With Whom You Play?” *Presence: Teleoperators & Virtual Environments* 15 (4): 381–92. https://doi.org/10.1162/pres.15.4.381.

Ryan, Richard M. and Deci, Edward L. 2000. “Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being.” *American Psychologist* 55 (1): 68–78. https://doi.org/10.1037/0003-066X.55.1.68.

Ryan, Richard M., Scott S. Rigby and Andrew Przybylski. 2006. “The Motivational Pull of Video Games: A Self-Determination Theory Approach.” *Motiv Emot* 30, 344–360. https://doi.org/10.1007/s11031-006-9051-8.

Schouten, D. G. M. 2014. "Shared Experience: The influence of the in-game social connection on the player experience in digital gaming. tue. nl."

Seligman, Martin E. P., Tracy A. Steen, Nansook Park, and Christopher Peterson. 2005. “Positive Psychology Progress: Empirical Validation of Interventions.” *American Psychologist* 60 (5): 410–421. https://doi.org/10.1037/0003-066X.60.5.410.

Schroeder, R., Steed, A., Axelsson, A. S., Heldal, I., Abelin, Å.,Wideström, J. et al. 2001. “Collaborating in networked immersive spaces: as good as being there together?”

Computers & Graphics. 25 (5): 781–788. https://doi.org/10.1016/S0097-8493(01)00120-0.

Sherry, John, Bradley Greenberg, Kristen Lucas, and Kenneth Lachlan. 2006. "Video game uses and gratifications as predictors of use and game preference." *International Journal of Sports Marketing and Sponsorship* 8: 213-224

Short, John, Ederyn Williams, and Bruce Christie. 1976. *The Social Psychology of Telecommunications*. New York, NY: John Wiley. https://ci.nii.ac.jp/ncid/BA22930972.

Siriaraya, P., and Ang, C. S. 2012. "Age differences in the perception of social presence in the use of 3D virtual world for social interaction." *Interact. Comput* 24: 280–291. https://doi.org/10.1016/j.intcom.2012.03.003.

Slater, Mel, and Sylvia Wilbur. 1997. "A Framework for Immersive Virtual Environments (FIVE): Speculations on the Role of Presence in Virtual Environments." *Presence: Teleoperators & Virtual Environments* 6: 603-616. https://doi.org/10.1162/pres.1997.6.6.603.

Steuer, J. 1992. "Defining virtual reality: dimensions determining telepresence." *Journal of Communication* 42 (4), 73–93. https://doi.org/10.1111/j.1460-2466.1992.tb00812.x.

Sullivan Gail M., Anthony R. Artino Jr. 2013. "Analyzing and interpreting data from Likert-type scales." *J Grad Med Educ* 5 (4): 541-2. doi: 10.4300/JGME-5-4-18.

Swann, Christian, Richard J. Keegan, David Piggott, Lee Crust. 2012. "A systematic review of the experience, occurrence, and controllability of flow states in elite sport." *Psychology of Sport and Exercise* 13 (6): 807-819. https://doi.org/10.1016/j.psychsport.2012.05.006.

Sweetser, Penelope and Peta Wyeth. 2005. "GameFlow: a model for evaluating player enjoyment in games." *Computers in Entertainment* 3 (3): 3. https://doi.org/10.1145/1077246.1077253.

Baltrušaitis, Tadas, Amir Zadeh, Yao Chong Lim, and Louis-Philippe Morency. 2018. "OpenFace 2.0: Facial Behavior Analysis Toolkit" *IEEE International Conference on Automatic Face and Gesture Recognition*.

Team17 and Ghost Town Games. 2018. *Overcooked! 2*. Wakefield: Team17.

Thorsteinsson, Elinar B., and Jack E. James. 1999. "A meta-analysis of the effects of experimental manipulations of social support during laboratory stress." *Psychology & Health* 14 (5), 869–886. https://doi.org/10.1080/08870449908407353.

Tondello, Gustavo F., Rina R. Wehbe, Lisa Diamond, Marc Busch, Andrzej Marczewski, and Lennart E. Nacke. 2016. "The Gamification User Types Hexad Scale." In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '16). Association for Computing Machinery*, 229–243. New York, NY, USA. https://doi.org/10.1145/2967934.2968082.

Valve. 2011. *Portal 2*. Bellevue, Washington: Valve.

Voida, Amy, Sheelagh Carpendale, and Saul Greenberg. 2010. “The individual and the group in console gaming.” In *Proceedings of the 2010 ACM conference on Computer supported cooperative work (CSCW '10). Association for Computing Machinery*, 271-380. New York, NY, USA. https://doi.org/10.1145/1718918.1718983.

Verhagen, T., van Nes, J., Feldberg, F., and van Dolen,W. 2014. “Virtual customer service agents: using social presence and personalization to shape online service encounters.” *Journal of Computer-Mediated Communication* 19 (3): 529–545. https://doi.org/10.1111/jcc4.12066.

Waddell, Julia C., and Wei Peng. 2014. "Does It Matter with Whom You Slay? The Effects of Competition, Cooperation and Relationship Type among Video Game Players." *Computers in Human Behavior* 38: 331-338. Accessed January 23, 2024. https://doi.org/10.1016/j.chb.2014.06.017.

Wallner, Günter and Simone Kriglstein. 2015. “An Introduction to Gameplay Data Visualization.” In *Game Research Methods: An Overview*, eds. P. Lankoski & S. Björk, 231–250. ETC Press. http://hdl.handle.net/20.500.12708/28619.

Walther, Joseph B. 1992. “Interpersonal effects in computer-mediated interaction: A relational perspective.” *Communication Research* 19 (1): 52–90. https://doi.org/10.1177/009365092019001003.

Walther, J. B. 1996. “Computer-mediated communication: Impersonal, interpersonal, and hyperpersonal interaction.” *Communication Research* 23 (1): 3–43. https://doi.org/10.1177/009365096023001001.

Weibel, David, Bartholomäus Wissmath, Stephan Habegger, Yves Steiner, and Rudolf Groner. 2008. “Playing online games against computer- vs. human-controlled opponents: Effects on presence, flow, and enjoyment.” *Computers in Human Behavior* 24 (5): 2274–2291. https://doi.org/10.1016/j.chb.2007.11.002

Wolpaw, Eric, and Faliszek, Chet. “Portal 2: Creating a Sequel to a Game That Doesn't Need One,” YouTube Video, 59:26, Sep 1, 2016, https://www.youtube.com/watch?v=BYFvwbby2YM&t=1817s

Zhan, Z., and Mei, H. 2013. “Academic self-concept and social presence in face-to-face and online learning: Perceptions and effects on students’ learning achievement and satisfaction across environments.” *Computers & Education* 69: 131–138. https://doi.org/10.1016/j.compedu.2013.07.002.

Appendix I

Informed Consent Form

Date of Experiment

Informed Consent Form

Title of Study: Study on the Relationship between Social Presence, Facial Action Units and Flow in Multiplayer Game Experiences

Principal Investigator: Michele Fanelli
Game Studies and Engineering
Alpen-Adria-Universität Klagenfurt
Universitätsstraße 65/67, 9020 Klagenfurt am Wörthersee

1. Taking part in the study

You are invited to voluntarily participate in a research study conducted by the investigator. This study aims to measure and assess the effects and the relationships between social presence, facial action units and flow on players partaking in a multiplayer game experience.

2. Use of information in the study

During the course of the study, the investigator will collect the following data.

1. **Recording of in-game behaviour:** Participants will engage in a multiplayer game experience in two distinct sessions. The first session will be played with a computer-controlled bot, and the second session will be played with a known person. The game experience will be recorded through screen capture. The investigator will make observations on the participants' recorded in-game behaviour during both sessions. This may include noting collaborative strategies, communication patterns, and other observable behaviours relevant to the study.
2. **Recordings of facial behaviour:** Throughout the game sessions, participants' facial behaviour will be recorded through an embedded webcam. The recordings will be processed using OpenFace software, a facial behaviour analysis tool. OpenFace does not store in any way the recordings. This software will be applied to analyze and extract relevant facial behaviour data such as the movements of the participants' eyebrows, eyes, mouth and head pose. The statistical analysis will be performed only on the numerical data extracted from the recordings.
3. **Measurements of Subjective Experience:** After each game session, participants will be asked to provide self-reports on their experiences. This will include assessments of gaming personality, social presence, flow, and immersion. Participants will be prompted to reflect on their feelings and perceptions immediately following each gaming session.

The facial behaviour recordings will be stored securely on the investigator's personal computer, and will be deleted at the completion of the academic work. The recordings of the in-game behaviours, the self-reports and the numerical data processed out of the facial recordings, will all be anonymized through code number labelling. The anonymized numerical data may be published in the final academic work.

3. Appliance to the General Data Protection Regulation

In accordance with the General Data Protection Regulation law that went into effect May 25, 2018, this study processes data lawfully (Art. 6), and doesn't process any special categories of data (Art. 9). The participant has the right to access (Art. 15) and to ask for the erasure of all or any portion of the recordings (Art. 17).

4. Voluntary Participation and Consent

In signing this form the participant agrees to consent to the acquisition and processing of the personal data specified in this document, for research purposes only. The participant claims to have understood and read the study information provided in the document, or that the information was read to them. The participant was able to ask questions about the study and the questions were answered to the participant's satisfaction.

Upon signing the participant consents voluntarily to take part in the study and understands that they can refuse to answer questions and withdraw from the study at any time, without having to give a reason.

Participant's Name: ________________________ Date: ________________

Participant's Signature: _______________________

Researcher's Certification:
I confirm that I have explained the nature and purpose of the study to the participant and answered any questions they may have had.

Researcher's Name: ______Michele Fanelli_ Date: ________________

Researcher's Signature: ________________________

Appendix II

Raw Self-Report Data Results

	1a	1b	2a	2b	3a	3b	4a	4b	5a	5b	6a	6b	7a	7b	8a	8b	9a	9b	10a	10b	11a	11b	12a	12b	13a	13b	14a	14b
p1	7	5	6	5	6	6	7	6	7	7	7	7	7	7	6	7	5	6	5	6	7	6	7	7	7	5	5	7
p2	5	6	6	5	6	7	7	5	7	7	7	6	7	7	6	7	5	6	6	5	6	6	6	7	7	6	6	7
s1	4	5	7	5	5	6	6	5	5	7	7	5	5	6	5	6	4	5	6	5	6	6	6	3	4	3	5	6

s2	2	6	7	5	5	5	6	5	4	7	6	6	5	6	4	5	4	4	7	5	5	5	6	3	5	5	5	6
a1	6	7	6	6	7	5	5	6	6	6	5	7	7	5	6	5	6	7	5	7	5	5	6	4	3	2	6	5
a2	7	7	5	4	7	5	6	7	7	4	6	7	7	7	5	6	6	6	5	7	6	6	7	7	6	6	6	6
r1	7	7	4	2	6	5	5	6	6	5	6	7	7	3	3	5	4	5	7	6	6	5	2	6	5	6	5	5
r2	7	7	4	3	5	5	5	5	6	6	7	3	6	4	4	5	4	3	6	6	6	6	3	6	3	6	6	7
m1	5	4	5	4	6	4	6	6	7	2	4	5	7	5	5	6	6	7	5	7	4	4	2	4	7	7	6	6
m2	7	6	4	6	5	5	6	6	7	2	5	5	7	3	6	6	7	7	3	7	6	5	3	6	7	7	6	6
d1	3	2	2	4	3	5	4	4	5	4	3	3	5	3	2	1	2	5	1	2	4	2	2	4	4	3	2	2
d2	2	2	2	2	4	4	5	5	4	4	3	1	1	2	2	1	2	5	2	3	3	3	2	4	4	5	2	2
i1b	1	4	2	2	2	1	3	4	1	1	2	4	1	1	2	3	1	2	2	1	4	1	2	1	1	5	1	1
f1b	5	5	4	5	4	4	5	5	5	3	5	5	5	4	4	5	5	5	4	5	5	5	4	4	4	5	3	5
i2b	1	1	2	2	1	4	1	2	3	5	2	1	1	2	1	4	1	4	1	1	1	1	3	1	3	1	2	1
i3b	4	2	1	1	1	1	1	2	2	2	3	1	4	1	1	2	1	2	1	1	1	1	1	1	1	1	2	1
f2b	5	5	3	2	3	2	2	2	4	2	4	2	4	2	4	4	2	3	2	3	4	5	2	2	2	4	3	4
f3b	5	5	2	3	2	2	2	4	4	2	3	4	4	2	4	5	2	5	2	4	3	1	1	2	5	1	2	5
f4b	5	5	4	4	3	4	5	5	4	2	5	5	5	4	4	5	5	5	4	5	4	5	3	2	3	5	3	5
i4b	4	4	5	4	4	3	3	4	2	1	4	3	3	1	3	2	3	2	3	5	2	3	3	1	2	3	2	5
f5b	4	5	2	2	3	3	2	3	4	1	3	3	3	2	3	5	2	4	2	4	4	4	1	2	1	3	3	1
e1b	1	2	1	2	1	2	2	1	1	2	1	1	1	1	1	2	1	1	1	1	1	1	2	1	1	1	1	1
b1b	5	1	4	4	5	3	5	2	4	3	4	3	4	1	4	4	5	5	3	1	3	5	3	5	1	5	3	5
b2b	5	4	2	3	3	2	4	2	2	2	3	3	4	2	3	3	1	3	4	1	2	3	4	5	1	1	3	3
e2b	1	2	1	2	1	3	2	1	1	1	3	1	4	1	1	1	1	2	1	1	1	1	2	1	1	1	1	2
b3b	5	1	1	3	2	1	3	2	1	2	2	2	3	4	1	2	1	3	3	1	1	3	2	5	1	1	2	2
b4b	5	5	4	4	4	3	5	4	4	4	5	4	5	4	3	5	5	5	2	4	3	4	3	5	1	5	4	3
n1b	1	1	2	1	3	4	3	3	1	3	5	3	2	1	1	3	1	1	1	1	1	1	3	1	1	5	2	1
e3b	4	4	4	2	2	3	5	4	2	2	4	1	4	1	1	1	3	3	3	1	2	1	2	1	1	5	1	3
e4b	1	1	2	1	1	1	1	1	1	4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
e5b	1	4	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1
n2b	1	4	3	1	1	2	1	1	1	3	1	1	1	1	1	5	1	2	1	1	1	1	1	1	1	1	1	2
n3b	1	4	4	4	1	3	1	1	1	3	2	1	1	1	1	5	1	2	1	1	1	1	1	1	1	1	1	2
e6b	1	1	2	1	2	3	3	3	1	1	4	2	1	1	1	2	1	4	1	4	1	1	1	1	1	1	3	2
b5b	5	4	4	4	5	3	5	3	4	4	4	2	3	2	4	4	4	5	2	1	2	2	2	5	1	4	3	4
b6b	5	1	2	3	2	2	4	2	1	4	3	3	3	2	4	4	2	4	2	1	2	2	1	5	1	1	2	3
n4b	5	4	1	2	3	4	5	4	4	3	5	4	3	1	4	5	3	4	3	4	3	3	4	5	1	5	4	4
n5b	5	5	5	1	2	5	5	5	5	1	2	4	3	1	4	4	3	3	4	5	3	3	5	5	3	4	5	2
i1f	2	4	2	1	1	1	2	4	1	5	2	3	1	1	1	3	1	2	3	1	4	1	2	1	2	5	1	2
f1f	5	5	4	5	5	5	5	5	5	5	5	5	5	5	5	5	4	5	5	5	5	5	4	4	4	5	4	5
i2f	1	1	2	2	1	2	1	2	1	2	3	1	4	2	1	2	2	5	1	1	1	1	2	1	1	1	1	2
i3f	1	1	1	1	2	1	1	1	3	1	5	1	4	1	1	1	1	4	1	1	1	3	1	1	1	1	3	1

f2f	5	4	2	2	5	4	4	4	4	5	4	5	4	3	2	5	3	5	3	4	5	5	2	4	2	2	4	5
f3f	5	5	4	4	4	4	5	4	4	1	2	4	5	3	4	5	4	5	3	4	4	1	2	2	4	2	3	5
f4f	5	5	4	4	5	5	5	5	4	5	4	5	5	4	5	5	5	4	5	5	5	5	3	4	4	5	4	5
i4f	5	5	4	4	4	4	5	4	3	5	2	4	2	4	3	4	3	2	4	5	3	4	4	1	4	5	3	5
f5f	5	5	3	2	5	4	4	2	4	4	4	3	3	2	4	4	3	4	3	4	4	5	2	1	4	4	2	1
e1f	4	5	4	4	3	4	5	3	4	5	5	2	4	1	3	5	3	4	4	4	3	4	4	1	3	3	4	3
b1f	5	5	4	5	5	4	5	4	4	1	5	2	3	5	4	5	5	5	2	1	3	5	4	5	3	5	4	5
b2f	5	2	4	5	5	4	5	3	3	3	4	4	3	4	4	5	3	5	3	1	3	3	4	5	2	4	4	4
e2f	4	3	4	4	4	4	5	3	3	4	4	3	4	4	2	5	3	4	4	4	4	5	4	1	4	5	4	4
b3f	5	2	4	4	4	4	5	3	2	5	5	4	4	4	4	5	2	5	4	2	3	4	4	5	3	4	4	4
b4f	5	4	4	4	4	4	5	3	4	5	5	4	5	5	4	5	5	4	4	4	3	5	4	5	1	5	4	5
n1f	2	2	1	1	3	3	2	3	1	1	1	3	2	1	1	1	1	4	1	1	2	5	1	1	1	5	3	3
e3f	4	5	5	5	5	4	5	4	4	4	5	4	4	5	3	5	4	3	4	5	4	5	4	1	4	5	4	4
e4f	4	4	1	2	3	3	4	2	1	3	4	1	2	1	1	3	3	4	2	3	1	1	2	1	3	1	3	3
e5f	2	5	1	2	3	4	4	2	2	4	4	1	2	1	3	4	3	5	3	4	1	1	3	1	3	1	3	3
n2f	5	4	1	4	3	3	4	3	2	3	4	4	3	1	3	3	2	4	3	3	4	4	5	5	4	3	4	5
n3f	1	4	2	4	2	3	4	3	2	4	5	2	3	1	2	3	1	3	2	3	2	4	5	1	3	4	2	5
e6f	4	4	3	3	4	4	5	4	1	3	4	1	2	2	4	3	1	4	4	4	2	2	3	1	1	4	4	5
b5f	5	5	4	4	4	4	5	4	4	1	5	3	2	4	4	5	5	4	4	5	2	4	4	5	5	4	5	5
b6f	5	5	4	4	2	4	5	3	4	5	4	3	2	5	4	4	2	5	4	2	3	4	4	5	5	4	4	5
n4f	5	4	3	1	2	3	3	4	2	1	1	5	4	2	3	1	2	4	2	4	4	5	1	5	3	5	4	2
n5f	5	5	3	1	4	5	2	3	5	1	4	5	4	3	3	3	4	4	3	5	5	3	2	5	3	5	3	2

Appendix III

Arithmetic Means of the Facial Action Unit Results - Computer-Controlled Agent scenario

	1a	**1b**	**2a**	**2b**	**3a**	**3b**	**4a**	**4b**	**5a**	**5b**	**6a**	**6b**	**7a**	**7b**	**8a**	**9a**	**10a**	**10b**	**11a**	**11b**	**12a**	**12b**	**13a**	**13b**	**14a**	**14b**
AU01	0.302	0.165	0.170	0.123	0.344	0.356	0.376	0.349	0.190	0.243	0.115	0.175	0.207	0.173	0.183	0.195	0.172	0.216	0.156	0.133	0.201	0.285	0.311	0.103	0.140	0.155
AU02	0.130	0.140	0.087	0.077	0.211	0.247	0.176	0.114	0.087	0.111	0.057	0.108	0.129	0.102	0.117	0.124	0.095	0.085	0.087	0.091	0.081	0.125	0.169	0.055	0.128	0.079
AU04	0.178	1.243	0.170	0.329	0.580	0.529	1.757	0.575	0.260	1.438	0.684	0.845	0.063	1.051	0.001	0.075	0.012	0.007	0.011	0.039	0.003	0.016	0.020	1.863	0.511	1.232
AU05	0.057	0.086	0.069	0.046	0.171	0.096	0.117	0.071	0.084	0.053	0.037	0.045	0.075	0.079	0.064	0.061	0.045	0.055	0.040	0.049	0.059	0.051	0.074	0.052	0.047	0.107
AU06	0.601	0.057	0.281	0.096	0.593	0.546	0.240	0.590	0.036	0.185	0.578	0.003	0.101	0.082	0.004	0.043	0.178	0.003	0.093	0.004	0.030	0.015	0.081	1.091	0.669	0.342

AU07	0.286	0.115	0.245	0.004	0.581	0.267	0.048	0.448	0.004	0.689	0.487	0.001	0.042	0.010	0.020	0.826	0.040	0.020	0.438	0.964	0.044	0.023	0.366	0.899	1.432	0.051
AU09	0.066	0.078	0.061	0.034	0.052	0.091	0.058	0.069	0.076	0.081	0.044	0.042	0.095	0.070	0.093	0.062	0.070	0.079	0.163	0.091	0.048	0.130	0.075	0.066	0.056	0.085
AU10	2.044	0.751	0.600	0.186	1.311	0.458	1.368	1.036	0.067	0.867	1.385	1.695	0.349	1.269	0.010	0.384	0.035	0.001	0.123	0.005	0.014	0.018	0.103	1.086	0.626	1.259
AU12	0.888	0.339	0.831	0.627	0.361	0.562	0.626	0.522	0.317	0.375	0.617	0.105	0.937	0.160	0.111	0.017	0.111	0.012	0.492	0.064	0.112	0.026	0.252	0.977	0.347	0.488
AU14	0.770	0.252	1.448	1.119	1.727	0.931	0.690	1.198	0.389	0.995	0.681	0.391	0.640	1.845	0.082	1.012	0.408	0.056	0.228	0.041	1.322	0.657	0.154	0.977	1.129	1.933
AU15	0.171	0.166	0.159	0.120	0.148	0.161	0.214	0.285	0.129	0.115	0.185	0.128	0.121	0.121	0.197	0.231	0.149	0.132	0.181	0.129	0.124	0.175	0.115	0.154	0.147	0.173
AU17	0.467	0.821	0.405	0.331	0.328	0.413	0.561	0.469	0.329	0.368	0.719	0.381	0.379	0.313	0.380	0.286	0.340	0.297	0.579	0.397	0.290	0.302	0.419	0.643	0.244	0.336
AU20	0.180	0.159	0.142	0.120	0.120	0.129	0.183	0.141	0.123	0.109	0.107	0.111	0.139	0.094	0.097	0.105	0.091	0.092	0.120	0.073	0.066	0.081	0.200	0.095	0.116	0.104
AS23	0.165	0.177	0.247	0.065	0.107	0.150	0.250	0.197	0.078	0.151	0.109	0.109	0.184	0.085	0.147	0.093	0.148	0.096	0.201	0.128	0.100	0.126	0.185	0.181	0.140	0.163
AU25	0.504	0.311	0.283	0.241	0.678	0.424	0.826	0.576	0.371	0.504	0.331	0.436	0.435	0.346	0.334	0.471	0.295	0.269	1.096	0.726	0.265	0.348	0.392	0.864	0.297	0.497
AU26	0.484	0.348	0.402	0.393	0.440	0.440	0.579	0.326	0.364	0.368	0.594	0.265	0.411	0.256	0.347	0.335	0.372	0.248	0.458	0.405	0.213	0.223	0.402	0.596	0.267	0.310
AU45	0.342	0.313	0.227	0.257	0.273	0.308	0.266	0.361	0.190	0.190	0.153	0.217	0.313	0.183	0.148	0.210	0.166	0.249	0.238	0.284	0.179	0.205	0.382	0.239	0.200	0.181

Arithmetic Means of the Facial Action Unit Results - Human-Controlled Avatar scenario

	1a	1b	2a	2b	3a	3b	4a	4b	5a	5b	6a	6b	7a	7b	8a	9a	10a	10b	11a	11b	12a	12b	13a	13b	14a	14b
AU01	0.279	0.185	0.171	0.140	0.215	0.473	0.323	0.352	0.221	0.173	0.144	0.187	0.190	0.273	0.304	0.124	0.144	0.250	0.169	0.126	0.200	0.376	0.217	0.157	0.141	0.142
AU02	0.113	0.147	0.092	0.095	0.106	0.248	0.183	0.126	0.079	0.081	0.073	0.130	0.095	0.136	0.220	0.087	0.088	0.121	0.075	0.074	0.097	0.156	0.091	0.081	0.163	0.080
AU04	0.118	1.028	0.142	0.221	0.664	0.790	1.536	0.560	0.368	1.338	0.729	0.815	0.028	0.707	0.015	0.021	0.002	0.011	0.009	0.009	0.010	0.008	0.005	1.881	0.379	0.993
AU05	0.056	0.098	0.064	0.075	0.054	0.086	0.107	0.087	0.074	0.059	0.049	0.050	0.074	0.061	0.097	0.048	0.039	0.050	0.043	0.050	0.050	0.050	0.045	0.057	0.055	0.067
AU06	0.964	0.326	0.800	0.205	1.375	0.711	0.669	1.092	0.120	0.622	0.941	0.005	0.296	0.950	0.071	0.100	0.086	0.012	0.204	0.043	0.146	0.018	0.105	1.604	1.014	1.127
AU07	0.562	0.341	0.531	0.006	0.956	0.398	0.157	0.693	0.024	0.696	0.880	0.002	0.033	0.105	0.050	0.789	0.027	0.036	0.626	0.888	0.032	0.018	0.246	0.989	1.908	0.243
AU09	0.084	0.093	0.066	0.047	0.103	0.156	0.056	0.093	0.080	0.124	0.043	0.051	0.070	0.072	0.093	0.075	0.053	0.093	0.155	0.079	0.066	0.104	0.067	0.059	0.058	0.091
AU10	2.024	1.121	0.930	0.244	1.800	0.920	1.427	1.199	0.327	1.110	1.434	1.690	0.551	2.124	0.080	0.536	0.020	0.007	0.373	0.009	0.085	0.018	0.230	1.662	0.862	1.825
AU12	1.484	0.981	1.481	0.967	0.993	0.899	1.310	0.988	0.747	0.914	1.231	0.144	1.356	1.024	0.435	0.066	0.097	0.045	1.056	0.243	0.207	0.034	0.324	1.457	0.695	1.326
AU14	1.337	0.807	1.771	1.345	1.900	1.136	0.636	1.314	0.727	1.394	1.564	0.420	0.784	2.515	0.103	1.294	0.428	0.133	0.460	0.197	1.410	0.751	0.280	1.430	1.346	2.302
AU15	0.218	0.240	0.177	0.127	0.146	0.175	0.270	0.397	0.115	0.152	0.130	0.143	0.129	0.196	0.300	0.155	0.147	0.127	0.213	0.155	0.193	0.190	0.127	0.137	0.185	0.240

AU17	0.821	0.848	0.485	0.295	0.435	0.588	0.606	0.509	0.310	0.439	0.907	0.346	0.365	0.408	0.364	0.272	0.335	0.322	0.521	0.341	0.220	0.271	0.320	0.766	0.314	0.834
AU20	0.151	0.126	0.128	0.099	0.114	0.155	0.198	0.197	0.095	0.104	0.102	0.108	0.126	0.087	0.071	0.092	0.089	0.104	0.125	0.072	0.067	0.090	0.167	0.095	0.137	0.142
AS23	0.255	0.231	0.230	0.080	0.199	0.202	0.158	0.217	0.130	0.207	0.112	0.109	0.176	0.138	0.092	0.085	0.146	0.147	0.170	0.060	0.119	0.115	0.144	0.132	0.142	0.220
AU25	0.716	0.598	0.395	0.337	0.671	0.514	0.924	0.818	0.451	0.765	0.322	0.541	0.442	0.622	0.570	0.372	0.425	0.313	1.384	0.673	0.323	0.307	0.401	1.039	0.499	1.255
AU26	0.722	0.367	0.511	0.389	0.390	0.375	0.761	0.345	0.356	0.404	0.316	0.272	0.465	0.313	0.342	0.281	0.441	0.261	0.508	0.351	0.245	0.219	0.286	0.526	0.321	0.606
AU45	0.385	0.371	0.258	0.228	0.260	0.364	0.253	0.358	0.209	0.201	0.125	0.256	0.279	0.263	0.133	0.217	0.146	0.209	0.214	0.229	0.130	0.165	0.281	0.252	0.219	0.168

Maximum Values of the Facial Action Unit Results - Computer-Controlled Agent scenario

	1a	1b	2a	2b	3a	3b	4a	4b	5a	5b	6a	6b	7a	7b	8a	9a	10a	10b	11a	11b	12a	12b	13a	13b	14a	14b
AU01	4.10	3.96	3.60	3.41	5.00	4.40	4.89	5.00	4.50	3.78	2.46	2.99	5.00	4.10	5.00	5.00	5.00	4.54	4.80	2.66	3.36	5.00	3.82	3.84	4.34	5.00
AU02	3.66	3.49	3.19	2.63	4.97	4.79	3.79	4.10	2.94	3.13	2.33	3.02	5.00	3.58	4.98	4.58	4.88	2.70	3.21	2.15	2.86	5.00	3.19	1.96	4.29	3.70
AU04	2.60	4.48	2.14	2.58	4.45	3.54	3.62	1.92	2.23	4.48	2.95	3.05	2.94	2.91	0.66	2.27	2.44	1.80	1.44	1.44	1.41	1.38	1.52	3.72	1.55	2.77
AU05	1.34	3.38	3.19	1.88	4.11	2.80	3.86	2.13	1.65	1.55	1.25	0.94	1.67	1.99	2.84	3.57	2.05	1.40	1.08	2.73	2.31	2.30	2.25	2.18	2.61	3.72
AU06	2.91	2.78	1.95	1.97	2.90	2.86	2.37	3.89	1.41	3.40	3.46	0.89	2.53	1.92	0.65	1.75	2.68	1.07	1.65	1.43	1.64	2.73	2.22	3.64	2.60	2.58
AU07	2.80	3.13	3.32	0.78	4.23	3.30	3.90	2.40	1.69	3.81	2.15	0.57	3.74	1.10	1.83	3.28	2.62	2.56	2.41	2.87	1.22	4.77	3.49	2.81	3.73	2.49
AU09	2.16	2.77	2.86	0.73	3.07	3.21	1.51	2.26	1.34	2.23	1.14	0.94	2.69	2.04	2.52	2.07	2.56	4.87	3.00	3.16	2.42	4.39	2.45	1.17	2.54	1.43
AU10	3.52	3.28	2.55	2.64	2.83	2.32	3.26	3.29	1.60	2.60	3.57	2.65	2.54	2.44	1.41	2.12	2.22	0.89	2.26	1.26	1.11	3.27	2.33	4.70	2.93	3.70
AU12	3.63	2.71	2.75	3.16	3.27	2.88	3.24	3.85	2.59	3.38	3.36	1.79	3.01	2.16	2.44	1.70	2.11	1.97	3.06	1.67	2.61	2.38	2.45	3.43	2.41	3.55
AU14	3.28	3.47	4.25	2.93	3.34	3.94	3.02	3.18	2.63	4.97	3.62	1.91	2.59	3.42	1.53	3.26	2.77	2.62	1.76	1.85	3.22	2.85	2.28	3.23	4.14	3.87
AU15	2.72	4.13	3.55	1.88	1.52	3.79	3.23	5.00	4.64	5.00	5.00	1.64	2.26	3.42	2.90	5.00	4.91	4.02	2.60	3.12	5.00	5.00	2.83	1.85	2.36	5.00
AU17	3.20	4.46	3.65	2.41	2.20	2.07	4.53	2.70	2.55	2.08	3.55	2.66	3.91	1.63	3.85	3.67	3.29	4.50	3.45	3.68	2.46	3.74	2.86	3.22	2.36	4.05
AU20	2.66	3.10	2.22	1.39	1.99	2.10	2.89	3.06	2.34	1.46	4.40	1.27	2.90	2.82	1.59	3.33	1.75	1.85	1.71	1.77	0.94	2.41	3.34	1.28	2.78	1.46
AS23	1.90	3.99	4.82	1.03	4.74	2.22	3.23	2.47	2.58	2.45	2.01	1.50	3.21	1.50	2.05	2.03	4.55	4.54	2.86	2.35	2.28	5.00	2.59	1.96	2.51	2.33
AU25	5.00	4.17	2.92	3.69	5.00	2.57	4.30	5.00	4.41	3.95	4.63	3.54	4.95	3.81	3.83	4.90	3.51	3.88	4.37	2.76	1.85	4.52	3.57	4.35	3.73	4.05
AU26	3.53	2.77	4.86	5.00	4.65	3.35	3.46	2.31	2.69	3.35	3.03	2.40	4.25	3.51	3.22	3.95	2.78	2.69	2.80	2.19	3.84	3.95	2.96	3.84	2.43	3.31
AU45	4.41	4.03	3.38	3.27	3.41	3.82	2.85	3.51	3.15	3.26	3.94	2.98	4.38	3.30	4.02	2.62	3.49	3.74	3.19	2.92	3.15	3.70	4.92	3.28	3.52	2.68

Maximum Values of the Facial Action Unit Results - Human-Controlled Avatar scenario

	1a	1b	2a	2b	3a	3b	4a	4b	5a	5b	6a	6b	7a	7b	8a	9a	10a	10b	11a	11b	12a	12b	13a	13b	14a	14b
AU01	4.72	2.20	4.49	3.44	3.81	5.00	5.00	5.00	4.18	3.09	4.44	3.19	4.99	4.97	5.00	3.81	5.00	5.00	5.00	2.05	3.67	4.53	5.00	3.65	3.27	3.79
AU02	3.66	2.48	3.68	3.40	2.96	4.92	4.43	4.09	3.51	2.45	2.31	2.50	4.30	4.82	4.97	2.78	4.81	4.92	3.88	1.59	2.62	3.86	3.40	3.03	3.22	3.80
AU04	2.32	3.16	2.03	3.01	3.97	3.46	3.17	3.03	2.06	4.21	2.28	3.21	1.69	4.47	2.42	1.10	2.41	4.60	1.37	1.03	1.97	3.45	0.79	3.37	1.58	2.77
AU05	1.30	3.71	2.69	3.65	2.41	2.55	3.67	2.39	2.19	2.83	2.09	1.99	2.22	2.27	2.86	2.09	2.57	1.22	0.98	1.79	1.42	2.75	1.35	1.29	1.20	2.45
AU06	3.26	3.79	3.62	2.72	3.74	3.31	2.85	4.15	1.81	3.18	3.83	1.10	2.02	3.18	2.27	2.72	2.02	2.27	3.68	1.47	2.06	2.47	2.59	3.97	3.59	3.79
AU07	3.34	4.05	2.95	1.74	4.65	4.23	2.29	2.81	2.55	4.00	2.59	1.70	2.01	2.75	2.16	3.30	2.66	3.02	3.54	3.08	1.53	3.13	3.63	3.13	4.90	2.29
AU09	0.93	2.99	2.24	1.98	2.39	2.10	1.74	2.47	3.60	3.81	1.57	1.16	1.86	3.34	2.13	2.94	2.08	4.06	3.21	1.43	1.62	5.00	3.13	1.19	1.33	3.62
AU10	3.24	4.55	3.09	2.55	4.02	3.40	3.16	3.32	3.04	3.44	3.56	2.75	2.21	3.17	2.36	2.54	2.26	2.05	3.29	1.25	2.86	1.94	2.78	4.28	4.48	4.71
AU12	4.05	3.80	5.00	4.04	3.92	3.70	4.01	4.01	3.64	3.84	3.71	1.66	4.02	3.32	3.82	3.46	2.53	2.74	3.30	3.02	2.67	3.22	3.75	3.65	3.79	4.12
AU14	3.57	3.48	4.56	3.53	4.92	4.20	2.90	3.83	3.85	4.16	3.38	1.92	2.44	4.49	1.90	3.29	1.87	2.91	2.92	2.19	2.99	3.96	3.37	4.61	4.49	4.55
AU15	2.25	4.24	3.78	3.38	2.30	5.00	3.65	5.00	3.30	5.00	2.97	2.32	2.46	5.00	3.54	5.00	2.69	3.87	3.88	1.45	2.52	4.59	2.47	4.22	4.67	5.00
AU17	2.78	4.39	2.86	3.20	2.60	3.07	3.84	3.16	2.93	3.76	3.84	3.56	3.11	3.52	3.24	3.92	2.82	4.44	4.89	1.70	2.79	2.98	3.16	2.67	3.61	3.78
AU20	1.91	2.79	2.09	2.49	1.75	3.76	2.23	4.57	2.39	2.40	1.90	1.65	2.33	3.89	1.70	1.76	2.47	3.14	1.84	2.19	1.04	2.97	2.87	1.90	2.63	1.96
AS23	1.99	3.64	5.00	3.32	3.39	4.15	2.36	1.84	5.00	2.32	4.50	1.99	3.09	4.43	2.06	4.35	3.28	4.47	2.34	1.33	1.99	2.33	2.18	2.23	1.87	4.53
AU25	4.93	4.05	5.00	4.49	4.49	4.01	3.19	5.00	4.93	4.94	4.45	3.90	3.24	5.00	4.31	4.89	3.15	3.60	4.87	4.81	4.05	5.00	4.40	4.53	4.61	5.00
AU26	4.30	3.17	3.26	5.00	2.93	3.52	3.74	3.41	3.22	3.53	3.15	2.99	4.40	5.00	3.67	2.60	2.96	2.90	4.45	2.61	2.93	3.03	4.24	3.57	3.92	4.22
AU45	4.50	3.66	3.68	3.18	3.72	4.17	3.06	3.61	2.95	3.31	3.10	4.08	4.11	3.93	4.03	2.60	3.05	4.00	3.08	3.17	2.58	3.72	4.67	3.87	3.31	3.69

Standard Deviation Values Values of the Facial Action Unit Results - Computer-Controlled Agent scenario

	1a	1b	2a	2b	3a	3b	4a	4b	5a	5b	6a	6b	7a	7b	8a	9a	10a	10b	11a	11b	12a	12b	13a	13b	14a	14b
AU01	0.562	0.296	0.350	0.270	0.779	0.729	0.717	0.626	0.441	0.515	0.218	0.410	0.494	0.438	0.459	0.460	0.480	0.389	0.331	0.236	0.377	0.493	0.536	0.214	0.293	0.305
AU02	0.379	0.313	0.252	0.235	0.654	0.645	0.465	0.352	0.267	0.327	0.147	0.248	0.423	0.325	0.405	0.379	0.313	0.219	0.233	0.208	0.203	0.318	0.428	0.123	0.350	0.236

AU04	0.213	0.348	0.232	0.200	0.612	0.433	0.427	0.317	0.247	0.487	0.268	0.336	0.272	0.287	0.026	0.152	0.129	0.064	0.076	0.139	0.057	0.095	0.098	0.317	0.176	0.357
AU05	0.142	0.266	0.221	0.126	0.562	0.263	0.394	0.177	0.226	0.138	0.100	0.109	0.197	0.229	0.194	0.208	0.144	0.135	0.100	0.142	0.178	0.131	0.229	0.143	0.140	0.326
AU06	0.570	0.253	0.251	0.263	0.531	0.456	0.434	0.615	0.146	0.438	0.434	0.039	0.281	0.240	0.041	0.150	0.278	0.039	0.280	0.041	0.120	0.120	0.248	0.896	0.365	0.449
AU07	0.403	0.299	0.366	0.033	0.645	0.352	0.216	0.486	0.061	0.499	0.380	0.014	0.226	0.062	0.106	0.498	0.193	0.119	0.477	0.407	0.117	0.182	0.447	0.547	0.541	0.140
AU09	0.140	0.229	0.151	0.063	0.127	0.272	0.127	0.175	0.160	0.197	0.096	0.082	0.229	0.153	0.249	0.157	0.190	0.192	0.424	0.263	0.106	0.399	0.202	0.143	0.158	0.200
AU10	0.407	0.399	0.335	0.321	0.408	0.389	0.505	0.569	0.217	0.359	0.417	0.253	0.256	0.411	0.089	0.260	0.174	0.025	0.342	0.054	0.082	0.141	0.239	0.730	0.380	0.503
AU12	0.633	0.405	0.305	0.483	0.566	0.482	0.706	0.653	0.361	0.531	0.614	0.166	0.394	0.285	0.261	0.111	0.287	0.104	0.773	0.149	0.204	0.168	0.424	0.920	0.265	0.527
AU14	0.591	0.423	0.494	0.540	0.482	0.633	0.558	0.711	0.370	0.535	0.666	0.263	0.371	0.423	0.203	0.477	0.244	0.223	0.363	0.125	0.287	0.341	0.294	0.893	0.418	0.500
AU15	0.272	0.308	0.320	0.192	0.239	0.281	0.362	0.650	0.267	0.194	0.287	0.208	0.188	0.196	0.413	0.440	0.296	0.257	0.312	0.214	0.249	0.352	0.224	0.221	0.210	0.410
AU17	0.411	0.736	0.458	0.286	0.292	0.344	0.690	0.472	0.287	0.323	0.569	0.360	0.391	0.246	0.527	0.286	0.408	0.324	0.558	0.369	0.244	0.269	0.481	0.516	0.216	0.305
AU20	0.329	0.325	0.275	0.216	0.233	0.242	0.393	0.302	0.240	0.196	0.220	0.203	0.291	0.184	0.198	0.232	0.178	0.172	0.250	0.141	0.106	0.161	0.400	0.163	0.263	0.180
AS23	0.338	0.365	0.493	0.129	0.239	0.301	0.496	0.380	0.164	0.311	0.222	0.205	0.389	0.160	0.335	0.195	0.345	0.233	0.420	0.266	0.228	0.358	0.417	0.337	0.309	0.323
AU25	0.763	0.302	0.327	0.372	0.934	0.460	0.878	0.858	0.631	0.641	0.484	0.511	0.469	0.583	0.469	0.612	0.391	0.344	1.175	0.640	0.287	0.475	0.555	0.986	0.435	0.687
AU26	0.575	0.332	0.512	0.523	0.527	0.518	0.612	0.362	0.413	0.421	0.628	0.298	0.432	0.302	0.500	0.470	0.396	0.301	0.543	0.394	0.243	0.274	0.460	0.616	0.313	0.413
AU45	0.744	0.607	0.445	0.557	0.520	0.607	0.493	0.593	0.377	0.359	0.315	0.348	0.579	0.391	0.340	0.407	0.353	0.500	0.518	0.529	0.324	0.445	0.871	0.518	0.394	0.336

Standard Deviation Values Values of the Facial Action Unit Results - Human-Controlled Avatar scenario

	1a	1b	2a	2b	3a	3b	4a	4b	5a	5b	6a	6b	7a	7b	8a	9a	10a	10b	11a	11b	12a	12b	13a	13b	14a	14b
AU01	0.497	0.308	0.373	0.329	0.396	0.797	0.680	0.603	0.385	0.312	0.245	0.373	0.360	0.552	0.677	0.216	0.431	0.526	0.339	0.220	0.347	0.620	0.427	0.352	0.262	0.262
AU02	0.297	0.321	0.258	0.305	0.290	0.630	0.493	0.365	0.258	0.196	0.166	0.293	0.258	0.377	0.635	0.214	0.324	0.387	0.198	0.183	0.232	0.384	0.262	0.227	0.390	0.203
AU04	0.216	0.338	0.201	0.269	0.364	0.664	0.397	0.348	0.324	0.438	0.209	0.478	0.119	0.459	0.146	0.069	0.049	0.116	0.079	0.054	0.107	0.082	0.035	0.400	0.197	0.371
AU05	0.136	0.297	0.182	0.228	0.147	0.241	0.340	0.239	0.225	0.169	0.149	0.129	0.189	0.160	0.304	0.158	0.126	0.122	0.105	0.138	0.120	0.133	0.116	0.144	0.142	0.183
AU06	0.747	0.549	0.582	0.440	0.874	0.662	0.639	0.724	0.276	0.695	0.408	0.047	0.421	0.554	0.293	0.308	0.234	0.106	0.444	0.168	0.253	0.129	0.356	1.048	0.642	0.742
AU07	0.595	0.526	0.494	0.048	1.016	0.559	0.350	0.632	0.146	0.517	0.328	0.041	0.128	0.230	0.181	0.606	0.124	0.169	0.683	0.507	0.110	0.142	0.370	0.606	0.841	0.336
AU09	0.163	0.224	0.146	0.147	0.249	0.341	0.114	0.235	0.216	0.322	0.093	0.104	0.156	0.164	0.207	0.181	0.137	0.236	0.415	0.175	0.144	0.321	0.203	0.126	0.148	0.234
AU10	0.476	0.569	0.456	0.378	0.430	0.483	0.590	0.599	0.330	0.448	0.358	0.278	0.353	0.382	0.336	0.374	0.116	0.081	0.612	0.068	0.190	0.131	0.391	0.837	0.473	0.991

AU12	0.818	0.622	0.728	0.624	0.980	0.737	1.030	0.741	0.553	0.817	0.587	0.173	0.727	0.678	0.686	0.336	0.312	0.220	0.884	0.483	0.306	0.196	0.545	1.022	0.596	1.066
AU14	0.690	0.624	0.689	0.611	0.601	0.677	0.524	0.905	0.363	0.661	0.523	0.275	0.407	0.560	0.180	0.510	0.281	0.368	0.422	0.310	0.279	0.386	0.470	0.883	0.532	0.660
AU15	0.332	0.432	0.326	0.218	0.216	0.290	0.485	0.815	0.214	0.300	0.205	0.208	0.211	0.370	0.475	0.302	0.246	0.232	0.370	0.234	0.286	0.373	0.218	0.218	0.291	0.489
AU17	0.661	0.870	0.472	0.243	0.342	0.483	0.605	0.491	0.273	0.415	0.666	0.301	0.341	0.316	0.421	0.233	0.346	0.370	0.520	0.268	0.212	0.259	0.350	0.600	0.270	0.644
AU20	0.274	0.258	0.251	0.182	0.203	0.286	0.363	0.420	0.192	0.190	0.186	0.194	0.252	0.167	0.131	0.215	0.184	0.219	0.263	0.133	0.111	0.175	0.328	0.174	0.273	0.262
AS23	0.445	0.446	0.456	0.186	0.429	0.413	0.289	0.382	0.294	0.390	0.251	0.208	0.328	0.288	0.187	0.219	0.294	0.349	0.350	0.117	0.235	0.239	0.314	0.270	0.285	0.417
AU25	0.963	0.747	0.641	0.561	1.002	0.651	0.860	1.240	0.709	1.003	0.466	0.627	0.520	0.837	0.930	0.529	0.443	0.447	1.346	0.836	0.425	0.410	0.563	1.025	0.826	1.208
AU26	0.735	0.380	0.590	0.535	0.414	0.449	0.685	0.386	0.412	0.441	0.342	0.329	0.475	0.379	0.415	0.369	0.433	0.316	0.584	0.397	0.258	0.236	0.373	0.553	0.331	0.669
AU45	0.754	0.649	0.457	0.483	0.477	0.625	0.484	0.547	0.402	0.374	0.257	0.442	0.535	0.500	0.313	0.401	0.290	0.442	0.450	0.459	0.245	0.336	0.684	0.527	0.408	0.321

Appendix IV

Emotional Facial Behaviour Observations - Computer-Controlled Agent scenario

	1a	1b	2a	2b	3a	3b	4a	4b	5a	5b	6a	6b	7a	7b	8a	9a	10a	10b	11a	11b	12a	12b	13a	13b	14a
Lips Tightner	2	0	6	0	3	2	6	3	3	3	2	2	4	0	4	1	0	0	2	0	1	1	2	0	3
Lips Part	10	1	1	1	4	6	13	8	7	13	3	1	7	4	0	1	3	6	9	5	0	4	1	3	5
Nose Wrinkler	0	1	0	0	0	4	2	0	0	1	1	1	0	0	1	0	3	0	0	0	0	0	0	3	1
Dimpler	2	6	7	2	5	1	3	0	0	0	1	0	1	0	2	2	0	1	0	0	0	0	1	4	1
Lips Puckerer	2	1	2	1	4	3	3	0	0	0	0	1	1	0	0	1	2	0	0	1	0	0	0	0	0
Lips Corner Raiser	0	0	0	3	1	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	1
Lips Corner Raiser	1	0	1	1	1	1	3	5	3	3	3	1	1	1	0	1	1	1	0	0	0	3	0	4	1
Lips Corner Raiser	2	0	1	6	4	3	0	1	0	1	3	0	1	0	3	2	4	0	3	0	1	0	4	7	1
Inner Brow Raiser	1	0	2	4	9	5	6	3	0	1	1	1	1	2	2	1	3	3	0	1	3	2	1	1	1
Brow Lowerer	0	0	1	1	1	2	4	1	0	4	1	2	5	2	0	0	0	0	0	0	0	1	2	2	1

Smiling Rest Face	No	No	No	No	Yes	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	Yes	No

Emotional Facial Behaviour Observations - Human-Controlled Avatar scenario

	1a	**1b**	**2a**	**2b**	**3a**	**3b**	**4a**	**4b**	**5a**	**5b**	**6a**	**6b**	**7a**	**7b**	**8a**	**9a**	**10a**	**10b**	**11a**	**11b**	**12a**	**12b**	**13a**	**13b**	**14a**
Lips Tightner	11	0	15	0	8	2	3	2	2	4	3	3	2	0	1	3	2	1	1	0	1	0	2	0	4
Lips Part	6	6	3	4	7	5	19	10	5	10	7	9	9	4	1	1	7	9	11	9	0	7	2	6	2
Nose Wrinkler	0	0	1	0	3	2	0	1	0	1	2	1	0	0	0	0	1	0	0	0	0	0	1	5	1
Dimpler	2	7	4	0	3	0	5	2	0	0	3	3	1	3	3	0	1	0	0	0	0	0	2	2	1
Lips Puckerer	0	3	2	0	2	3	0	2	0	0	0	0	0	1	1	0	0	0	1	1	0	0	0	0	0
Lips Corner Raiser	4	2	7	5	3	4	3	6	6	5	2	0	2	5	4	2	2	1	1	1	2	0	1	10	3
Lips Corner Raiser	3	5	5	4	4	2	4	7	9	2	3	2	5	4	1	2	2	0	3	3	3	1	2	16	2
Lips Corner Raiser	4	1	4	5	6	6	8	2	0	7	4	0	2	9	5	4	4	2	2	2	3	2	5	9	4
Inner Brow Raiser	1	1	4	6	3	3	4	3	2	1	3	2	0	8	8	0	4	0	0	0	1	6	2	1	1
Brow Lowerer	2	0	0	4	0	0	1	0	2	3	1	11	3	6	3	0	0	2	0	0	0	2	1	6	0
Smiling Rest Face	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	Yes	Yes	No	No	Yes	Yes	No	Yes	No	Yes	Yes	No

Appendix V

Print of Google Colad Notebook Page

A Study on the Relationship between Social Presence, Emotional Facial Behaviour and Flow in Multiplayer Game Experiences

This Google Colab Notebook serves as an auxiliary tool for Michele Fanelli's homonym master's thesis, a study on how introspective experiences like social presence and flow, and observable physiological phenomena like emotional facial behaviour, are affected by multiplayer game experiences.

The study consists of an experiment in which participants are isolated in different rooms with no possible communication during the session, and are asked to play a cooperative game experience, first against a bot and then against a known friend playing in the same conditions in another room.

The introspective experiences are measured through self-report tools, mainly through the Hexad-12 questionnaire administered at the start of the experiment, and a tailored questionnaire composed of key scales of the Game Experience Questionnaire, administered after each game session.

The emotional facial behaviour is analysed through the Facial Action Units extraction software Open Face which provides quantitative measurements to the movements of the specific facial muscle groups.

This Google Colab Notebook has only demostrative purposed, as it showcases the collected data of the mentioned measurements and provides simple Python code cells to calculate its descriptive statistics and plot them accordingly.

The datasets in CSV format as well as the whole master's thesis can be found on this GitHub Repo.

The following code cell loads the datasets from the GitHub Repo and prepares them as Pandas DataFrames.

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

# The CSV files containing the datasets are loaded from the GitHub repo of the project
self_report_data_URL = "https://raw.githubusercontent.com/Phanelly/Relationship_SocialPresence_EmotionalFacialBehaviour_Flow_Study_Data/
fau_avg_data_URL = "https://raw.githubusercontent.com/Phanelly/Relationship_SocialPresence_EmotionalFacialBehaviour_Flow_Study_Data/mai
fau_max_data_URL = "https://raw.githubusercontent.com/Phanelly/Relationship_SocialPresence_EmotionalFacialBehaviour_Flow_Study_Data/mai
fau_std_data_URL = "https://raw.githubusercontent.com/Phanelly/Relationship_SocialPresence_EmotionalFacialBehaviour_Flow_Study_Data/mai

df_self_report = pd.read_csv(self_report_data_URL)
df_fau_avg = pd.read_csv(fau_avg_data_URL)
df_fau_max = pd.read_csv(fau_max_data_URL)
df_fau_std = pd.read_csv(fau_std_data_URL)

pd.set_option("display.precision", 1)
```

Self-Report Data

The self-report data consists of the results of the Hexad-12 questionnaire, the results of the tailored Game Experience Questionnaire administrated after the bot (computer-controlled agent) game session and the results of the same tailored Game Experience Questionnaire administrated after the friend (human-controlled avatar) game session.

The Hexad-12 questionnaire consists of six 7-point Likert scales: Philanthropist, Socializer, Achiever, Player, Free Spirit and Disruptor. Each scale is composed of two items, each labelled according to the scale: 'p' for Philanthropist, 's' for Socializer, 'a' for Achiever, 'r' for Player, 'm' for Free Spirit and 'd' for Disruptor.

The tailored GEQ consists of five 5-point Likert scales: Immersion, Flow, Psychological Involvement (Empathy), Psychological Involvement (Negative Feelings) and Behavioural Involvement. Each scale from four to six items, each labelled according to the scale: 'i' for Immersion, 'f' for Flow, 'e' for Empathy, 'n' for Negative Feelings and 'b' for Behavioural Involvement.

Additionally, the items of the tailored GEQ administered after the bot game session are characterized by the suffix 'b' for bot, while the items of the tailored GEQ administered after the friend game session are characterized by the suffix 'f' for friend.

More information on the self-report tools and their scales can be found in the master's thesis document.

The following code cell creates a specific DataFrame containing the items of the Hexad-12 questionnaire. Then it creates a DataFrame containing the values of the scales, calculated as the arithmetic mean of the items composing it.

The code cell afterwards plots the data.

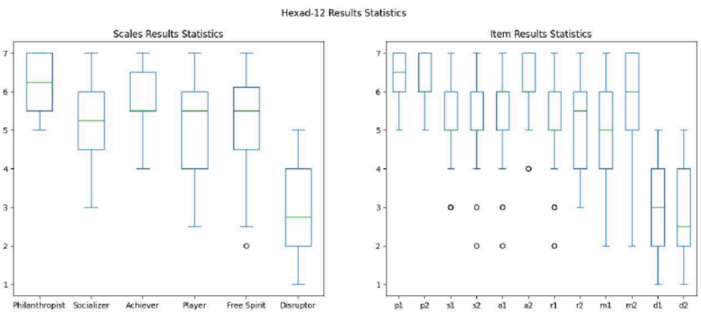
The label codes correspond to the following questions:

```
p1: It makes me happy if I am able to help others
p2: The well-being of others is important to me
s1: I like being part of a team
s2: I enjoy group activities
a1: I like mastering difficult tasks
a2: I enjoy emerging victorious out of difficult circumstances
r1: Rewards are a great way to motivate me
r2: If the reward is sufficient, I will put in the effort
m1: It is important to me to follow my own path
m2: Being independent is important to me
d1: I see myself as a rebel
d2: I dislike following rules
```

```
# Select from the Self Report Data Results the columns corresponding to items of the Hexad-12 Questionnaire
hexad_columns = ['p1', 'p2','s1', 's2', 'a1', 'a2', 'r1', 'r2', 'm1', 'm2', 'd1', 'd2']

# Create the DataFrame containing the Hexad-12 Questionnaire items
df_hexad = pd.DataFrame(data=df_self_report, columns=hexad_columns)

# Create the DataFrame containing the Hexad-12 Questionnaire sclaes
# calculated as the arithmetic mean on their items
df_hexad_scales = pd.DataFrame({
    'Philanthropist': df_hexad[['p1', 'p2']].mean(axis=1),
    'Socializer': df_hexad[['s1', 's2']].mean(axis=1),
    'Achiever': df_hexad[['a1', 'a2']].mean(axis=1),
    'Player': df_hexad[['r1', 'r2']].mean(axis=1),
    'Free Spirit': df_hexad[['m1', 'm2']].mean(axis=1),
    'Disruptor': df_hexad[['d1', 'd2']].mean(axis=1)
})

# Print the descriptive statistics of both DataFrames
pd.set_option("display.precision", 1)
pd.concat([df_hexad_scales, df_hexad]).describe().loc[['mean', 'std', 'max', 'min']]
```

	Philanthropist	Socializer	Achiever	Player	Free Spirit	Disruptor	p1	p2	s1
mean	6.2	5.2	5.8	5.2	5.4	3.0	6.3	6.2	5.3
std	0.7	1.0	0.8	1.2	1.3	1.1	0.8	0.7	1.0
max	7.0	7.0	7.0	7.0	7.0	5.0	7.0	7.0	7.0

```
# Plot the data of the two Hexad-12 DataFrames
fig, axes = plt.subplots(nrows=1, ncols=2, figsize=(16, 6))

df_hexad.plot(kind='box', title='Item Results Statistics', ax=axes[1])
df_hexad_scales.plot(kind='box', title='Scales Results Statistics', ax=axes[0])

plt.suptitle('Hexad-12 Results Statistics')
plt.show()
```



The following code cell creates two DataFrames containing the items of the tailored Game Experience Questionnaire, one for the bot game setting and one for the friend setting. Then it creates two DataFrames containing the values of the scales, calculated as the arithmetic mean o the items composing it.

The code cell afterwards plots the data.

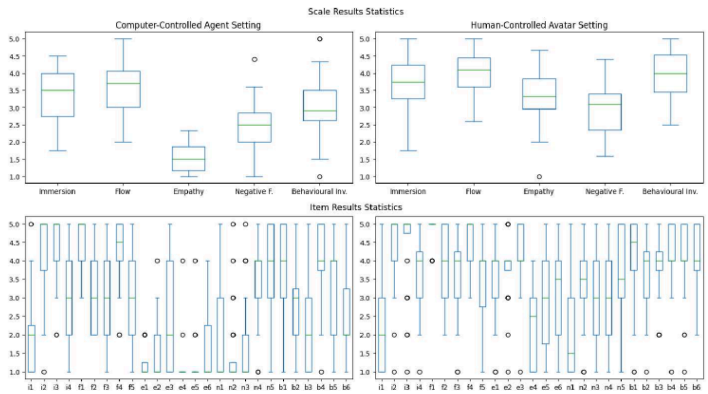
The label codes correspond to the following questions:

```
i1: I was interested in the game’s story
i2: I thought about other things
i3: I found it tiresome
i4: It felt like a rich experience
f1: I was fully occupied with the game
f2: I forgot everything around me
f3: I lost track of time
f4: I was deeply concentrated in the game
f5: I lost connection with the outside world
e1: I empathized with the other
e2: I felt connected to the other
e3: I found it enjoyable to be with the other
e4: When I was happy, the other was happy
e5: When the other was happy, I was happy
e6: I admired the other
n1: I felt jealous about the other
n2: I influenced the mood of the other
n3: I was influenced by the other moods
n4: I felt revengeful
n5: I felt schadenfreude (malicious delight)
b1: My actions depended on the other’s actions
b2: The other's actions were dependent on my actions
b3: The other paid close attention to me
b4: I paid close attention to the other
b5: What the other did affected what I did
b6: What I did affected what the other did
```

```python
# Select from the Self Report Data Results the columns corresponding to items of the GEQ
geq_bot_columns = ['i1b', 'i2b', 'i3b', 'i4b', 'f1b', 'f2b', 'f3b', 'f4b', 'f5b',
                   'e1b', 'e2b', 'e3b', 'e4b', 'e5b', 'e6b', 'n1b', 'n2b', 'n3b',
                   'n4b', 'n5b', 'b1b', 'b2b', 'b3b', 'b4b', 'b5b', 'b6b']

geq_friend_columns = ['i1f', 'i2f', 'i3f', 'i4f', 'f1f', 'f2f', 'f3f', 'f4f', 'f5f',
                      'e1f', 'e2f', 'e3f', 'e4f', 'e5f', 'e6f', 'n1f', 'n2f', 'n3f',
                      'n4f', 'n5f', 'b1f', 'b2f', 'b3f', 'b4f', 'b5f', 'b6f']

# Create the dataframes containing the GEQ items
df_geq_bot = pd.DataFrame(data=df_self_report, columns=geq_bot_columns)
df_geq_friend = pd.DataFrame(data=df_self_report, columns=geq_friend_columns)

# Rename the columns to improve readability
new_bot_colums = {col: col[:-1] for col in geq_bot_columns}
new_friend_colums = {col: col[:-1] for col in geq_friend_columns}

df_geq_bot.rename(columns=new_bot_colums, inplace=True)
df_geq_friend.rename(columns=new_friend_colums, inplace=True)

# The items i2 and i3 need to be inverted due to the item negative formulation
df_geq_bot['i2'] = 6 - df_geq_bot['i2']
df_geq_bot['i3'] = 6 - df_geq_bot['i3']
df_geq_friend['i2'] = 6 - df_geq_friend['i2']
df_geq_friend['i3'] = 6 - df_geq_friend['i3']

# Create the dataframes containing the GEQ sclaes calculated as the arithmetic mean on their items
df_geq_bot_scales = pd.DataFrame({
    'Immersion': df_geq_bot[['i1', 'i2', 'i3', 'i4']].mean(axis=1),
    'Flow': df_geq_bot[['f1', 'f2', 'f3', 'f4', 'f5']].mean(axis=1),
    'Empathy': df_geq_bot[['e1', 'e2', 'e3', 'e4', 'e5', 'e6']].mean(axis=1),
    'Negative F.': df_geq_bot[['n1', 'n2', 'n3', 'n4', 'n5']].mean(axis=1),
    'Behavioural Inv.': df_geq_bot[['b1', 'b2', 'b3', 'b4', 'b5', 'b6']].mean(axis=1)
})

df_geq_friend_scales = pd.DataFrame({
    'Immersion': df_geq_friend[['i1', 'i2', 'i3', 'i4']].mean(axis=1),
    'Flow': df_geq_friend[['f1', 'f2', 'f3', 'f4', 'f5']].mean(axis=1),
    'Empathy': df_geq_friend[['e1', 'e2', 'e3', 'e4', 'e5', 'e6']].mean(axis=1),
    'Negative F.': df_geq_friend[['n1', 'n2', 'n3', 'n4', 'n5']].mean(axis=1),
    'Behavioural Inv.': df_geq_friend[['b1', 'b2', 'b3', 'b4', 'b5', 'b6']].mean(axis=1)
})

# Print the descriptive statistics of the scales
pd.concat([df_geq_bot_scales, df_geq_friend_scales], axis=1).describe().loc[['mean', 'std', 'max', 'min']]
```

	Immersion	Flow	Empathy	Negative F.	Behavioural Inv.	Immersion	Flow	Empathy	Neg
mean	3.4	3.5	1.6	2.5	3.1	3.6	4.0	3.3	
std	0.7	0.8	0.4	0.7	0.9	0.8	0.6	0.8	
max	4.5	5.0	2.3	4.4	5.0	5.0	5.0	4.7	

```python
# Plot of the four Game Experience Questionnaire dataframes
fig = plt.figure(constrained_layout=True, figsize=(14, 8))
subfigs = fig.subfigures(nrows=2, ncols=1)
fig.suptitle('Game Experience Questionnaire Results Statistics')

top_axes = subfigs[0].subplots(nrows=1, ncols=2)
subfigs[0].suptitle("Scale Results Statistics")

df_geq_bot_scales.plot(kind='box', title='Computer-Controlled Agent Setting', ax=top_axes[0])
df_geq_friend_scales.plot(kind='box', title='Human-Controlled Avatar Setting', ax=top_axes[1])

bottom_axes = subfigs[1].subplots(nrows=1, ncols=2)
subfigs[1].suptitle("Item Results Statistics")

df_geq_bot.plot(kind='box', ax=bottom_axes[0])
df_geq_friend.plot(kind='box', ax=bottom_axes[1])

plt.show()
```



Facial Action Units Data

The Facial Action Units data consists of the physiological measurements of the subtle movements of the facial muscle groups codified in the Facial Action Coding System. This FAU data has been acquired by applying the OpenFace software on the recordings of the participants' faces. Each participant has been recorded twice, once for each game session.

For each frame of the recording, the OpenFace software identifies the presence of an action unit, if it is visible in the face, and its intensity, from minimal to maximal on a 5-point scale, with 0 denoting the absence of the relative facial muscle activation, 1 denoting the presence at the minimum intensity and 5 the presence at the maximum intensity.

Due to the mastodontic nature of the data extrapolated with the recordings, it was deemed favourable to not work with the raw data but instead with some key descriptive statistics, namely the arithmetic mean, the maximum value and the standard deviation. The three provided datasets thus contain one of the three different descriptive measurement of choice calculated for each action unit, for each participant, on the whole recording length.

In the dataset the action units are given the suffix 'b' for the extrapolations of the recordings of the bot setting and 'f' for the extrapolations of the recordings of the friend setting.

The FAU datasets present two less entries compared to the self-reports dataset. This is due to the recording of two participant being lost due to human error.

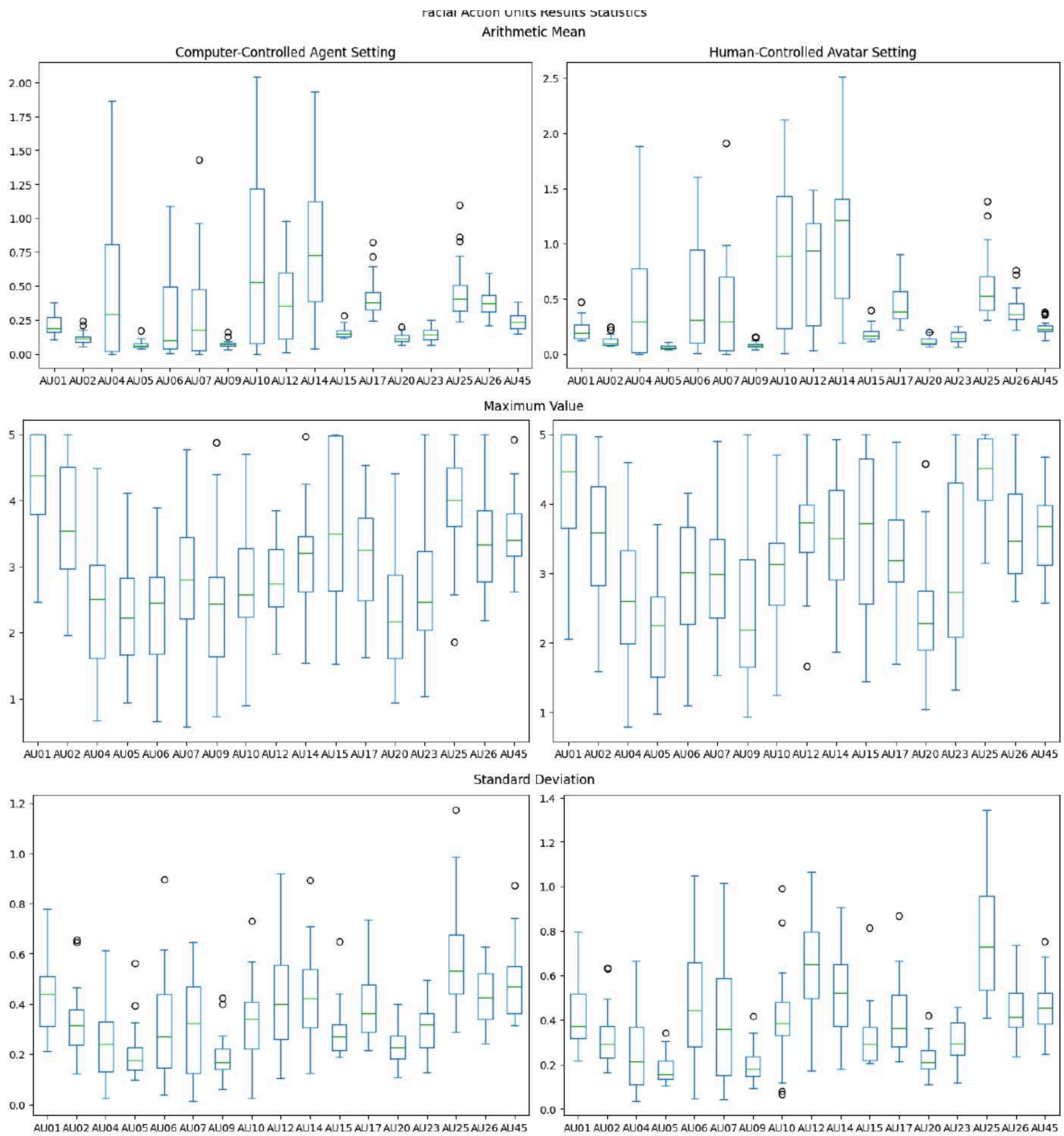
More information on the Facial Action Unit measurements and the choice behind the descriptive statistics can be found in the master's thesis document.

The following code cell creates two DataFrames for each FAC dataset, one containing the measurements of the bot setting and the other the measurements of the friend setting. Then the data is plotted.

The mousce group codified by each action unit is:

```
AU01: Inner brow raiser
AU02: Outer brow raiser
```

```
 AU04: Brow lowerer
 AU05: Upper lid raiser
 AU06: Cheek Raiser
 AU07: Lid tightener
 AU09: Nose wrinkler
 AU10: Upper lip raiser
 AU12: Lip corner puller
 AU14: Dimpler
 AU15: Lip corner depressor
 AU17: Chin raiser
 AU20_ Lip stretcher
 AU23: Lip tightener
 AU25: Lips part
 AU26: Jaw drop
 AU45: Blink


# Let's make a bot and a friend dataframe for each of the tree FAC dataframes

# Select the correct columns
bot_columns = ['AU01b', 'AU02b', 'AU04b', 'AU05b', 'AU06b', 'AU07b', 'AU09b',
               'AU10b', 'AU12b', 'AU14b', 'AU15b', 'AU17b', 'AU20b', 'AU23b',
               'AU25b', 'AU26b', 'AU45b']

friend_columns = ['AU01f', 'AU02f', 'AU04f', 'AU05f', 'AU06f', 'AU07f', 'AU09f',
                  'AU10f', 'AU12f', 'AU14f', 'AU15f', 'AU17f', 'AU20f', 'AU23f',
                  'AU25f', 'AU26f', 'AU45f']

df_fau_avg_bot = pd.DataFrame(data=df_fau_avg, columns=bot_columns)
df_fau_avg_friend = pd.DataFrame(data=df_fau_avg, columns=friend_columns)

df_fau_max_bot = pd.DataFrame(data=df_fau_max, columns=bot_columns)
df_fau_max_friend = pd.DataFrame(data=df_fau_max, columns=friend_columns)

df_fau_std_bot = pd.DataFrame(data=df_fau_std, columns=bot_columns)
df_fau_std_friend = pd.DataFrame(data=df_fau_std, columns=friend_columns)

# Rename the columns to improve readability
new_bot_colums = {col: col[:-1] for col in bot_columns}
new_friend_colums = {col: col[:-1] for col in friend_columns}

df_fau_avg_bot.rename(columns=new_bot_colums, inplace=True)
df_fau_avg_friend.rename(columns=new_friend_colums, inplace=True)

df_fau_max_bot.rename(columns=new_bot_colums, inplace=True)
df_fau_max_friend.rename(columns=new_friend_colums, inplace=True)

df_fau_std_bot.rename(columns=new_bot_colums, inplace=True)
df_fau_std_friend.rename(columns=new_friend_colums, inplace=True)

# Plot the data
#fig, axes = plt.subplots(nrows=2, ncols=2, figsize=(16, 7))
fig = plt.figure(constrained_layout=True, figsize=(14, 15))
subfigs = fig.subfigures(nrows=3, ncols=1)
fig.suptitle('Facial Action Units Results Statistics')

top_axes = subfigs[0].subplots(nrows=1, ncols=2)
subfigs[0].suptitle("Arithmetic Mean")

df_fau_avg_bot.plot(kind='box', title='Computer-Controlled Agent Setting', ax=top_axes[0])
df_fau_avg_friend.plot(kind='box', title='Human-Controlled Avatar Setting', ax=top_axes[1])

bottom_axes = subfigs[1].subplots(nrows=1, ncols=2)
subfigs[1].suptitle("Maximum Value")

df_fau_max_bot.plot(kind='box', ax=bottom_axes[0])
df_fau_max_friend.plot(kind='box', ax=bottom_axes[1])

bottom_axes = subfigs[2].subplots(nrows=1, ncols=2)
subfigs[2].suptitle("Standard Deviation")

df_fau_std_bot.plot(kind='box', ax=bottom_axes[0])
df_fau_std_friend.plot(kind='box', ax=bottom_axes[1])

plt.show()
```



Facial Action Units Results Statistics
Arithmetic Mean
Computer-Controlled Agent Setting
Human-Controlled Avatar Setting
Maximum Value
Standard Deviation
AU01 AU02 AU04 AU05 AU06 AU07 AU09 AU10 AU12 AU14 AU15 AU17 AU20 AU23 AU25 AU26 AU45

Declaration of Use of AI

During the writing process of this document, both Grammarly and ChatGPT were used extensively. Grammarly was used in its freemium version with the Google Chrome Extension for Google Docs. This version offers a spelling check tool and synonyms via double-click function. The spelling check was kept on for most of the writing process, while the synonyms function stopped working when the document size approached 50 pages in number. ChatGPT has been used extensively for brainstorming and acquiring preliminary knowledge on unfamiliar topics. For instance, the software has been repeatedly questioned to find possible statistical analysis suitable for the data type. Due to the nature of the software, all answers were taken with the needed scepticism and approached as indicators rather than trustworthy material. In many other instances, ChatGPT has been used as a linguistic auxiliary tool. It would, for example, be asked to provide adjectives or nouns better describing concepts to formulate, or it would be asked to reformulate given sentences and change the construction from passive to active. The most frequent prompt type would be "How do I write [colloquial sentence construction] in academic English". The use made of ChatGPT is far from being irreplaceable by the principal investigator's linguistic efforts, yet it does speed up the overall writing process.

Overall, AI tools have been used with conscience and consideration, in ways the principal investigator believes to be ethical use of current-time AI tools. The AIs have been used as facilitators and enhancers of the principal investigator's linguistic and cognitive skills, never, in any way, substitutive for human work.