UNIVERSIDADE DE LISBOA
INSTITUTO SUPERIOR TÉCNICO

Improving Believable Interactions in Real-time Multi-party Interactive Experiences

Ricardo Eugénio Proença Rodrigues

Supervisor : Doctor Carlos António Roque Martinho

Thesis approved in public session to obtain the PhD Degree in Computer Science and Engineering

Jury final classification: Pass with Distinction

2024

TÉCNICO LISBOA

UNIVERSIDADE DE LISBOA
INSTITUTO SUPERIOR TÉCNICO

Improving Believable Interactions in Real-time Multi-party Interactive Experiences

Ricardo Eugénio Proença Rodrigues

Supervisor : Doctor Carlos António Roque Martinho

Thesis approved in public session to obtain the PhD Degree in Computer Science and Engineering

Jury final classification: Pass with Distinction

Jury

Chairperson : Doctor Luís Eduardo Teixeira Rodrigues, Instituto Superior Técnico, Universidade de Lisboa;
Members of the Committee :

Doctor Catherine Pelachaud, CNRS - Institut des Systèmes Intelligents et de Robotique, Sorbonne Université, França;
Doctor Maria Luísa Torres Ribeiro Marques da Silva Coheur, Instituto Superior Técnico, Universidade de Lisboa;
Doctor Carlos António Roque Martinho, Instituto Superior Técnico, Universidade de Lisboa;
Doctor Verónica Costa Teixeira Pinto Orvalho, Faculdade de Ciências, Universidade do Porto.

Funding Institutions - Fundação para a Ciência e a Tecnologia

2024

Abstract

Believable interactions between synthetic characters are an essential factor defining the success of a virtual environment relying on human participants being able to create emotional bonds with such characters. It is important not only for the characters to be believable but also for interactions with and between them too. This work proposes *3motion*, a model for synthetic characters' interactions based on anticipation and emotion, in which the traditional atomic action is deconstructed into three stages: anticipation, action, and follow-through. With these divisions, the model allows for precise non-verbal affective communication, which improves the believability of both synthetic characters and their actions.

We improve the model by expanding on its emotion expression capabilities while adapting it to *Adfectus*, a 3D real-time video game use case in which two characters battle one another. While the characters are controlled by the players, their emotions are expressed independently. To validate and understand the emotion expressions implemented, we created the *TAI method*, which through the use of a triad comparison and multiple data analysis techniques, creates a model of how people perceive expressions.

To validate whether *3motion* in *Adfectus* actually improves believability, we perform several evaluations in which we compare our emotional models to a more reactive approach common in video games. The results show a high level of interaction believability regardless of the version, but no significant differences between emotional models. We did however find that anticipatory models (taking advantage of action subdivision) create characters that are perceived as more aware of others and more capable of expressing unique personalities.

Keywords

Virtual Agents, Synthetic Characters, Believable Interactions, Emotion, Anticipation

Resumo

Interacções credíveis entre personagens sintéticas são um factor essencial que define o sucesso de um ambiente virtual, no qual os participantes humanos são capazes de criar laços emocionais com tais personagens. É importante não apenas que os personagens sejam credíveis, mas também as interacções com e entre eles. Este trabalho propõe o *3motion*, um modelo de interacções entre personagens sintéticas baseado em teorias de antecipação e emoção, no qual a tradicional acção atómica é desconstruída em três estágios: antecipação, acção e conclusão. Com essas divisões, o modelo permite uma comunicação afectiva não-verbal e precisa, o que melhora a credibilidade tanto das personagens sintéticas quanto das suas acções.

Melhoramos o modelo ao expandir as suas capacidades de expressão emocional, ao mesmo tempo, em que o adaptamos ao *Adfectus*, um caso de uso e videojogo 3D em tempo real no qual dois personagens lutam entre si. Embora os jogadores controlem os personagens, as suas emoções são expressas de forma independente. Para validar e compreender as expressões emocionais implementadas, criamos o *método TAI*, que, usando comparações triádicas e múltiplas técnicas de análise de dados, cria um modelo de como as pessoas percepcionam as expressões.

Para validar se o *3motion* no *Adfectus* realmente melhora a credibilidade, realizamos várias avaliações nas quais comparamos os nossos modelos emocionais a uma abordagem mais reactiva comum em videojogos. Os resultados mostram um nível elevado de credibilidade das interacções, independentemente da versão, mas não encontrámos diferenças significativas entre os modelos emocionais. No entanto, descobrimos que os modelos antecipatórios (que fazem uso da subdivisão da acção) criam personagens percepcionadas como mais conscientes dos outros e mais capazes de expressar personalidades únicas.

Palavras Chave

Agentes Virtuais, Personagens Sintéticas, Interacções Credíveis, Emoção, Antecipação

Dedicated to my parents, **Baltazar** and **Gena**, for raising me as you did,
to **Raquel**, for supporting and caring for me every step of the way,
to my long-time friends, **Inês** and **João**, for clearing my head in tough times,
to my colleagues (and friends), **Diogo** and **Manuel**, for making me less socially awkward,
to my new found friends, **Raquel**, **Inês**, and **Pedro**, for helping me get back on my feet, and
to **Carolina**, for all the love she gives.
Without you, none of this would have been possible.

The oldest and strongest emotion of mankind is fear,
and the oldest and strongest kind of fear is fear of the unknown.

H.P. Lovecraft

Acknowledgments

This work was partially supported by national funds through Fundação para a Ciência e a Tecnologia (FCT) with ref. UIDB/50021/2020, and FCT grant from project Tutoria Virtual with ref. TDC/IVCPEC/3963/2014.

Thank you to professor Carlos António Roque Martinho, for the long meetings that made this thesis such a good experience and for always being available, and for taking time from his family to help me.

Thank you to the Masters' students with whom I had the pleasure to work, André Lima, Ricardo Silva, Ricardo Pereira, Taíssa Ribeiro, João da Silva, Vitor do Vale, Guilherme Fernandes. Your work is invaluable and amazing.

Thank you everyone at *GAIPS*, *LabJogos*, and *GameDev Técnico* for your companionship and support.

Contents

List of Figures

List of Tables

1

Introduction

Contents

"In a hole in the ground there lived a hobbit."

J.R.R. Tolkien

In the first film of "The Lord of the Rings"[188], near its end, we see Boromir in a very difficult situation. He is trying to save two little Hobbits from being captured by surrounding orcs. While in difficulty, Boromir continues to fight off its opponents. At the same time, we see, at a distance, Lurtz, a powerful orc, pulling the string of his bow and aiming it at Boromir. For a couple of moments, the camera switches between Boromir, unaware of the new threat, and Lurtz, readying his shot. For a moment, then, time almost stops, while Boromir, in dismay, finally acknowledges Lurtz and the orc releases the arrow.

To know what happens next, you must see the film for yourself, but this moment helps showcase how well-timed interactions can make the viewer feel engaged and the characters rich. This first chapter of the thesis motivates how these moments impact the audiences and how some try to recreate the experience in interactive media, such as video games. We identify some problems with current solutions and propose a hypothesis on how to tackle the problem. Here, we also discuss our main contributions of the thesis and give an outline of the rest of the work.

1.1 Motivation

In today's cinematographic world, many movies conquer audiences and create an imaginary world where the audience loses themselves. This is called immersion, a state of being deeply engaged or involved [1]. But how can films create such immersion? Let us look again at "The Lord of the Rings"[188] films and their fervorous battles. When a character moves or attacks mid-combat it is always clear to the viewer their intention, either by the movement of their eyes or their body: when Legolas shoots his arrows at a distant target, the target is usually shown first, then the camera passes to Legolas preparing and shooting the arrow, and then the target again, being shot[1]. This flow creates anticipation in the audience and a sense of presence, where the audience can feel as being there, which refers to what is called the suspension of disbelief, the notion that the implausibility of something can be suspended for the sake of enjoyment [2].

Many video games try and succeed in creating the suspension of disbelief by introducing pre-scripted scenes and narrowing the player's playable area and actions, giving it a more cinematic feel (see Figure 1.1). Take, for example, the first-person shooter games genre (e.g. *Call of Duty: WWII* [190], *Battlefield V* [184], and *Halo Infinite* [177]), with a campaign where one is put in a soldier's perspective in the midst of a modern/futuristic war, with frenetic and over-the-top scenes on par with many of today's action movies. In this setting, the player can easily feel immersed and feel that they are cooperating with their synthetic companions, mainly due to the constraints of the game. However, in the open-world role-playing games genre (e.g. *The Witcher 3: Wild Hunt* [182], *Horizon Zero Dawn*

[1] Multiple examples of this character's behavior can be viewed in the video with the link `https://www.youtube.com/watch?v=Ousk2CSMvWI`.

[2] Image collected from the video of *Call of Duty: WWII* gameplay with the link `https://www.youtube.com/watch?v=F1Y2h2evV1Y`.

[3] Image collected from the video of *The Witcher 3: Wild Hunt* gameplay with the link `https://www.youtube.com/watch?v=F1Y2h2evV1Y`.

(a) In-game cinematic showcasing two characters interacting in the game *Call of Duty: WWII* [190][2].

(b) In-game interaction with a plot-relevant character in the game *The Witcher 3: Wild Hunt* [182][3].

Figure 1.1: Examples of pre-scripted interactions in video games that showcase believable interaction between characters.

[185], and *Fable II* [187]) this kind of immersion is hard to achieve, mainly because of the interactions between synthetic characters or Non-Player Characters (NPCs) and the player. Often the player can only interact with a few plot-relevant characters and the way one can interact with them is by a set of dialogue options or predefined actions (see Figures 1.1b and 1.2), making such interactions appear unnatural and breaking the suspension of disbelief.

Figure 1.2: In Fable II [187], it is possible to interact with most characters using a set of predefined actions that will increase or decrease a value on an “attractiveness” scale.

The same issues apply in more action-oriented interactions (e.g. combat, racing, fleeing, parkour). In many cases, there is no verbal or nonverbal communication between characters, but much of human communication is made nonverbally, by the movement of hands or eyes, indicating the other’s intent [3–5]; this creates a mechanical and oversimplified battle sequence with no sense of immersion. Emotion expression is also essential to correctly perceiving and delivering an intention [4], and, in many video games and interactive experiences, the expression of emotions is limited to pre-scripted scenes. In *The Lord of the Rings* movies, during a battle, characters frequently show their intentions through gestures, guiding their colleagues to safety, and building on the audience’s expectations.

These interactions lead to the topic of believability, for now, we will use its more obvious linguistic denotation that something can be believed by someone [6] and that when an interaction is believable, it is maintaining the suspension of disbelief.

The study of believability in academia is often geared towards the study and replication of human behavior [6–8], also regularly connected with emotions [9–11]. In this area, the exploration of multi-party interactions is common [10, 12] and focus is given to the simulation of behavior [13, 14]. While the role of real-time interactions is present in many works [7, 12, 15], the focus is not often on the enjoyment of the experience (as in the Game Industry), but rather on determining how believable the characters or the experience are.

Believable interactions are explored differently in academia and the Game Industry and with different ends, we hope to join the real-time interactions more often seen in the Game Industry with the multiparty interactions more often explored in academia to create Believable Interactions in an interactive environment.

1.2 Problem

Video games today strive to lead their player audience into a sense of presence by creating more believable characters and interactions, promoting the suspension of disbelief. Yet we are still far from achieving real-time interactions that the audience would classify as believable. Elsewhere, the academia is striving to explore multi-party interactions and is creating more believable characters interacting with each other, yet they are still far from achieving a level of believability seen in pre-scripted scenes. Part of this problem arises from the limited verbal and nonverbal communication between characters, whether they are controlled by the player or NPCs, particularly during the performance of a significant action.

Faced with this problem, **how can we enable believable multiparty real-time interactions between synthetic characters?** Let us break the question down into its different components. *Believability* refers to interactions that give an illusion of life (it will be discussed in detail in Section 2.1). *Multi-party* refers to the presence of multiple parties in a scene (two or more) that interact with each other, for example, having two NPCs interacting with each other and with the player character (note the importance of interactions between synthetic characters and not only with the player and their character). *Real-time* in this context refers to the dynamic development of interactions, where players experience the scene while actively interacting with it (as opposed to the use of prescripted scenes that remove control from the player while displaying an almost cinematic video). With these concepts in place, we can now tackle this problem in its proper context.

1.3 Hypothesis

We hypothesize that by focusing on improving individual actions, instead of the whole interaction, and that by *subdividing actions according to the traditional principles of animation* [9] and allowing for *emotional reactions and interruption of behavior* during their expression, the overall *real-time interaction will be perceived as more believable*.

We propose the division of an action into three stages: anticipation, action, and follow-through. Each stage takes some time to play out, allowing for emotion expression and interpretation to be done under different contexts of the same action, either by the performer or anyone (human or synthetic) that spectates the action unfold (often also performing an action themselves). By explicitly modeling a split of an action into the three stages, we are able to communicate the intentions of a character in a clearer way and give a richer emotional context for all the characters involved in the scene, consequently improving believability.

Let's demonstrate a scenario showcasing the use of action subdivision. In this hypothetical fantasy-inspired scenario, we will see two characters, Rua and Gorm, during a fight in an arena until one falls.

> Gorm is holding a two-handed axe and Rua a long spear. The battle starts and both fighters run toward each other. As they get closer Gorm decides to perform a very slow but highly destructive overhead attack by pulling his axe back, revealing his intention in the process. Rua perceives this and shows confidence because she believes that the attack will miss. She then decides to thrust her spear towards Gorm. As a result, and still, during the anticipation stage, Gorm becomes scared and decides to cancel his attack to defend himself. Rua's attack then advances to the action stage and is blocked by Gorm, leading Gorm to feel relieved and Rua irritated. Both fighters move away from each other during the follow-through, to prepare for the next attack, and the battle proceeds.

As shown in this scenario, our model ensures that the characters express emotions with the correct timing, which we argue will make interactions more believable.

We proposed *3motion* [16] as a model that implements action subdivision; we will expand on this model in an interactive scenario and use it to support the evaluation of our hypothesis. Within this agent model, we defined the emotional model *BALTAS*, in which we compute the agents' emotions through anticipatory mechanisms, taking full advantage of the action subdivision. Furthermore, we also implemented an approach with an atomic action (i.e. without the subdivision into stages), thus considering only the beginning of the action and its end. This approach simulates a classic view on actions, seen in many interactive experiences today. With these two approaches, we can compare how each performs in terms of believability in a future evaluation.

1.4 Contribution

The main contributions of this work are:

- Review of the state of the art on the concepts of believability and emotions and their expression, specifically in synthetic characters;

- Definition of a computational model, *3motion*, for synthetic characters communicating through an action split between anticipation, action, and follow-through;
- Definition of an emotional model, *BALTAS*, for anticipatory emotional reactions, making use of the action subdivision;
- Implementation of *Adfectus*, an arena game where two characters battle one another, as a use case to test our hypothesis;
- Definition of the Triad Affect Interpretation (TAI) Method, a method to model how people perceive different emotions in specific synthetic characters;
- Evaluation with users to assert the viability of the proposed approach when compared with a classic approach.

During the duration of this thesis, we published several works. Hereafter, we present a list of each published work, categorized by the chapter to which they contribute the most.

- Chapter 3: *Multi-party interactions with* 3motion
  - [16] R. Rodrigues and C. Martinho, "Towards Believable Interactions Between Synthetic Characters," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, J. Beskow, C. Peters, G. Castellano, C. O'Sullivan, I. Leite, and S. Kopp, Eds. Cham: Springer International Publishing, 2017, vol. 10498 LNAI, pp. 385–388.
  - [17] R. Rodrigues and C. Martinho, "Measuring Believable Interactions, a Case Study: 3Motion," in Interaction with Agents and Robots: Different Embodiments, Common Challenges Workshop (IVA 2017), 2017.
  - [18] R. Rodrigues, "Enriching Discrete Actions with Impactful Emotions," in *2019 8th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos*, ACIIW 2019, 2019, p. 39.
- Chapter 4: *A Case Study on Expression of Emotions*
  - [19] A. P. Cláudio et al., "Empathic mediators for distance learning courses," in *Proceedings ICGI - 2018. International Conference on Graphics and Interaction*, pp. 177–178, 2018.
  - [20] R. Rodrigues, R. Silva, R. Pereira, and C. Martinho, "Interactive Empathic Virtual Coaches Based on the Social Regulatory Cycle," in *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*, Sep. 2019, pp. 69–75.
  - [21] R. Rodrigues, R. Silva, R. Pereira, and C. Martinho, "A cautionary tale of side-by-side evaluations while developing emotional expression for intelligent virtual agents," in *Proceedings of the 22nd ACM International Conference on Intelligent Virtual Agents*, Sep. 2022, pp. 1–8.

- [22] T. Ribeiro, R. Rodrigues, and C. Martinho, “Modeling the Interpretation of Animations to Help Improve Emotional Expression,” in *Proceedings of the 2023 International Conference on Autonomous Agents and Multiagent Systems*, 2023, pp. 2301–2303.

- Chapter 5: *Designing Believable Interactions in* Adfectus
  - [23] J. Silva, R. Rodrigues, and C. Martinho, “Believability, Anticipation, and... Timing Improving believability through timing manipulation,” in *Proceedings of the 18th International Conference on the Foundations of Digital Games*, Apr. 2023, pp. 1–11.
  - Multiple works on the final evaluations are currently under development.

1.5 Summary

We started this chapter by exploring the concept of immersion in films and video games. We discussed how clear intentions and well-timed interactions in films create anticipation and a sense of presence, leading to the suspension of disbelief. Video games achieve immersion through scripted scenes and limited player actions. However, open-world role-playing games struggle with immersion due to unnatural interactions with NPCs. In this setting, the lack of verbal and nonverbal communication between characters during action-oriented interactions results in mechanical and oversimplified sequences that break immersion. Faced with this problem, we wondered *how can we enable believable multiparty real-time interactions between synthetic characters?*

We hypothesized that by *subdividing actions according to the traditional principles of animation* [9] and allowing for *emotional reactions and interruption of behavior* during their expression, the overall *real-time interaction will be perceived as more believable*. To demonstrate action subdivision, we presented a scenario with two characters that battle each other. To test our hypothesis, we proposed the use and expansion of *3motion*, an agent model that makes use of action subdivision, along with *BALTAS*, an emotion model with anticipatory computations to take advantage of the action subdivision.

We then described this thesis’ contributions, including a review of the state of the art, a definition of a computational model and a use case, and the evaluations to assert the hypothesis.

1.6 Outline

In the next **Chapter 2**, we start by defining believability and what believable characters are. We then move on to emotions, their expression, and their connection with synthetic characters. Connecting both believability and emotions, we discuss character animation through the principles of traditional animation. We close the chapter by discussing recent works that, like ours, tackle emotional multi-party interactions among synthetic characters. The **Chapter 3** detailing the implementation of the *3motion* model is then presented, considering the subdivision of actions, an illustrative scenario, and preliminary evaluations on a text-based application use case. We then move to **Chapter 4**, discussing the Triad Affect Interpretation (TAI) method to understand the expression of emotions in virtual coaches and how it impacts the way perception of emotions, furthermore, we discuss the implementation of virtual coaches and their emotional system. This is followed by **Chapter 5** which presents a discussion on the implementation of *Adfectus*, where we incorporate *3motion* and *BALTAS* based on the emotional system of Chapter 4, and its multiple evaluations on Believability and on the perception of emotions using the TAI Method. Finally, **Chapter 6** presents a conclusion to this work along with our main findings.

2

Related Work

Contents

"A room without books is like a body without a soul"

Marcus Tullius Cicero

In this chapter we present different but intertwined subjects that are relevant to our implementation and the context it is in. We start by discussing believability, its definition, and what are believable characters. We then move to emotions, their expression, and their connection with synthetic characters. Connecting both believability and emotions, we discuss character animation through the principles of traditional animation. We then discuss other recent works that discuss emotional multi-party interactions among synthetic characters. All these topics help create the context for the definition of 3motion *and the creation of* Adfectus*.*

2.1 Believability

In its more common linguistic denotation, believability means that something can be believed by someone [6], yet there is still no generally agreed or precise definition of believability, instead there is a "family of related meanings denoted by the same word"[6]. The entertainment industry gradually linked believability with the audience's engagement during a performance, while in the context of Artificial Intelligence and video games, we can add that something about a character or even the character itself is believed to be living by someone [7, 8]. Some authors take a *realistic approach*, stating that believability is defined as "someone believes that some character or bot is real" [6, 24] and many recent works give attention to a realistic portrayal of virtual humans [25–28] aligned with this view on believability.

Others take an *artistic approach*, focusing on the ability of a character to provide an *illusion of life* [8, 9, 11, 29], meaning "something [...] that appears to be living, that has an inner strength, a vitality, a separate identity – something that speaks out with authority – a creation that gives the illusion of life"[9]. To give the *illusion of life* does not mean that the character has to be a real living organism, for example, it may not have the many typical channels of expression, such as a mouth or eyes [30]. This can be seen in characters like the flying carpet from the movie Aladdin [192], where, even if it is just a carpet, through its movement it is able to show a personality and even express emotions (see figure 2.1). Furthermore, this definition is regularly associated with the audience's *suspension of disbelief* [11, 13, 29], or the ability to accept "limitations in the presented story, sacrificing realism, and occasionally logic and believability, as well as the media content's aesthetic quality for the sake of enjoyment"[2], and, in interactive scenarios, the player's immersion, "where players forget they are in front of a screen [...], escaping their own physical limitations and [...] assuming their character's identity such that it is 'they themselves'[...] who are in the game"[31].

Togelius *et al.*[6] states that behind the definition of believability there are two broad classes of examples: *Player Believability*, "someone believes that the *player* controlling the character is real, i.e. that a human is playing as that character instead of the character being computer-controlled"[6]; *Character Believability*, someone believes that the character itself is real in a certain context.

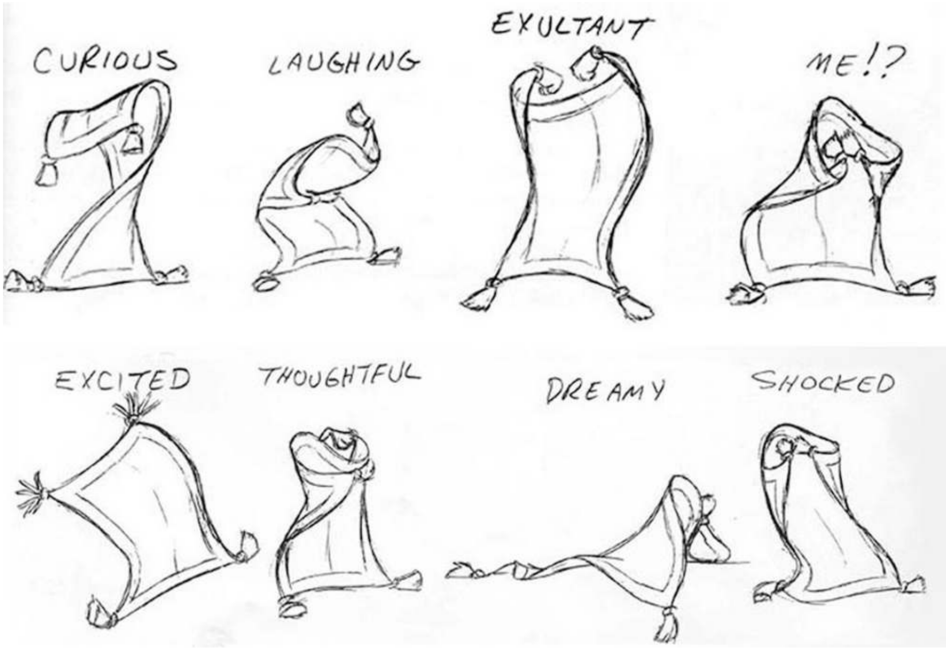


Figure 2.1: Emotion expressions from the flying carpet from the movie Aladdin [192] showing how an object is able to convey emotion.

Player Believability assumes the observer knows the character isn't real and that he "believes that a human has an ongoing input to and control over these processes, and that the human's control is interactive in the sense that the human is aware of what the character is doing in the game"[6].

Character Believability has its research mainly inspired by the roots of two other fields: *drama* and *animation*. When referring to *drama* we can go back to the ancient Greeks, where, according to Aristotle, a believable character "should be able to (1) capture, (2) represent, and (3) project believable states. Whether a believable character possesses any of the properties is irrelevant: it only needs to appear to have them"[7]. Additionally, Prendinger and Ishizuka [32] claim that realistic-looking characters' performance has high expectations from the audience, meaning that little deficiencies lead to user irritation and dissatisfaction, as opposed to an obvious synthetic embodiment. This is also discussed in the problem of the *uncanny valley* where almost, but not completely, real characters tend to be "creepy"[33] and elicit negative emotions in humans.

In *animation*, let us take as an example Chuck Jones' Road Runner[1], a cartoon where a coyote (Wile E. Coyote) repeatedly attempts to catch and subsequently eat the Road Runner without success. During a chase when the Road Runner runs off a cliff and Wile E. Coyote follows, they are both able to stop in mid-air, but it is only when the coyote looks down, becoming aware of his situation, that he falls, while the Road Runner continues running and escapes. "This is the audience's believable world, and the audience will never question the fact that it is not realistic. It is just the way how the world works"[7]. If the rules were broken, believability would be lost and the audience's suspension of disbelief as well: it wouldn't be realistic in the world they created.

Therefore a believable character has to be consistent with its world and also give the illusion of

[1] Wile E. Coyote and The Road Runner. Chuck Jones.

life, "the change of shape shows that a character is thinking, but it is the thinking that gives the illusion of life, and it is life that gives meaning to the expression"[9, 34].

2.1.1 Definition of Believability and its Components

Having explored some different views on believability, we can now define what our working definition will be. Giving focus on the more artistic approach, we define ***believability*** **as "the belief that an environment, its characters, and their interactions give the illusion of life"**.

While Character Believability has been discussed, the *environment* and the *interactions* need to be detailed. An environment gives the audience a context and helps to create expectations of what is or isn't possible. A *believable environment* must create a consistent context where the characters will interact in and interact with. This context encompasses visual style – where matching it with the characters' design is important to place the characters in the world –but also encompasses the affordances it provides to the characters, even if they don't employ them all (e.g. if the setting is realistic, having a character do super-human actions should be justified, otherwise it will appear uncanny). Therefore, a *Believable Environment* is defined as "a place that establishes a context that allows for consistent believable interactions by characters and with characters that give the illusion of life."

This leads us to the question of what is the definition of Believable Interactions? Similar to the previous definitions, believable interactions should provide consistency, in this case of behavior. The interactions performed by a character should be consistent with the character and their intentions. If the interaction is with the environment, then it should respect the context provided; if more characters are involved, the interaction (or multiple concurrent interactions performed by all the characters) should be coherent with the characters, the context, and the relationships between them. Therefore, a *Believable Interaction* is defined as "an interaction that allows for the expression of an action while respecting the restrictions of the environment and being aligned with the characters' intentions and their relationships." A diagram depicting the connection between the three presented topics on believability can be seen in Figure 2.2.

In this work, we focus on interactions, specifically those between Believable Characters (independently of who controls the characters, human or machine) with special attention given to emotion expressions and non-verbal communication. The environment will be given less attention but should be able to provide enough context so that the audience can focus on the characters and their interactions (e.g. provide a floor and walls the characters can navigate and be situated in). Essentially, we focus on *emotional non-verbal multi-party believable interactions*.

2.1.2 Believable Agents

When one combines the artistic aspects of believability and Artificial Intelligence (AI), the result is what is known as believable agents. Their goal is to simulate believable characters in a virtual environment. As in the definition of believability, several authors, from different fields, gave their definition of a believable agent [8, 35, 36]. Bates [35] states that believable agents require "only that they not be clearly stupid or unreal". Such broad, shallow agents must "exhibit some signs of

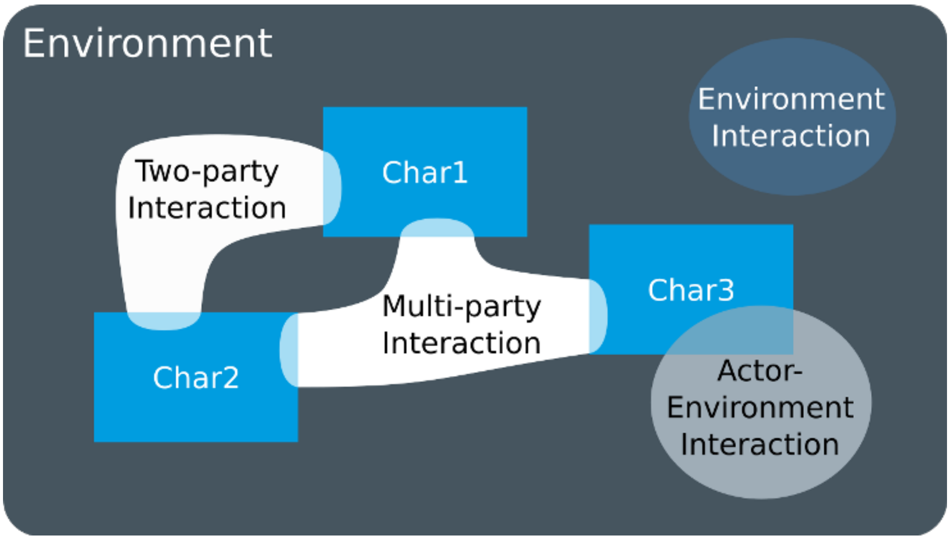


Figure 2.2: Diagram depicting the connection between believable environment (surrounding rectangle), characters (blue rectangles), and interactions (rounded shapes). A believable character exists within an environment, as without it there is no grounding to create the suspension of disbelief (the environment can be simple, but must ground the characters in it). Interactions exist between characters, not only between a couple (two-party interaction between Char1 and Char2) but also multi-party (between Char1, Char2, and Char3), and between them and the environment (actor-environment interaction with Char3). The model also considers interactions without characters, as if the environment interacts with itself (e.g. a tree falling in the forest without anyone being there).

internal goals, reactivity, emotion, natural language ability, and knowledge of agents [...] as well as of the [...] micro-world"[35]. Ortony states "believability entails not only that emotions, motivations, and actions fit together in a meaningful and intelligible way at the local (moment-to-moment) level, but also that they cohere at a more global level – across different kinds of situations, and over quite long time periods"[36]. Nayak *et al.*[8] give us a definition more closely linked to the illusion of life, defining believability as "the ability of the agent to be viewed as a living, although fictional, character", furthermore, the authors link the illusion of life and the suspension of disbelief to the creation of believability, quote "only if the user views the agent as a living character will the user seriously mentally engage him/herself with the agent."[8].

These definitions point to an agent that in order to be believable must have its own goals, be reactive and emotional, and be aware of himself and its world while remaining consistent at a local and global level.

2.1.3 Awareness and Situatedness

Along with believability, concepts emerge that bind the characters into their environment, two relevant concepts are those of Awareness [37, 38] and Situatedness [39–42]. Awareness can be defined as one's ability to "show they perceive the world around them"[37] and "is an essential aspect of believability for virtual characters capable of interacting with humans users"[38]. Situatedness has been defined by multiple authors [39–41]. We will be making use of the following definition as "a theoretical position that posits that the mind is ontologically and functionally intertwined within environmental, social, and cultural factors"[42]. For an agent to be believable it is required that it is perceived as having both awareness and situatedness. Independently of the agent's ability to perceive its environment “correctly”, it should react according to the context. If an agent hears a door close-by opening unexpectedly it is expected that the agent will react by looking at it.

2.1.4 Theory of Mind and Empathy

Two concepts that are critical for creating believable agents are those of the Theory of Mind and Empathy, as they are essential in the creation of believable behavior [10, 14, 43]. These concepts encapsulate or compare with multiple others in psychology, such as *Emotion Contagion* [44, 45], *Interpersonal Affect Regulation* [46], and *Imitation* [47]. Theory of Mind can be defined as “the cognitive capacity to attribute mental states to self and others”[48, 49] and is traditionally associated with logical thinking and critical for simulating social emotions [50]. Empathy can be defined as an “affective response more appropriate to someone else’s situation than to one’s own” [51] and is more often associated with emotional interpretation. Therefore, these concepts are important in the context of the behaviors we will implement. Empathy is given a special focus when discussing Virtual Tutoring in Chapter 4.

2.2 Assessment of Believability

Different approaches arise when discussing the measurement of believability. Hamdy and King [52] defend that “[w]hat can be evaluated [...] are players’ opinions regarding the character’s behavior, interaction, the relationship it managed to forge with them, and the whole experience.” This “[p]layer experience can be measured through subjective, objective, or gameplay-based approaches” [6, 52]. Subjective approaches refer to the measurement through user reports [16, 29, 37, 53], objective approaches refer to the measurement of participants’ physiological responses [54, 55], and gameplay-based approaches refer to measurements based on logging of gameplay statistical data [56–58]. In this work, we will focus on assessment through user-report and through gameplay.

2.2.1 Metrics for Character Believability

We will make use of a user-report assessment based on the work of Gomes *et al.*[37] that has participants play or watch gameplay of a game and then have them express their agreement with a set of metrics about their perception of the agents and how the agents perceived other agents. The metrics, or believable dimensions, are the following:

- **Behavior Coherence:** The audience will evaluate the coherence of a character’s behavior, which is one key aspect of believability [59].
- **Change with Experience:** In the context of interactive narrative, it represents how an agent changed because of a story event, a significant change in a life value of a character.
- **Awareness:** The audience should perceive the agent as being aware of his surroundings.
- **Behavior Understandability:** The audience must understand the agent’s behavior, therefore the agent must express itself in a way that its thoughts and motivations are clearly understood.

- **Personality:** The agent should be perceived as a unique individual. Its behavior should suggest unique personality traits.
- **Emotional Expressiveness:** The agent should be able to express its emotions so that the audience can perceive them correctly.
- **Social:** The audience must be able to acknowledge a social relationship between the agents.
- **Visual Impact:** "The agent should draw the attention of the participant."[60]
- **Predictability:** An agent's behavior must be moderately predictable to the audience, meaning that a very predictable agent will harm believability as much as an unpredictable one, affecting behavior coherence [59]. The extremes should be avoided.

The audience's perception is asserted using Likert scales, one scale per dimension (see the templates for the scales in Appendix A.1). As for emotional expressiveness, participants are asked what emotions are displayed by the agent in specific moments, such as joyfulness or sadness. If the participant's answer matches the emotion the system was trying to reproduce, this dimension would score higher, lowering if the answers did not match. More details on emotion expression are presented in Section 2.6.

2.3 Emotions

When talking about believable characters it is impossible not to talk about emotions, being one of the major factors that make a character believable [9] (with a work going as far as defining believability as "to what extent do you think the character is feeling the emotion"[61]). Unfortunately, there isn't an exact definition for emotion. Kleinginna *et al.*[62] compiled ninety-two disparate definitions for emotions into distinct categories pertaining to the more basic psychological theory they supported (e.g. affective, cognitive, physiological, adaptive) and although there is no concrete conclusion to the definition of emotion, there is a consensus of the view that emotion is considered by most theorists, "as a bounded episode in the life of an organism, characterized as an emergent pattern of component synchronization preparing adaptive action tendencies to relevant events as defined by their behavioral meaning and seeking control precedence over behavior"[63]. We use this definition as it bounds an emotion to anticipatory behavior (shown through action tendencies) and to previously experienced events, allowing for the creation of emotion through information given by the context one is in and their personality, aligning with definitions of believability.

Note that in this definition of emotion, there is no reference to the compartmentalization of emotions, yet in society and in many works (most notably Ekman *et al.*[64, 65]) we find multiple words that represent different categories of emotions, for example Ekman *et al.*[64] defined the six basic families of emotion: *Fear*, *Anger*, *Disgust*, *Surprise*, *Happiness*, and *Sadness* (see Figure 2.3). This act of categorization allows for a better understanding of emotions and their expression in different contexts. In her work, Barret [66] shows that these categorizations are not universal but bound to culture and people, and she hypotheses that “emotion words hold the key to understanding how children learn emotion concepts in the absence of biological fingerprints and in the presence of tremendous variation. Not the words in isolation, mind you, but words spoken by other humans in the child’s affective niche who use emotion concepts. These words invite a child to form [...] concepts for ‘Happiness,’ ‘Fear,’ and every other emotion concept in the child’s culture” [66]. From this hypothesis, we gather that understanding any emotion (defined by a word) is bounded by culture and language. Therefore, we know that the feeling of emotions is varied and takes many shapes depending on the episodes in our lives, and only when we name feelings into words do we compartmentalize and fit them into known “appropriate” words in our culture.

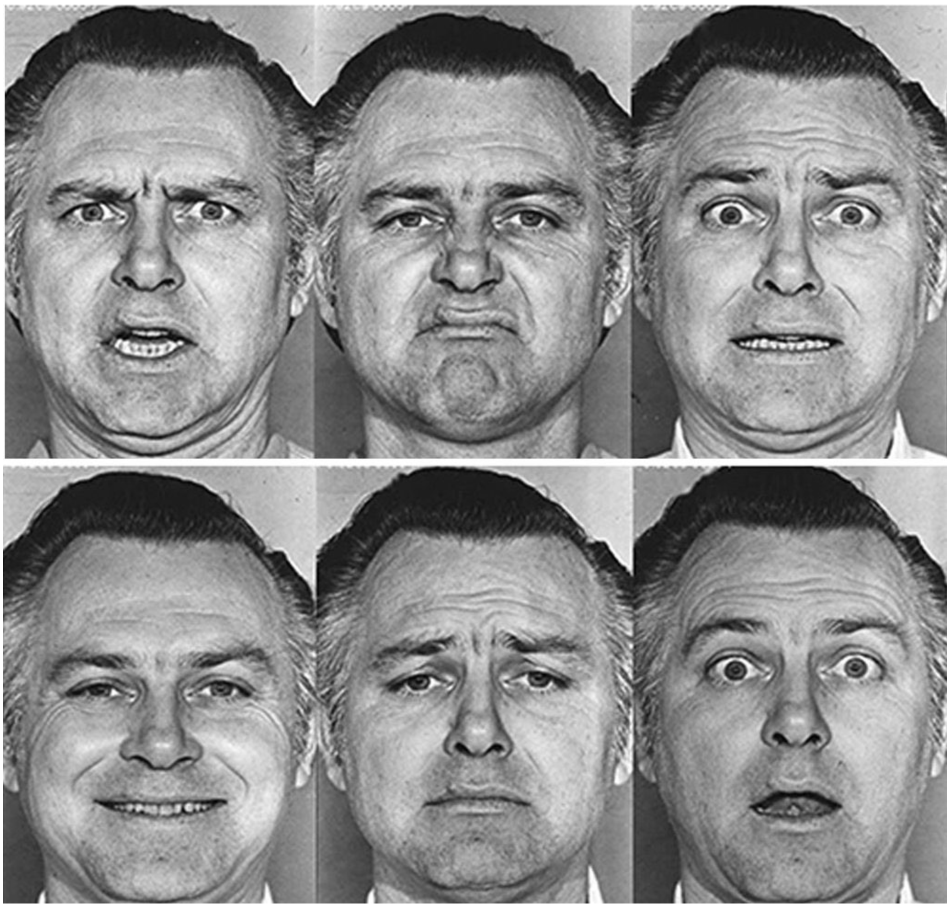


Figure 2.3: The expression of the six basic families of emotion [65, 67], from top left to bottom right: Anger, Disgust, Fear, Happiness, Sadness, and Surprise.

The use of the previously mentioned six basic families of emotion [64, 65] is common in AI works [68–70] as they offer a good compromise between variability and expressivity. Many emotional expressions have been identified for each emotional family [67], and all contain variations in the various emotional modalities associated with them. And although the expression of emotion may not reflect the inner feeling of said emotion [66, 71], in western culture these basic emotions are well known (e.g. in movies such as *Inside Out* [193]) and accompanied by a prototype of the emotion expression, or a mental representation of the best example of emotion expression (e.g. the prototype of sadness expression may be of one crying and sobbing, yet not all instances of sadness are expressed that way [66]), allowing for an easier expression of emotion that will be familiar to those within the culture.

2.4 Emotion and Anticipation

Anticipation and emotions are closely related. One of the emotions' principal functions is precisely that of anticipating events, especially when those events involve the well-being of the organism. *"If I am walking in the woods and, suddenly, 'something' ahead on the path lets out a loud roar, my heart races, my muscles tense, I 'feel' afraid and ready to run away"*[34], in this example, the emotions helped reduce the number of possible actions, by eliminating most of the consequences of each action from consideration a-priori. Therefore creating an **action tendency** or in other words *a desire to behave through select important or communicative actions that are connected to a particular emotion*.

Yet, the anticipation of an event may also elicit an emotion. Let us rephrase the previous example, *If I am* ***hunting*** *in the woods and, suddenly, 'something' ahead on the path lets out a loud roar, my heart races, I found my prey, today I will feed my family, I am happy*. In this case, the emotion was elicited by anticipation. "Through prediction, your brain constructs the world you experience"[66]. These **Anticipatory Emotions** [72] are often related to expectations, commitment toward important goals, and the validation or invalidation of both expectations and goals. Therefore, the same outcome can lead to a wide range of emotional experiences, based on different types of expectations. "To be prepared for what is to come is a crucial factor in survival"[34].

Delving deeper into anticipation (or predictions), in [66], the author posits that we humans actually perceive the world through our predictions of it, which are then tested against incoming sensory input or events that "merely tune your predictions". "Through prediction and correction, your brain continually creates and revises your mental model of the world. [...] But predictions aren't always correct, when compared to actual sensory input, and the brain must make adjustments."

Your brain holds a colossal and ongoing storm of predictions and corrections, arranged in prediction loops [66]. "Predictions become simulations of sensations and movement. These simulations are compared to actual sensory input from the world. If they match, the predictions are correct and the simulation becomes your experience. If they don't match, your brain must resolve the errors." Thus allowing us to "focus our attention on novel stimuli and information"[72].

Thus, in summary, emotions can create predictions of future events and the success or failure of a prediction allows us to feel emotions [7, 72].

2.5 Emotion in Synthetic Characters

When discussing emotion in synthetic characters, among others, there are two perspectives to consider, the *internal representation of emotions*, or how emotions are created, used, and/or processed internally without explicit visual cues, and the *expression of emotions*, or physical behavior that is associated with emotions. Both are important to consider when portraying emotions in a synthetic character.

2.5.1 Internal Representation of Emotions

To answer the question of how one can compute emotions, multiple authors proposed different works [59, 73, 74]. In an effort to incorporate these different affective models, Scherer *et al.*[63] created five categories: *Appraisal Theory Approaches*; *Anatomical Approaches*; *Rational Approaches*; *Communicative Approaches*; and, *Dimensional Theory Approaches*. Each category differentiates itself in what particularity it wishes to convey special relevance or the psychological theory they are backed by. A summary of each category is described in Appendix A.2. We give special attention to the Appraisal Theory Approaches and one in specific, the Emotivector.

Appraisal Theory Approaches postulates that "all emotions come mostly from our own interpretations of events"[73], where our appraisal of the situation is the emotional response. The theory is best used in connecting awareness with emotion, focusing on the individual and their psychological response, where his own judgment of a situation is to blame as the source of his emotional response.

Ortony *et al.*[59] describe an emotional classification that states that emotion is structured into the categories of *Fortune-of-others*, *Prospect-based*, *Well-Being*, *Attribution* and *Attraction*, or more largely grouped into *consequences of events*, *actions of agents*, or *aspects of objects*. This is known as the **OCC Model** and many virtual emotional models that opt for the appraisal theory often base themselves on this model. The OCC Model attempts to incorporate all emotions, but with no relationship between them other than categorical, but not all models based on the OCC model incorporate all emotions. One good example is the model proposed by Ochs *et al.*[13] which focuses on the believability of the NPCs. They attempt to improve the experience by focusing on the personality, social relations, and roles of the NPCs inside the game. The emotions modeled, using the OCC, were joy/distress, hope/fear, and relief/disappointment. There is also an emotional decay component implemented to revert the emotional state to a neutral state after some time period. This model focuses on the NPCs and tries to increase believability through simulation of social relations.

He *et al.*[75] proposed a fuzzy emotional model for virtual agents based on the OCC model. The model incorporates emotions such as hope/fear, satisfaction/fear-confirmation, and relief/disappointment. These emotions are determined by three variables: *desirability*, *importance*, and *likelihood*. Desirability relates to the benefit or detriment of an event to an agent, importance is linked to the motivation intensity of a goal, and likelihood refers to the possibility of actions by other virtual agents.

Using a partial implementation of the OCC model, the work of Jacobs *et al.*[76] focuses on the importance of the link between reward/punishment and emotional response. They propose the use of an emotion label system that converts each agent's state transition, an initial joy/distress mapping that converts into a hope/fear mapping over time, related to the agent's previous knowledge, hopefully allowing the gathering of useful information for planning and decision making of the agent. This model's objective is focused on the modeling of the agent behavior but is important to note how the modeling anticipatory behavior improved the tested agent's performance.

FatiMA [74] offers an interesting appraisal theory application, where the OCC model is implemented by storing appraisals (valence based) in a numeral intensity value (-10;10). This model acting along with goal mechanisms and perceived events, models the complete range of emotions inside the OCC, being able to give individual personalities and coping mechanisms to deal with specific goals. This model however not take into account anticipation of behavior or emotions.

The appraisal theory approaches show potential when used in a more static NPC or Environment emotional association, giving a simpler and robust emotional model to the in-game interactions. We now focus on Emotivector, a model used in our implementation.

2.5.2 Emotivector

Emotivector [7] is a system that focuses on the modeling of emotions by incorporating both anticipation and expectation. By splitting a sensorial input into several categories, according to sensations, a pair between expectation and valence, the model creates a range of emotions that an agent can express. “An increase in a positive sensation or a better reward than expected leads towards excitement, a decrease of a positive sensation or a worse reward than expected leads towards discontentment; a stronger punishment than expected leads towards depression, and a lower punishment leads to pleasure”[7]. Other categories can be made, such as expecting a reward and receiving a punishment can lead to sadness and frustration. This approach offers an appraisal model for virtual agents based on anticipation, a concept also explored in Section 2.4, and is the model used for our implementation of an emotional model. Due to its importance to our work, we will describe it in more detail based on what is described in [7]. It is also important to note that in [20] we developed an emotional model that made use of Emotivectors, the implementation will be discussed in Chapter 4. A final implementation was performed in Chapter 5.

2.5.2.A Behavior of an Emotivector

An emotivector is an “anticipatory mechanism attached to a sensor” and associated with a “one-dimensional aspect of perception”. When an event is perceived by the agent (anything that happens can be considered here, even the behavior of the agent itself), the associated emotivector of that event performs the following steps: (1) using the history of the event, the emotivector computes the next expected value for that event; (2) by confronting the expectation with the actual sensed value, the emotivector computes a preliminary salience for that instance of the event, and; (3) “a sensation is generated according to a model inspired in the psychology of emotion.” This process allows for when a sensation of an event is processed, it carries recommendations such as “you should seriously take this signal into consideration, as it is much better than we had expected” or “just forget about this one, it is as bad as we predicted.”

A – Salience or Attention To determine how relevant a particular emotivector is (how much attention should be given to it), a salience value is calculated. Two components, exogenous and endogenous, are computed to determine the final value. The exogenous component is based on the estimation error and reflects the principle that the least expected (or more surprising) is more likely to attract the attention. If no further information is given, the exogenous value is the only factor contributing to salience, however, if a "searched" value exists, the endogenous component is computed. It reflects whether the change in the distance to the "searched" value is better or worse than the expectation and the level of mismatch. The combination of both exogenous and endogenous components defines the salience. However, "an emotivector with a search value can also provide a qualitative interpretation of the percept", as explained in the next section.

B – Sensation or Emotion In [7], emotions are assumed to be "conditioned responses of primary sensations" and the model is focused on the generation of said sensations. These are defined across two dimensions: valence and change. "The emotivector estimation is used to anticipate a reward or punishment which, when confronted with the actual reward or punishment, triggers one of the [...] sensations. The intensity of each emotion is given by the endogenous component of the percept."

The four most simple sensations considered in the model were the following:

S+ (positive increase) If a reward is anticipated and the effective reward is stronger than the expected, an S+ sensation is thrown.

$+ (positive reduction) If a reward is anticipated but the effective reward is weaker than the expected, a $+ sensation is thrown.

S- (negative increase) If punishment is anticipated and the effective punishment is stronger than expected, an S- sensation is thrown.

$- (negative reduction) If punishment is anticipated but the effective punishment is weaker than expected, a $- sensation is thrown.

The original work (through the use of meta-predictor discussed below) and others that made use of Emotivectors [20, 77, 78], would go on to use more complex versions of the system with more sensations ([20] will be further detailed in Chapter 4).

With these sensations, we can generate emotions based on anticipation (e.g. feeling sad when punishment is greater than expected, S-), but it is also possible to compute emotions based on a combination of multiple sensations to create more believable characters (e.g. perceiving $+ and S- and feeling fear, because not only were you not rewarded as expected, but also you were punished harder than expected).

C – Prediction or Anticipation "The computation of the emotivector salience relies on the capacity of the emotivector to predict its next state." As there is no a-priori knowledge of the signal, [7] it is assumed that the intensity of a signal will change by a random small amount at each discrete time step, before suddenly changing to a random new value. "In other words, the sensed value will tend to remain constant except for certain points in time."

This assumption lead to the implementation of a "Simple Predictor" based on the Kalman filter, "a set of mathematical equations that provides efficient computational recursive means to estimate the state of a process in a way that minimizes the mean of the squared error.” The authors defend that while “this predictor is not as optimal as the ones it is inspired by, it provides a good efficiency/adaptation relation that performs well in real-time over unpredictable signal dimensions, and does not require any previous fine-tuning to work.”

Note that this predictor is applicable in scenarios where data is gathered in high frequency and has low variance (it is also useful when spacial prediction is involved, which was the case in [7]). Other works using Emotivector [20, 78] assume the signal changes with high variance, thus the previous predictor is not fit in those use-cases. In its place, these works use varied algorithms based on moving averages to deal with the difference in frequency.

2.5.2.B Emotivector Management

“[A]ll emotivectors are kept together in a salience module, responsible for their management as a whole”. This manager also needs to determine which percepts are “*a-priori* more relevant to the [...] the agent.” The following strategies are presented in the work: **(1) Winner-Takes-All** that selects the percept with the highest salience, while convenient it “hides” other percepts even if they are just “a little less salient”; **(2) Salience Ordering** creates a salience-descending ordered list that allows for more than one percept to be analyzed while prioritizing the most salient ones; and, **(3) Meta Anticipation** was the approach used in their use case and it makes use of a meta-predictor “whose role is to assess the salience of its associated predictor.” Instead of receiving a percept to predict, it receives the prediction errors of another predictor. “If the prediction error is higher than the meta-predictor estimated error, the percept sensed by the emotivector is marked as relevant.”

The meta-predictor also created the concept of uncertainty on each percept, or “how trustworthy the prediction is”. This allowed the system to take into account other sensations that take into account if the prediction is as expected (e.g. “this percept is as bad as we expected it to be”). All possible sensations are displayed in Figure 2.4.

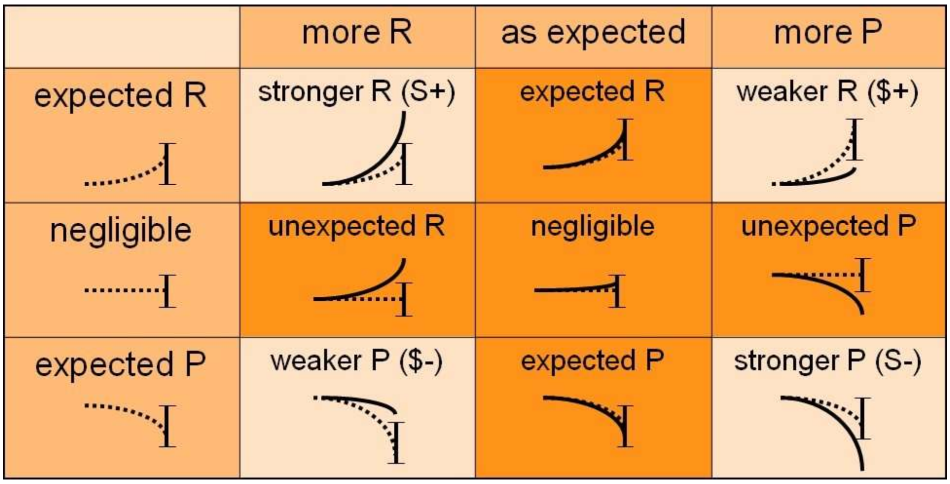


Figure 2.4: Nine-sensation model as seen in [34]: In the figure, R stands for reward and P for punishment. The first line shows possible outcomes when the emotivector is expecting a reward: ‘significantly better than expected’, ‘better, as expected’, and ‘significantly worse than expected’.

As a closing remark, this model is designed to support the creation of believable behavior and is not intended to "substitute any current approach but rather complement the agent architecture for life-like characters."

2.6 Expression of Emotions

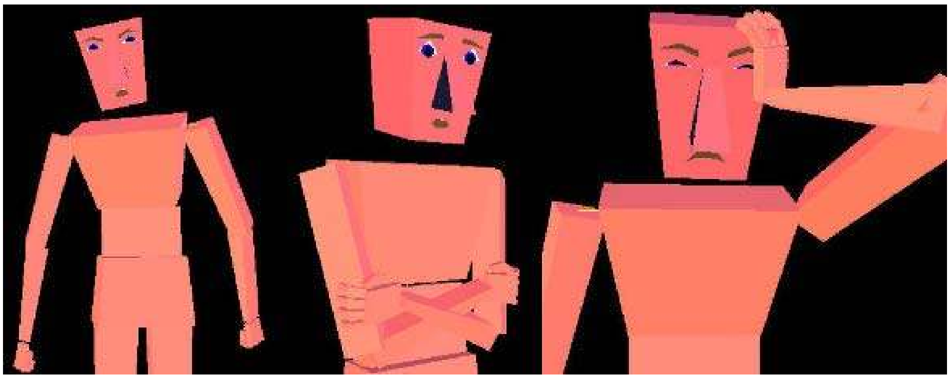


Figure 2.5: Example of expressions of the character used in Nayak's system [8] expressing (left to right) anger, defensiveness, and headache.

An emotional expression is a behavior that communicates an emotional state or attitude (see example in Figure 2.5). Emotions can be conveyed through multiple modalities, such as facial expressions [9, 20, 65, 79], back-channeling [4, 5, 80], and body language [3, 8, 61, 81, 82]. Synthetic characters can make use of most of these modalities and even exaggerate them for the benefit of believability. Furthermore, "it was found that humans tend to interact with computers as they do with real people"[61].

Facial expressions are one of the briefest signals associated with emotion expression, usually lasting only mere seconds [64]. Those that last for longer periods tend to be associated with more intense feelings. Although they do not necessarily correlate with actual emotional experience [66, 71], they are very useful in portraying each of the basic families of emotions [65, 67] (Fear, Anger, Disgust, Surprise, Happiness, and Sadness; see Figure 2.3 and Table 2.1) and, in addition to improving conversation, can also be used to convey dominance and affiliation. One of the more prominent techniques for coding facial expressions is the Facial Action Coding System (FACS) [83]. Developed for measuring facial movement, FACS can objectively score each of the six families of basic emotions, effectively coding them as a series of action units (AU). These have been built over the years. In particular, both Arya *et al.*[68] and Makarainen *et al.*[69] suggested models that combined AUs from the six families of basic emotions in a way that would allow the creation of perceptually valid facial expressions.

Facial expressions are also explored in other mediums, such as in comics, in this setting [84] explores the challenges of designing these expressions, including different kinds of expressions, the muscle groups that are active when expressing them, different strategies to render them graphically, and how these graphics are then applied in comic sequences to make sense and correctly perceived. The author makes use of Ekman's emotions [65] and their combinations to design a range of possible diverse emotions. Perhaps the most relevant subject for this work explored in [84] is how the author

Table 2.1: The facial expressions prototypical characteristics of the six basic emotions [65].

Emotion	Associated facial expression
Fear	The eyebrows appear raised and straightened. The eyes are opened and tense during fear, the upper eyelid raised, and the lower eyelid tense. The mouth opens in fear, but the lips are tense and may be drawn back tightly.
Anger	The eyebrows are drawn down and together. The eyelids are tensed, and the eye appears to stare out in a penetrating or hard fashion. The lips are in either of two basic positions: pressed firmly together, with the corners straight or down; or open, tensed in a squarish shape as if shouting. There is ambiguity unless anger is registered in all three facial areas.
Sadness	The inner corners of the eyebrows are raised and may be drawn together. Raising the lower lid increases the sadness conveyed. The corners of the lips are down or the lip is trembling.
Surprise	The eyebrows appear curved and high. The eyes are opened wide during surprise, with the lower eyelids relaxed and the upper eyelids raised. The jaw drops during surprise, causing the lips and teeth to part.
Happiness	Corners of lips are drawn back and up. The cheeks are raised. The lower eyelid shows wrinkles below it and may be raised but not tense.
Disgust	The brow is lowered, lowering the upper lid. The nose is wrinkled. The upper lip is raised. The lower lip is also raised and pushed up to the upper lip or is lowered and slightly protruding.

describes different graphic strategies to portray emotion: *Realism*, “reproducing the real-life appearance of expressions with realistic tones and details”; *Simplification*, “searching for a few key lines or shapes which clearly convey an expression”; *Exaggeration*, “amplifying the key features that make an expression recognizable”; and, *Symbolism*, “images that depict emotions symbolically rather than with real-world resemblance” (see Figure 2.6). Simplification shows that even with fewer details we can express emotions effectively. This concept was largely adopted in the creation of Emojis (see Figures 2.7), which use iconic representations of emotions in simple but recognizable faces, increasing recognizability with low development effort.

Another important non-verbal component of emotion expression, during social interactions, is *back-channeling* signals. These not only help us deliver our communicative intents [80], thus promoting effective communication [4], but also help us establish our personal tendencies, which in turn help define us as truly unique individuals [5]. Non-verbal behavior gives the listener a chance to provide basic communicative functions toward what the speaker is saying [5], and its absence could lead to negative experiences and incorrect interpretations of the situation [4]. The two most common back-channeling signals are gaze and head motion. *Gaze* establishes the rhythm of conversations, indicates interest in objects/people, improves listeners’ comprehension, expresses complex emotions, and facilitates interpersonal processes [85]. It is worth noting that gaze is not simply relegated to eye movement. Head and body movement/position can be an active component in gazing. These are not independent of one another and, if treated as such, will appear random and disjointed [86]. *Head motion* is an integral part of communication. Nodding can be used to show our agreement/distaste of

[2] Faces by Devendra Karkar from NounProject.com.

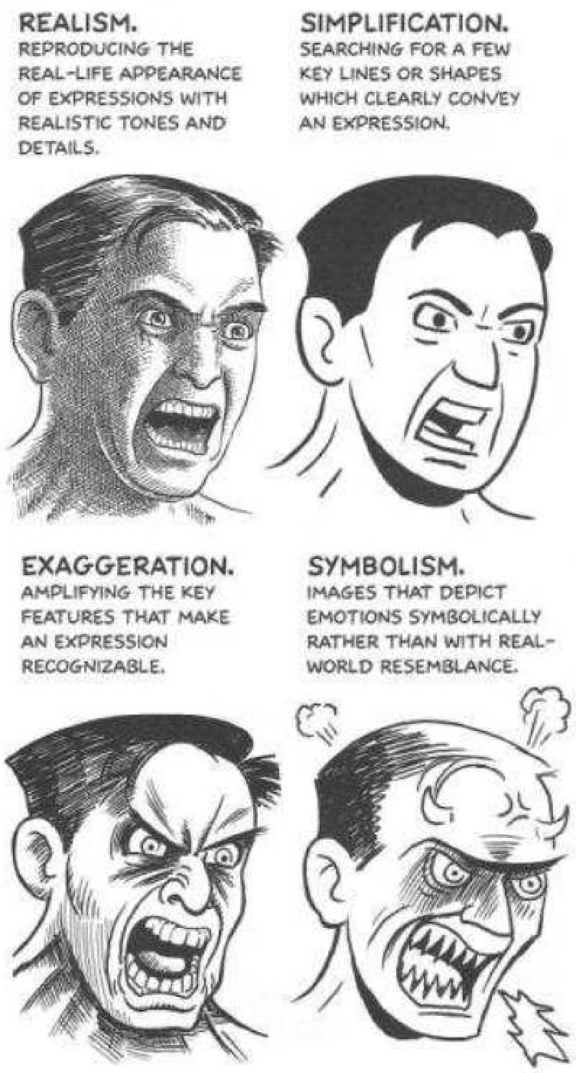


Figure 2.6: Different graphic strategies to portray emotion.

Figure 2.7: Example of different expression of emotions through emojis[2], as a representation of iconic imagery.

a subject matter, as a reinforcement when making a point, or even as a way to avoid unwanted gaze. It also plays an important role with regard to expression. An angry or distressed person will make erratic movements, while a calm, collected person will show very little head motion. This means it can be used to assist in discriminating between emotions [87].

"At the present, most researchers agree that words serve primarily to convey information and the body movements to form relationships"[3]. *Body Language* is often used to express emotion, specifically named *Emotional Body Language*, and is one of the most important forms of nonverbal communication [3]. Furthermore, it is interpreted similarly when portrayed by agents or humans in terms of recognition [3]. They include movements of hands, head, and other parts of the body that allow individuals to communicate a variety of feelings, thoughts, and emotions. Even "[b]reathing is known to participate to the expression of specific emotions"[61]. Do note that "to correctly interpret body language as indicators of emotional state, various parts of body must be considered at the same time"[3]. One can also assume that the correct expression of emotional body language is more important than that of facial expression, as when both "convey conflicting emotional information, judgment of facial expression is hampered and becomes biased toward the emotion expressed by the body"[88]. Noroozi *et al.*[3] provides an example of general movements protocol for six basic emotions presented in Table 2.2.

Table 2.2: The General Movement Protocols for the Six Basic Emotions [3].

Emotion	Associated body language
Fear	Noticeably high heartbeat rate (visible on the neck). Legs and arms crossing and moving. Muscle tension: Hands or arms clenched, elbows dragged inward, bouncy movements, legs wrapped around objects. Breath held. Conservative body posture. Hyper-arousal body language.
Anger	Body spread. Hands on hips or waist. Closed hands or clenched fists. Palm-down posture. Lift the right or left hand up. Finger point with right or left hand. Finger or hand shaky. Arms crossing.
Sadness	Body dropped. Shrunk body. Bowed shoulders. Body shifted. Trunk leaning forward. The face is covered with two hands. Self-touch (disbelief), body parts covered or arms around the body or shoulders. Body extended and hands over the head. Hands kept lower than their normal positions, hands closed or moving slowly. Two hands touching the head and moving slowly. One hand touching the neck. Hands closed together. Head bent.
Surprise	Abrupt backward movement. One hand or both of them moving toward the head. Moving one hand up. Both of the hands touching the head. One of the hands or both touching the face or mouth. Both of the hands are over the head. One hand touching the face. Self-touch or both of the hands covering the cheeks or mouth. Head shaking. Body shift or backing.
Happiness	Arms open. Arms move. Legs open. Legs parallel. Legs may be stretched apart. Feet pointing to something or someone of interest. Looking around. Eye contact relaxed and lengthened.
Disgust	Backing. Hands covering the neck. One hand on the mouth. One hand up. Hands close to the body. Body shifted. Orientation changed or moved to a side. Hands covering the head.

Another example of a generalization of emotional body language is provided by Nayak *et al.*[8] and chooses seven main characteristics to portray a person's mental state (see Figure 2.5): confidence, anxiety, interest, thought, anger, defensiveness, and pain. Each characteristic allows the programmer to set its value on a continuous scale and that affects the movement of the character. When multiple characteristics affect the same body movements, some of them can be averaged out to create a mixed expression. A small description of the general movements protocol for the seven characteristics is presented in Table 2.3.

Table 2.3: The General Movement Protocols for the Seven Characteristics of Nayak's work [8].

Characteristic	Associated body language
Confidence	A complete lack of confidence will lead to a self-doubting hunched-over character. As confidence increases, the character's posture straightens with his chest out and chin up.
Anxiety	As anxiety increases it creates the effect of the character becoming increasingly nervous and fidgety. As anxiety increases, the character's shoulders gradually become increasingly tensed up or shrugged.
Interest	If the character has no feelings towards any object for an extended period of time, he gets bored and may yawn. However, if the character is interested in or is averse to an object, then it holds the character's attention and therefore his gaze, and rotates towards and away from the object, respectively. As the feelings become even stronger, the character begins to walk towards or away from the object.
Thought	The scale goes from lack of thought to deep thought. As the character becomes more deeply lost in thought, his natural fidgeting decreases. When maximally deep in thought, he is almost completely frozen.
Anger	The natural fidgeting of the character is directly proportional to the level of anger. As the anger level rises, the character's body starts trembling, gradually begins to gaze toward the entity at which he is directing his anger. If the anger level of the character rises above a certain point, the character performs the hostile gesture of clenching his fists tightly. If he is walking, he will swing his clenched fists as he walks.
Defensiveness	When defensive, the character will cross their arms and periodically gaze away from the entity of which he is defensive, preventing confrontation and avoiding the impression of aggressiveness. However, on high levels of defensiveness, the character is facing an imminent threat and will not gaze away, instead placing both hands in fists in the air in front of him to physically protect himself in self-defense.
Pain	Pain in the head, stomach, or feet is debilitating. Hence, the character's posture and, if he is walking, the walk speed are both inversely proportional to the level of pain. The character hunches over and slows down as pain increases, also holding the body part in pain.

2.6.1 Secondary Channels

Aside from the aforementioned modalities, secondary channels, such as dialogue or colors, can also be used to convey information and emotion. When considering synthetic characters, verbal dialogue can be communicated not only through audio (e.g. prerecorded dialogues or dictation technology) but also through text which can be used to reinforce the emotional state of a synthetic character.

More than its intrinsic meaning, dialogue in such applications can be enriched by other elements, such as color, kinetic typography, and speech bubbles, which can be used in tandem with characters, as a secondary communication channel.

Starting with *color*, Kaya *et al.*studied its properties in association with emotion [89, 90]. The main findings relate to how symbolism is dependent on each individual, and which aspects (objects or physical space) they associate each color with. Unfortunately, these associations are dependent on personal experience, as well as a social and cultural background such as age, gender, and nationality. Despite being a complex topic and difficult to properly utilize, the richness of their symbolism allows for the conveying of certain emotions [91].

An example of color use is in the film *Inside Out* [193], where emotions are portrayed as characters, and each emotion is associated with a color: *Joy* is yellow; *Sadness* is blue; *Fear* is purple; *Disgust* is green; *Anger* is red. The use of distinct colors for each emotion helps to visually distinguish and emphasize their presence throughout the film. It also aids in conveying the emotional journey of the main character and the interactions between the different emotions within her mind. A possible unintended effect of the movie is the dissemination of an association between a color and an emotion in societies exposed to it (mainly Western society). Another example is in Plutchik's Wheel of Emotions [92] (see Figure 2.8), a model that visually represents eight basic emotions and their various intensities. These eight emotions form the primary colors of the wheel. Different intensities or combinations of these emotions are represented by mixing the primary colors.

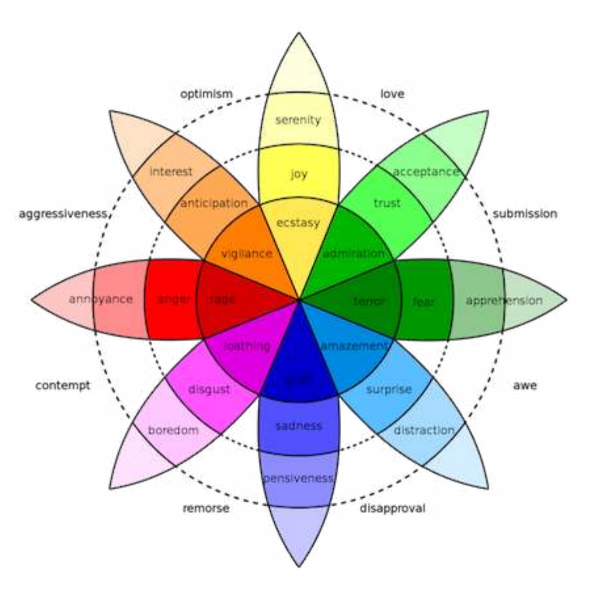


Figure 2.8: Plutchik's Wheel of Emotions. The wheel is organized in a circular pattern, and adjacent emotions are considered related or similar in nature.

Kinetic typography, is a communicative medium that adds some of the expressive properties of film to static text [93]. Studies on such animations show that one of the best ways to convey emotion is by imitating body and sound movements. The results presented by Malik *et al.*showed that specific motions, such as shaking, fading, and bouncing are, depending on the intensity, able to convey emotion, regardless of context [94]. It is important to notice that kinetic typography cannot normally replace or nullify the intrinsic emotive content of a sentence. Instead, it should be used to reinforce it.

Speech balloons have been used in comics for a long time to specify which character is speaking and how they are talking. However, artists create their conventions when conveying the physical or emotional state of a character [95]. Despite the diversity, there are a few bubble shapes that are commonly used in the art form - speech, whisper, thought, and scream. If properly used, these types of balloons can help to convey specific emotions to the user. The previous statements also apply to the use of *backgrounds*. In comics, the background of a scene is also constantly used to specify the state of mind of a character [95], through the use of color, icons, or a combination of both (see Figure 2.9). However, the more ambiguous and lesser-known a shape is, the less meaning users can derive from it.

(a) Expression of terror when a character cannot overcome its despair.

(b) Expression of confidence when a character overcomes a challenging situation.

Figure 2.9: Two expressions of emotions in the game Darkest Dungeon [189]. This game uses a mechanic of accumulation of stress that can make the characters despair (image on the left) or, in rare cases, overcome the stress and become confident (image on the right). To express either situation, the developers made use of a physical expression through the posture of the body and facial expression (when visible), but also through the use of the background, mixing color (red for negative motifs and white for good motifs) and symbols (circle with spikes expressing stress, circle modeled like clouds to express confidence).

All elements discussed have unique properties that allow them to express or reinforce the emotion being expressed, creating different emotional communication channels. The challenge arises from attempting to utilize them in unison to effectively express emotional states. This topic is discussed in more detail in Chapter 4.

2.6.2 Emotion Expression in Video Games

Different media, such as movies, cartoons, and video games, use different ways of expressing emotions. They used the previously mentioned methods, through facial and body expressions, but, as we have seen in the last section (see Figure 2.9), they also make use of other secondary channels. Here we will summarily explore different emotion expressions seen in video games that will be useful for this work.

Story-driven games, such as the ones presented in Chapter 1 (see Figures 1.1a and 1.1b), make use of facial and body expressions, mainly when in cinematic sequences. When outside of these sequences and when players have control of the character, some games make use of Control Modulation [82], “the process by which we take a player’s input and apply filters to it before applying the result of those filters to the virtual character the player is controlling”. This modulation allows for *Emotional Body Language* through the use of prototypical postures to display different emotions. For example, in the game *Tomb Raider* [183], the animations adapt to the character’s current state, where she walks slower and with a hand on her waist when hurt (see Figure 2.11).

Alternatively, the nature of games with fast-paced settings, such as fighting or shooting games, where players will likely be immersed in the action, the use of subtle expressions of emotion is less effective. Other factors, such as the distance to the camera and other elements on the screen, might also impact the perception of the character’s expressions. In these cases, the use of emotional body language needs to be exaggerated and almost iconographic (in the sense that the expression can be viewed easily and easily associated with the respective emotion). The same applies to the facial expressions used. Video games make use of Emotes that portray the characters in iconic and identifiable poses to express emotions or moods (see Figure 2.10).

In other video games, for example, those that use a third-person camera perspective, the main character’s facial expressions may be obstructed. To solve this issue, some games make use of HUD[3] elements to express the character’s expressions (see Figure 2.12). Other HUD elements are also used to provide information on hidden elements, such as enemy characters or other objects requiring the player’s attention (see Figure 2.13).

All these modalities help the player be aware of the characters, their expressions, and the environment around them. While some are more important to the emotion expression, all contribute to the overall perception of the current state and are useful for this work.

2.7 Recognition and Assessment of Emotion Expression

When discussing emotion recognition it is important to discuss the concept of Emotional Intelligence, meaning the ability to know one’s own and other individuals’ emotions and intentions, and the capacity to use such information to adapt to certain environments [96, 97]. Many models for Emotional Intelligence have been proposed with different views on the base definition. A particular model was presented in the book “Executive EQ” and defined 4 cornerstones: “*emotional literacy*, which includes knowledge of one’s own emotions and how they function; *emotional fitness*, which includes emotional hardiness and flexibility; *emotional depth*, which involves emotional intensity and potential for growth and *‘emotional alchemy’*, which includes the ability to use emotion to spark creativity”[96]. This concept and model are important to take into account when discussing the evaluations with participants, as different people have different degrees of Emotional Intelligence and that can impact their

[3]In video games, the heads-up display is the method by which information is visually relayed to the player as part of a game’s user interface.

Figure 2.10: Emote of a character in *Overwatch* [181] expressing a burst of happiness. The exaggerated display makes it recognizable even if far from the character.

Figure 2.12: Heads-up Display (HUD) showcasing characters' expressions. On the top of the game *Warcraft III: Reign of Chaos* [179], you can see a character and its stats, like health and mana, while performing facial expressions and talking. At the bottom of the game *Doom* [186], the character's health is expressed through both a percentage and the facial expression of the character.

Figure 2.11: Example of animation adapting to the state of the character in *Tomb Raider* [183]. It shows the character walking normally (left) and walking while it is hurt (right), note the arm holding the area where the character is hurt. Lighting is also used to highlight the gravity of the situation, using darker tones to portray a bad feeling.

Figure 2.13: Screenshot from *Assassin's Creed Odyssey* [191] highlighting a radial element of the HUD that indicates an incoming ranged attack. This element provides a visual cue to otherwise hidden interactions.

perception of the expression of emotions.

Studies found that neither younger nor older adults are able to correctly identify most emotions with 100% certainty. Additionally, Mill *et al.*found that the correct recognition of negative valences decreases with age, starting at the age of 30 years old [98]. This suggests that recognition varies not only from emotion to emotion but also from individual to individual. Regarding probable causes, Bassili [99] presented a discussion in which he points to similarities/ambiguities between facial expressions, alongside poor actor performance, as the likely contributors to erroneous recognition. Additionally, Barrett defends that without context it is almost impossible to recognize emotion [66]. Another possible issue is the use of multiple emotional channels. Mower *et al.*[100] presented evidence that adding a second channel can create channel bias, which manifests itself when one channel is more predominant and therefore captures the receiver's attention, or when two channels transmit contradictory cues. This is an important concern when combining multiple distinct channels, like the ones discussed in Section 2.6.

Having discussed emotion recognition, we can now discuss the assessment of Emotion Expression. Lottridge *et al.*[101] reviewed relevant research on emotion and its measurement, discussing how one can measure a participant's affective state through physiological measurements [102–104] and self-report [61, 105–107]. Focusing on the participants' assessment of emotions of others, a short review of works revealed that the participants' input is often extracted through reports, either about other people [61, 108–110], virtual agents [61, 111], or anything in general (usually for labeling) [112, 113]. Other works extract the participants' assessment through their physiological responses to interactions with others or, more specifically, with virtual agents [114, 115]. In this work, we will focus on the assessment of emotions of others through self-report to evaluate our agents.

Two common methods to evaluate emotions of the self or others through self-report are the *Geneva Emotion Wheel* [116] and the *PANAS* [117].

2.7.1 Geneva Emotion Wheel

The *Geneva Emotion Wheel* [116] is a psychological framework that aims to categorize and understand human emotions. The wheel provides a comprehensive representation of emotions by organizing them into a circular structure (see Figure 2.14). It consists of three main components: emotion dimensions, emotion categories, and intensity levels. The emotion dimensions are the fundamental aspects along which emotions are organized. These dimensions include valence (positive vs. negative), potency (strong vs. weak), and activity (active vs. passive). They capture the subjective experience of emotions.

The emotion categories are represented as segments on the wheel, each representing a specific emotion. There are six main categories: surprise, joy, anger, sadness, fear, and disgust. Each category is further divided into subcategories, resulting in a total of 20 discrete emotions. For example, the joy category includes subcategories such as happiness, relief, and pride.

The intensity levels represent the strength or degree of an emotion, ranging from low to high. They are indicated by the size of the segments on the wheel. Larger segments indicate stronger emotions,

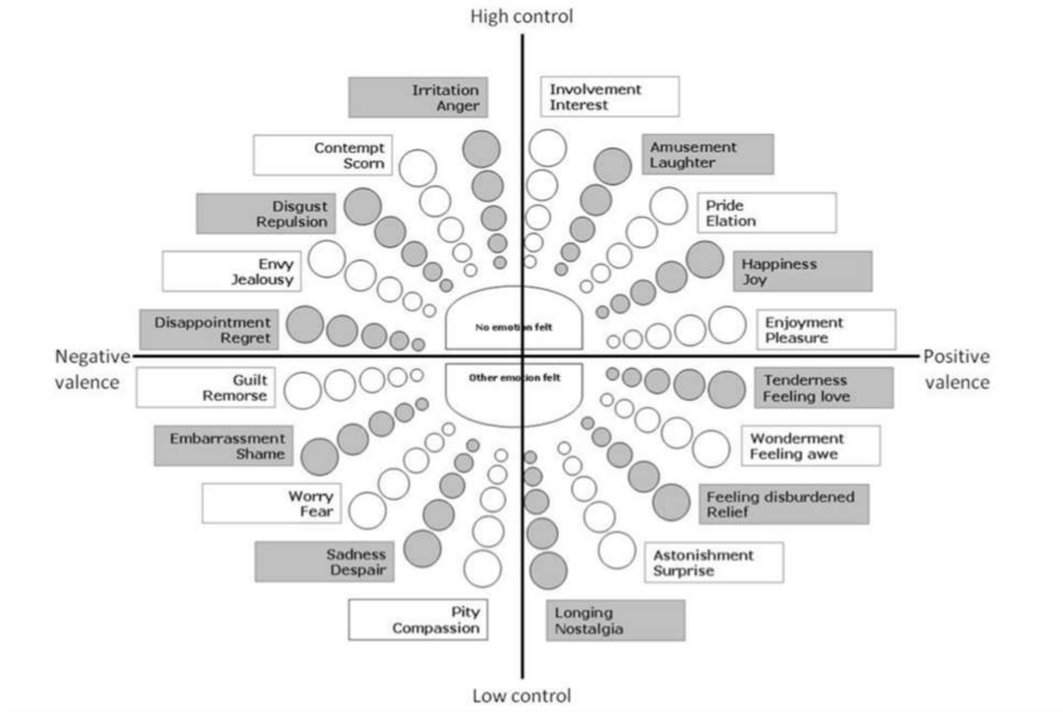


Figure 2.14: The Geneva Emotion Wheel. Showcasing all represented emotions in the wheel depending on valence (positive or negative) and control (high or low).

while smaller segments represent milder or less intense emotions.

The Geneva Emotion Wheel provides a visual framework that helps researchers and psychologists analyze and study emotions. With the wheel, participants can pinpoint how they perceive certain emotion expressions, and researchers can check if the expression is being correctly perceived.

2.7.2 PANAS

The *Positive and Negative Affect Schedule* (PANAS) [117] is a psychological measurement tool used to assess an individual's positive and negative affect, which are two components of emotional experience. The PANAS consists of two separate scales: the Positive Affect Scale and the Negative Affect Scale. Each scale consists of a list of adjectives that describe different emotions or feelings. The individual being assessed rates the extent to which they have experienced each emotion or feeling over a specific period of time, typically in the past week. The ratings are usually done on a Likert scale ranging from 1 (not at all) to 5 (very much).

The Positive Affect Scale includes adjectives such as "interested," "excited," "enthusiastic," and "happy." It measures the extent to which an individual experiences positive emotions or feelings. The Negative Affect Scale includes adjectives such as "distressed," "nervous," "angry," and "guilty." It measures the extent to which an individual experiences negative emotions or feelings.

The PANAS provides researchers and psychologists with a standardized and reliable way to measure and compare positive and negative affect across individuals and groups.

2.8 Principles of Traditional Animation

People are “remarkably adept at recognizing the actions performed by others” [118] and some works “point to the existence of neural mechanisms specialized for registration of biological motion” [118]. With no more than 12 point-lights portraying biological motion, “people can reliably discriminate male from female actors, friends from strangers” [118].

When we apply biological motion to animation, it requires work to appear believable. The animated character, while not needing to be realistic, should move and behave consistently with the way it looks [61, 119], and discrepancies in behavior were found to be disturbing [119]. The principles of traditional animation, first introduced in the book *The Illusion of Life: Disney Animation* [9], mitigate this problem by defining standardized practices followed by Disney's animators to create believable animations (see Figure 2.15), both traditionally drawn and digital [31, 120]. Below follows a description of some of the principles.

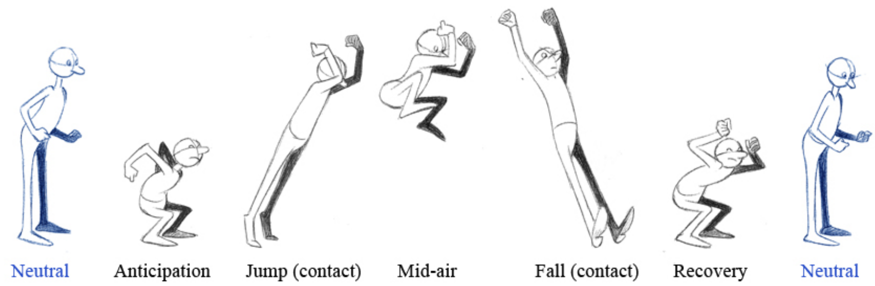


Figure 2.15: Stages of a jump animation. In this figure, we can see some of the principles of traditional animation, such as *Anticipation*, notice the stretching of the arms and legs in the anticipation, preparing the viewer for the jump, which also showcases the *Staging*. When landing, the *Follow-through* principle is shown in the exaggerated motions of the fall and its recovery back to the neutral position.

- ***Timing***, or speed of an action, defines how well the idea behind an action will be read by an audience. More importantly, timing defines the weight of an object, as in the example “a giant has much weight, more mass, more inertia than a normal man; therefore he moves more slowly. (...) he takes more time to get started and, once moving, takes more time to stop.”[120].

  One can also define the emotional state of a character by its movement, where the varying speed of an action indicates whether the character is lethargic, excited, nervous, or relaxed.

- ***Anticipation*** is the preparation for the action, for example, if a character wishes to grab a cup of coffee they first raise their arm and stare at the cup, broadcasting their intention, which leads those watching to expect the character to pick up the cup before the action is done. Without anticipation, many actions are abrupt, stiff, and unnatural.

  An exaggerated anticipation can also emphasize the weight of an object when a person has to bend down to be able to pick up a heavy crate, or used to show a character's emotional state when one is scared or anxious about doing something he must do.

Additionally, in video games, the anticipation in NPC actions or attacks (called telegraphing) is desirably longer, as it informs the player that they must react to something incoming [31].

- ***Staging*** "is the presentation of an idea so it's completely and unmistakably clear"[120]. This principle declares that to clearly stage an idea the audience must be led to be paying attention exactly to what the creator wants them to, otherwise the idea will be missed.

  When staging an action, it is important that only one action be passed to those watching, to do that there should be a contrast between the object to focus on and the rest of the scene, for example, in a big crowd walking in the side-walk, a person standing still will attract the viewer's attention.

- ***Follow-through and Overlapping Action*** – Most times an action does not come to a sudden stop after it is complete, in many movements like a jump there is the termination of the action or Follow-through. You can see an example in the recovery in Figure 2.15, where the action is carried past its termination point.

  Overlapping action "covers the notion that different parts of a character's body will move at different rates"[31]. It can be variations added to the timing and speed of the loose parts of objects or of different body parts during the animations. This makes the objects seem more natural and maintains a continual flow between the phases of an action.

- ***Exaggeration*** is self-explanatory, but it has to be done with care. It can work with every component, but not in isolation. The exaggeration of various components must be balanced, where some elements are exaggerated and others are used as natural elements for the viewer to use as a comparison so that the scene remains believable.

  When animating characters, exaggeration is very important to transmit their emotional state. If a character is sad, make him sadder; if he is wild make him frantic. The exaggeration allows for a better interpretation of the emotion being displayed [61].

- ***Secondary Action*** "is an action that results directly from another action"[120]. It is important since it adds realistic complexity to the scene, but must always be kept subordinate to the primary action. "Secondary action is something that is under the character's control that embellishes or enhances the primary action in a way that adds character or personality."[4] While the Overlapping Action is an involuntary response to the motions of the primary action, the secondary action is a voluntary response by the character to express personality.

  Although secondary, this type of actions will be very important to this work, since we will consider the reply to the primary action, of the character and of those who watched, to be secondary actions.

There are other principles that were not described above as they are not a focus of this work: **Squash and Stretch** principle defines the rigidity and mass of an object by distorting its shape during

[4] `https://medium.com/frame-by-frame/anticipation-follow-through-overlapping-action-and-secondary-action-f80069aba725`

an action; **Straight Ahead Action and Pose-To-Pose Action** are two approaches to the creation of movement; **Slow In and Out** principle that specifies the spacing between frames to achieve subtle movements and timing; **Arc** is a visual path of action for a believable movement; **Appeal** principle describes how to create an action that the audience enjoys watching; and, **Solid Drawings**, this principle refers to the importance of creating characters or objects with a sense of volume, weight, and three-dimensional solidity.

All of these principles help create believability in the characters, their interactions, and the environment around them. To close this discussion, here is a quote from the authors of the book,

> *"It is more than a drawing and more than an idea, possibly it is the love we feel for characters so heroic, so tender and funny and exciting – all of them entertaining, yet each different, each thinking his own thoughts, and experiencing his own emotions. That is what makes them so real, and that is what makes them so memorable. It is also what gives them the astounding illusion of life."*[9]

2.8.1 Five Fundamentals of Game Animation

"The 12 animation principles are a great foundation for any animator to understand [...]. Ultimately, however, they were written with the concept of linear entertainment like TV and film in mind"[31]. Having a more dynamic medium in mind, in the book *Game Anim* [31], the author proposes the *Five Fundamentals of Game Animation*. A set of principles "unique to game animation that [...] complement the originals"[31] and form a basis of interactive characters that not only are well portrayed (making use of the 12 principles) but also take into account the possible interactions with (or being controlled by) the player. Below follows a short description of the proposed principles.

- ***Feel*** focuses on creating animations that have a responsive and tactile quality. It involves making the character or object's movements feel satisfying and connected to player input, enhancing the overall game feel. The animation should provide a sense of weight, impact, and responsiveness to the player's actions. The visual feedback of these animations is also important to consider, as it will guide the perception of the actions and inform the player how the action impacts the world.

- ***Fluidity*** refers to the smoothness and natural flow of animations. It involves creating seamless transitions between different actions and movements, avoiding abrupt or jerky transitions. Smooth animations contribute to the immersive experience and enhance the visual appeal of the game.

- ***Readability*** emphasizes the clarity and readability of animations to convey information effectively to the player. The animations should clearly communicate the intended action, state, or reaction, allowing players to understand and respond accordingly. Distinct poses, clear timing, and well-defined silhouettes contribute to the readability of animations.

- ***Context*** fundamental involves considering the in-game context and gameplay mechanics when designing animations. Animations should align with the overall game design and contextually fit the specific interactions, environment, and narrative. Contextual animations enhance the player's understanding and immersion in the game world.
- ***Elegance*** refers to finding efficient and streamlined ways to implement and integrate animations into the game, taking into account the limitations and resources available. Animators should strive to create and/or use animation systems and workflows that are straightforward, and effective, considering factors such as budget and development time. "Good design involves 'as little design as possible' "[31].

Along with the previously presented principles, the 5 fundamentals of game animation help solidify the basis on which to design interactive characters and, as before, help create believability in the characters, their interactions, and the environment around them.

2.8.2 Anticipation and Execution

We will make use of the principles to define our action subdivision and emotion expression model. Our work will follow a similar approach to the work of Costa [60]. In his work, he proposes a new approach to agent communication and cooperation by dividing an action into two stages, *anticipation* and *execution*, following part of the principles of the traditional animation.

The *anticipation stage* serves the purpose of broadcasting the intent of an agent so that every other "is expecting it and can prepare accordingly". After the broadcast, the agent may choose to either execute or cancel their action, based on the other agents' responses. If the agent chooses to execute the action it enters the *execution stage*, where processes like path-finding and animation playing will allow the agent to execute the action. The agents' stages may overlap at any time, thus the agents must be aware of each other's current state of action, so they can effectively cooperate.

To determine if an agent continues from anticipation to execution, Costa proposed the use of a *Confidence value*. For each action, each agent has a confidence value that anticipates if the agent's action will be successful. When the agent broadcasts his intention to perform an action, his confidence will increase or decrease depending on the other agents' reactions. If the confidence value is below a certain threshold, then the agent cancels the action, feels frustrated, and the process starts all over again.

The confidence threshold is dependent on the agent's personality. "Just as feedback impacts different people differently in the real world, it should impact different agents differently"[60]. Additionally, the outcome of the action can also influence the confidence associated with future intents, creating an adaptive agent.

This approach gives a good foundation to build upon, incorporating the notion of a multi-staged action and the broadcast of intentions.

2.9 Emotional Multi-party Interactions among Synthetic Characters

Until now, through the exploration of related work, we created a context that allows for the discussion of a set of topics, i.e. multi-party interactions, emotion expression, real-time interactions, and believability. These topics are present in both the problem we want to solve as well as in our proposed hypothesis (recall Section 1.2). Considering their relevance to our work, are there other works that tackle all or most of these topics?

We collected relevant works from AI-related conferences[5] held in 2021. From the collected works, those tackling multi-party interactions tend to focus on decision-making, but lack emotion expression [121–126], while those showing emotion expression tend to focus on human-agent interactions (with a single agent), but often lack in multi-party interactions [127–136]. Real-time interactions are regularly present in human-agent interaction works, but often it is not the object of study (e.g. in [129] a study portraying real-time interactions is conducted to examine "whether social agents can elicit the social emotion shame as humans do"). Believability is not commonly present in recent works, yet when present it appears as an objective of the work [122, 126, 137–139].

Works tackling most of the relevant topics are not common, but a few exist [121, 122, 140, 141]. In [121], communication between reinforced learning agents is explored and a new communication mechanism is presented to complement other existing ones. The authors ran simulations (in the game StarCraft II [180]) to measure the success of the application of the communication mechanism. This mechanism focuses on the communication of information (so that other agents learn from that) based on novelty, in other words "Communicate what surprises you"[121] (the common approach is to "Communicate what rewards you"). Results show a greater degree of success when agents use both approaches simultaneously, that is they share what rewards and what surprises them. This work showcases what information is relevant to share and its approach on surprise is similar to what is seen in the Sensation salience in Emotivector (see Section 2.5.2).

In [122], while focusing on decision-making, the authors take special attention to the believability of behavior and real-time interactions in multi-agent environments. In this work "a method that allows [for NPCs acting as] guards to exhibit a real-time search behavior for an opponent that better exploits awareness of the level geometry" was introduced. They explore a novel approach to path-finding and search behavior based on a skeletal graph representation to better propagate the probability of potential opponent locations. Their approach is shown to display interesting behavior for multi-guards searching for an opponent on several maps. While no believability evaluation was conducted, the concern for believable behavior (in favor of realistic or optimal behavior) is relevant to our work and their approach can later inform how we want to enable decision-making in our agents and portray believable behavior.

In [141], the authors offer a "platform that researchers and developers can use as a starting point

[5] International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS), IEEE Conference on Games (CoG), International Conference on Intelligent Virtual Agents (IVA), Conference on Affective Computing and Intelligent Interaction (ACII), International Conference on Human-Agent Interaction (HAI).

to setup their own multi-IVA applications." In it, they implement multiple integrated components, such as topic section engines, dialogue engines, and behavior realizers, so that the agents can seamlessly interact with each other. Furthermore, "the viability of the platform has already been shown by uptake not only in student teaching, but also in at least 4 funded research projects that employ the full setup". This work helps us understand how a multi-agent platform is created and how every component is connected.

In [140], an exploration of how virtual agents react to non-verbal communication is performed. Three agents are depicted interacting with each other, one performs a wave (pre-programmed action of an agent waving his arm), while the others decide to wave back (or not) if they perceive that the wave is directed at them (through the analysis of body direction and other motion features). The work is still early in its development[6], but the authors say that in future works "[i]n a multi-agents scenario, [this approach] would allow for the design of not fully predefined non-player character reactive behaviors." The work offers an interesting view on non-verbal communication between synthetic characters, showcasing the potential of using pre-scripted animations in less scripted (or more spontaneous) interactions. These types of interactions appear to be closely linked to the expression of back-channeling signals discussed in Section 2.6, further motivating these almost instinctive (or less conscious) behaviors.

These works help us understand the current state of the art surrounding the previously mentioned topics and solidify the validity of previously mentioned subjects (such as back-channeling and salience). These works also showcase the existing gap we are filling with our work, further motivating investigation on the subject.

2.10 Summary

This chapter focused on the exposition and discussion of works that are relevant to our problem, hypothesis, and implementation. As a review, our problem is focused on how can one enable believable multi-party real-time interactions between synthetic characters. To address this problem in the correct context we presented related work on believability and its different definitions by different authors. We proposed our own definition for believability and also believable environments, characters, and interactions. Surrounding this topic, we delve into what believable agents are, how they need awareness and situatedness to be grounded in the world, and how theory of mind and empathy allows characters to bond and be more believable. We then discussed works that tackle the assessment of believability.

One of the crucial elements that make a character believable is emotion, thus we then explored how emotions are defined and how they are expressed. We discuss their link with anticipation and how one influences the other, through action tendencies, and vice-versa, through context. We then move to how emotions are used in synthetic characters, looking into their internal representation, how emotions are created, and their external representation, how emotions are expressed. Special

[6]As of July 3rd, 2023, no new works seem to have been published as a continuation of [140].

attention was given to Emotivector, an emotion model that makes use of anticipation to generate emotions, and emotional body language, which describes how we view emotion expression in body movements. To close the study on emotions, a review on recognition and assessment of emotion expression is shown.

Connecting believability, emotion, and expression we discuss the principles of traditional animation. These principles are a set of standardized practices followed by Disney's animators to create believable animations and can be applied to digital environments. Afterward, we discuss the subdivision of an action applied to an interactive experience to help create more believable interactions and characters.

Finally, we review a selection of recent works that explore multi-party interactions through communication and emotion expression between synthetic characters.

3

Multi-party interactions with *3motion*

Contents

"We do this not because it is easy, but because we thought it would be easy."

Raquel Oliveira

In this chapter we discuss 3motion*, first developed in [142] and expanded here, is a model that splits actions into separate stages based on the principles of traditional animation [120] (see Section 2.8), aligned with an emotion expression model based on anticipation, it allows the creation of synthetic characters that can more easily react to others and their actions, making use of non-verbal communication. What follows is a discussion of its inner workings, as well as a summary of the first evaluation, performed using a text-based application use case.*

3.1 Action Subdivision

To enable believable multi-party real-time interactions between synthetic characters, we created *3motion*[1] [16, 142], which approach deconstructs the traditional atomic action, generally used when implementing synthetic character behavior in virtual environments, into three distinct stages[2]: *anticipation*, *action*, and *follow-through* (see Figure 3.1).

Each stage may take a certain time to play out, and interactions at different points in the sequence of events will have a different meaning for the other agents participating in the scene, as well as for the viewer passively watching or actively interacting with the scene, allowing for the creation of a rich interaction with multiple agents and/or users (certain stages can be interrupted, allowing for more unique interactions to be created).

The *anticipation stage* serves the purpose of communicating the intent so that all other agents understand it and can expressively prepare for it [60]. At this stage, an agent broadcasts its intent and associated emotion and receives feedback in the form of emotions from the other agents who are aware of the expression. This allows an agent to interpret the emotional reactions in the context of its intentions and modulate its behavior accordingly. This stage is further subdivided into two sub-stages: an *interruptible stage* in which the agent is still able to cancel the initiated action, and an *uninterruptible stage* in which the action reached a point where it cannot be stopped. The precise timing of the expression and affective responses is crucial in these stages, otherwise, the audience may not be able to understand how decision-making took place in this shared affective context.

The *action stage* is instantaneous and only exists conceptually. This stage represents the moment the action is resolved and the state of the virtual world changes.

The *follow-through stage* is entered after the action is resolved and its result is broadcast, which will meet or challenge the expectations of the agents. As in the anticipation stage, the agent sends its affective appraisal to others, allowing them to feel happy, sorry, etc., and receives their affective feedback to perform a final appraisal of the action. This stage is also subdivided into two sub-stages: an *uninterruptible stage* in which the action is unable to be stopped, and an *interruptible stage* in which the agent regains the ability to terminate the action and start a new one. While in the interruptible

[1] The name *3motion* originates from the combination of Emotion and the three (3) stages of a subdivided action: *anticipation*, *action*, and *follow-through*.

[2] This approach was inspired by the work of Costa [60] where they subdivided an action into two stages: **anticipation** and **action**.

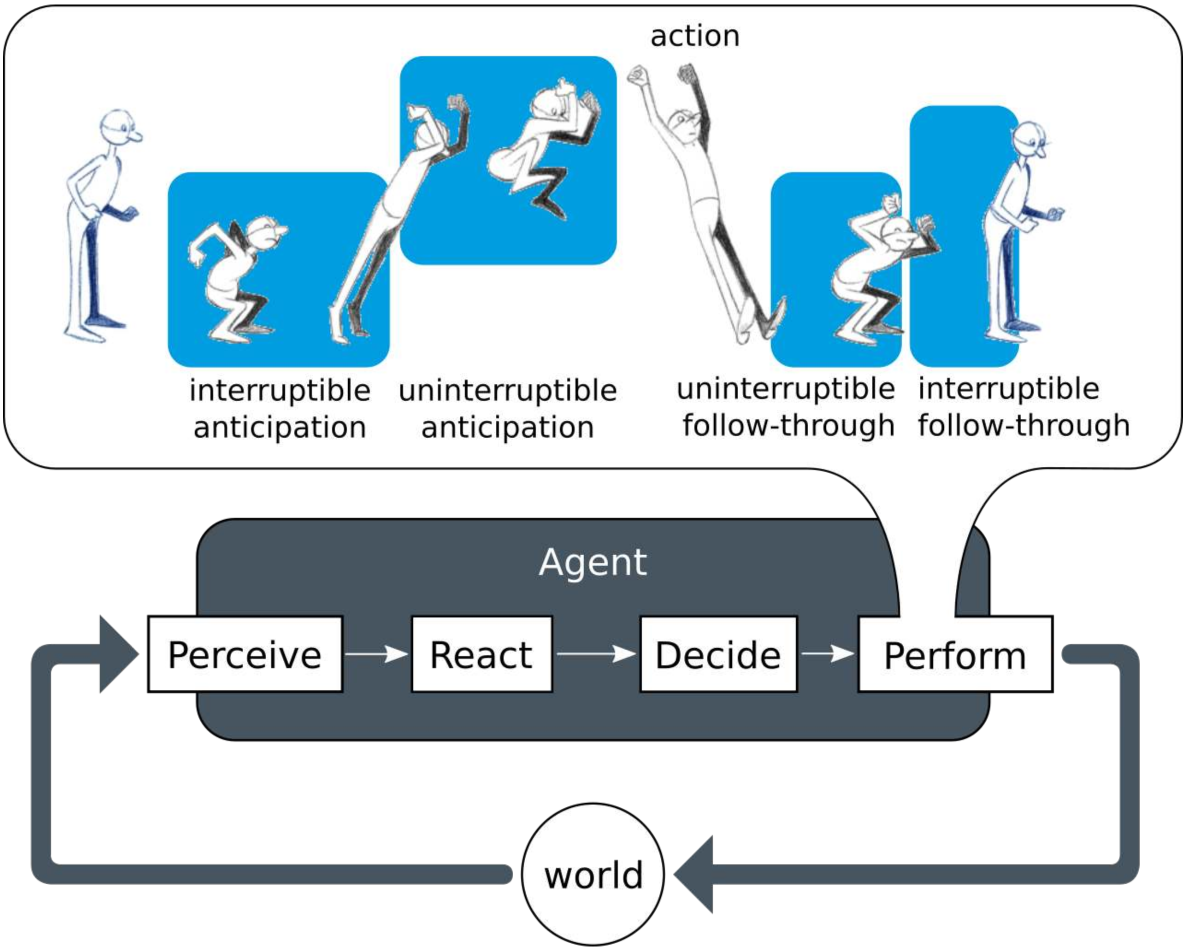


Figure 3.1: *3motion* conceptual approach depicting the four stages of an action and its connection to the agent's behavior. Note that there can be multiple iterations of the agent's behavior cycle happening while a single action is evolving. The 4 behavioral stages of an agent are also presented at the bottom of the figure. The agent perceives events from the world, that are then interpreted to create an emotional reaction and a decision on what action to perform, then the action starts or continues executing which impacts the world.

stage, if the character does not start a new action (e.g. swing with a weapon), the action follows its natural follow-through, yet if a new action begins, this action is completed (its animations are interrupted, but its purpose has been completed). The follow-through stage marks the end of the action, asserting whether the acting and observing characters' expectations are fulfilled. As such, characters express their emotions according to how their expectations were or were not fulfilled with respect to the expected result of the action.

Each of these action stages is associated with a different timing and duration, thus creating different moments for possible emotional responses. As a result, characters may have multiple reactions to a single action instead of reacting to an action as a whole.

3.2 The Endless Loop

To allow for the expression of an action and its stages, we modeled an agent's behavior in a 4-step *anticipation-based* cycle (see the bottom of Figure 3.1): (1) the agent *perceives* changes in the world based on its *expectations*, (2) *reacts* to them emotionally based on what was *anticipated*, (3) *decides* what to do next and *anticipates* what will happen, (4) and *performs part* of an on-going action (it is here that we apply the action subdivision), then repeats. This cycle occurs multiple times in the course of an action allowing the virtual agents to perceive events and express different emotions, among other expressions, while performing the same action.

3.2.1 Perceive

In the *perceive* step, the agent perceives events around them in the world. Special attention is given to two events, those from actions and emotions, that will help inform the next steps and give context to what is happening. Since characters will take some time to understand what others are doing and feeling, we explicitly model a delay in the perception of new events that represents the time required for the event to be understood by the virtual agent. The delay is used to emphasize the non-immediate response of agents, improving Believability, and its values are informed by research on vision, more specifically the Sustained System [143], and start at 200ms (humans take around 200 to 350ms to perceive new information, we can extend these values to improve readability and believability, see Appendix A.3). Furthermore, [144] analyzed the behavior of participants who were performing tasks on computers and found that when users are experiencing negative emotions both mouse movement and speed are relatively slowed down, suggesting that our emotions influence our reaction speed. The delay, which varies from one event to the next and may be modulated by the perceiving agent, allows for the expression of interesting behavior such as the expression of surprise or being actively waiting for something. This also translates well to the viewers' understanding of what is happening in the scene, since the viewers themselves will generally take some time to understand what is going on [143].

Note that the agents in our model can always perceive the correct emotion and action performed by another, if they do not perceive it correctly it is a deliberate design choice. This approach is akin to theater, where all cast members know what the others will perform and react accordingly. This is done to favor the believability of behavior and reinforce others' expressions.

3.2.2 React

The *react* step selects the emotion the agent is feeling as a result of the perceived events and its expectations, in the context of the actions being performed at the time by each of the virtual agents in the scene. This step is called throughout the execution of an action, allowing an agent to feel different emotions in the course of a single action, which allows the creation of a more believable behavior. The implementation of the appraisal of events and the generation of an emotion is based on Emotivector, the affective model by Martinho and Paiva [7] (see Section 2.5.2), where for each action they will expect a reward or a punishment; with this approach, the agent can create expectations regarding the action's success and react upon them.

Emotivector allows the computation of Sensations (the mismatch of perceived and predicted events), which can then be interpreted and will enable the creation of emotions. Our implementation can make use of several Emotivectors calculating predictions for continuous measures (e.g. perceived walking speed of opponent) and more discrete measures (e.g. a probability of defeat based on health and attacks received).

For continuous measures we assume that the “sensed value will tend to remain constant except for certain points in time”[7], thus we make use of the simple predictor proposed by Martinho and Paiva. Regarding discrete measures, we assume the data collection is more scarce and with high variance (similar to what is seen on the Virtual Coach, see Section 4.3.1), thus we make use of weighted moving averages and their first derivative. For both cases, a “searched” value might exist, depending on the context, in which case it will help generate more relevant sensations (e.g. expecting the opponent to be in a certain location to trigger a trap).

Having emotivectors, the management of multiple emotivectors is highly dependent on the use case. It is feasible to assume that one can use the solutions proposed by Martinho and Paiva [7], Winner-Takes-All (also used in the Virtual Coaches), Salience Ordering, and Meta Anticipation. We plan to make use of Meta Anticipation with additional rules that override the selection and allow the management to be more specific for the context (e.g. if the character is defensive, it would give more importance to emotivectors that refer to preserving its health and less to those referring to damaging the opponent).

Regarding the generation of emotions, based on the implementation of the Virtual Coach (see Section 4.3.1), we have an initial approach to the generation of emotions through a mismatch between the most salient Sensation and a previous mood. It is a simple solution that showed good results and is a good first implementation for this model.

3.2.3 Decide

In this step, an action is chosen to be executed. If the character is controlled by a human player, this step simply transforms the player’s input into an action that can be later manipulated (if need be) to better express an emotion (e.g. having a character walk slowly if they are scared). If otherwise the character is being controlled by an AI, then a decision-making algorithm decides what action it wants to perform (in our first use case we implemented a simple rule-based algorithm to select actions, see Section 3.5).

3.2.4 Perform

The *perform* step allows the execution of an action and the expression of an emotion. It is at this step that the action subdivision is implemented and where emotion expression will be applied. This step will manage the action being portrayed and keep track of what stage it is in, meanwhile it will blend the action with the emotion expression. Given the high dependence on the setting of a use case, performing an action needs to be reviewed case by case and its implementation needs to take into account the perception of the user (similar to what is discussed in Section 3.2.1, but where the focus is on the expression itself).

3.3 Behavior Illustration

Going back to the illustrative scenario first introduced in the Hypothesis (see Section 1.3), here we will expand it and use it to showcase the action subdivision. The scenario, now depicted in Figure 3.2, is set in a fantasy-inspired setting where two characters, Rua and Gorm, fight in an arena until one falls. Gorm is holding a two-handed axe and Rua a long spear. The figure is followed by a text description of the same scenario with annotations of numbers that link back to specific moments in the figure.

> The battle starts and both fighters run toward each other. As they get closer Gorm decides to perform a very slow but highly destructive overhead attack by pulling his axe back, revealing his intention in the process **(1)**. Rua becomes aware of this and shows confidence because she believes the attack will miss **(2)**. She then decides to thrust her spear towards Gorm **(3)**. As a result, Gorm becomes scared and decides to cancel his attack to defend himself **(4)**. Rua's attack continues **(5)** and is blocked by Gorm, leading Gorm to feel relieved and Rua irritated **(6)**. Both fighters move away from each other to prepare for the next attack.
>
> The first one to attack again is Rua, now with a slower spear swing **(7)**. Gorm decides not to attack and focuses on dodging the attack. Unfortunately for Rua, she miscalculated the distance and the attack misses completely **(8)**. Gorm takes advantage of this opportunity and attacks with a much quicker axe swing leading Rua to feel scared for her life as she cannot recover quickly enough to defend himself from the attack **(9)**. Gorm's axe swing hits Rua defeating her in the process. Now Gorm feels happy since he has become the winner of this battle **(10)**.

In this scenario not only do two characters fight (common in video games) but also any wrong move can end the fight tragically (more common in movies), this creates tension and intense emotions during the combat, a good example where the use of *3motion* helps improve believability. The use of *3motion* ensures that the characters express emotions at the correct timing, which helps create more believable interactions. Another scenario depicting the use of *3motion* is shown in Appendix C.1.

3.4 Mental State

For the agent to be able to react to others' actions more effectively, each agent must have *theory of mind* and/or *empathy* (see Section 2.1.4). In our approach, each agent has some theory of mind and keeps mental states associated with itself and with other agents: each agent stores what actions are being performed by which agent at a certain time and what stage they currently are in, as well as the emotions felt and perceived at each stage of the action. This not only allows the agent to be able to predict certain behavior and express emotions accordingly but also allows the creation of rich and coherent behavior based on the dynamics of what was perceived by each agent.

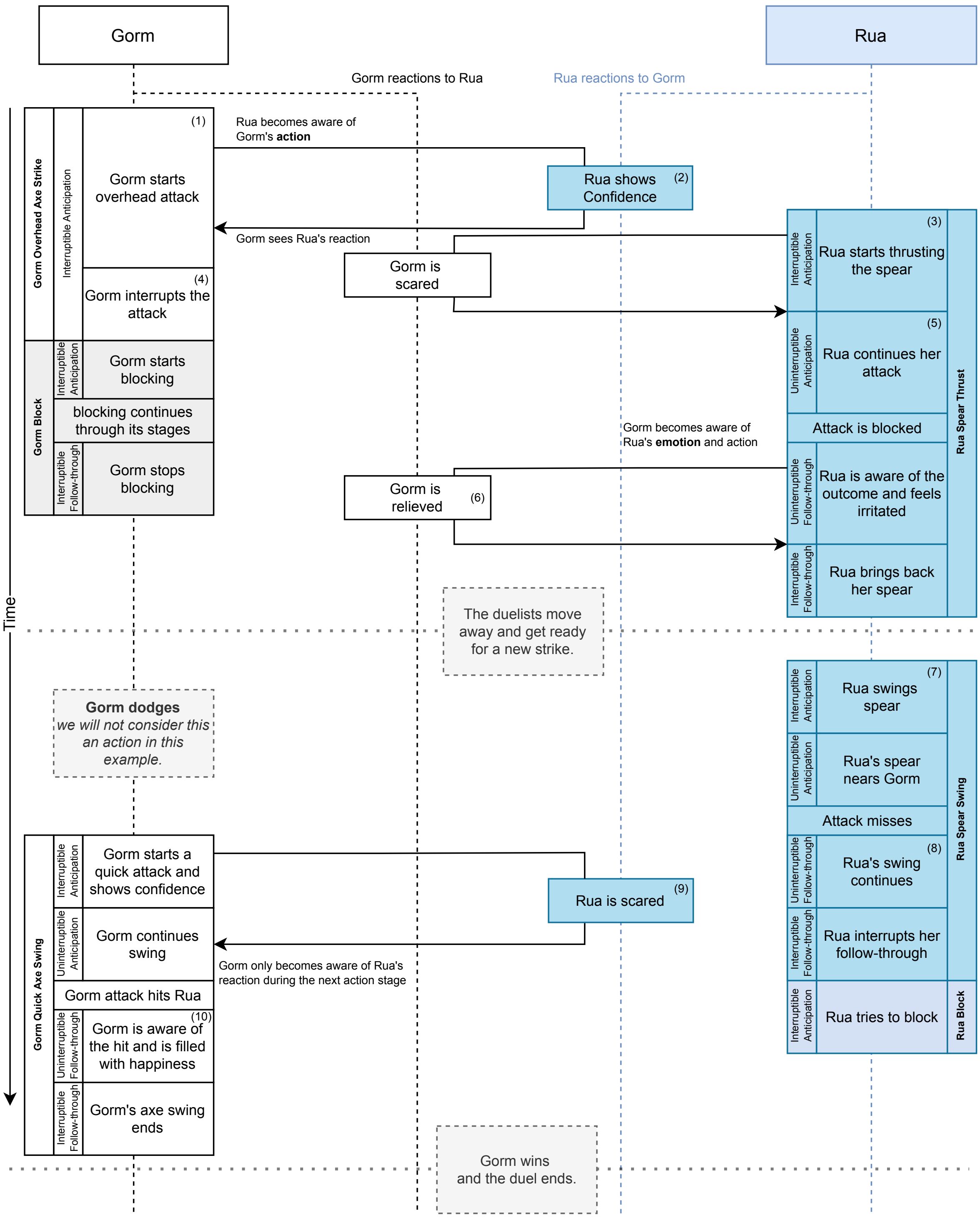


Figure 3.2: Action Flow of the Gorm dueling with Rua and their reactions. The figure depicts the subdivision of the action into several stages that allow for emotional non-verbal communication. Furthermore, it showcases the emotional response of the opponent during one's actions. Due to limited space and its similarity to the block action, the dodge action is not expanded in this example. Some blocks are annotated with numbers (e.g. "Rua shows Confidence" is annotated with the number *2*) to link them with the text description of the scenario.

3.5 Use Case: Text-based Application

As an early attempt of understanding the expressive power of our approach [16], a real-time text-based representation of a scene was developed with interacting agents and performed a user study with different variations of our model. Progress bars were used to represent playing an animation of an action and/or an emotion (see Figure 3.3). A small text next to each progress bar would describe the associated animation. Several progress bars would be displayed at the same time to mimic the concurrent execution of several actions in the scene.

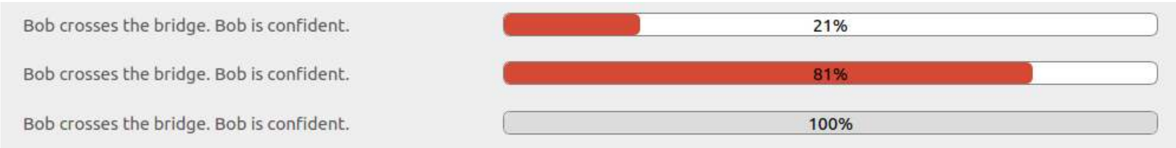


Figure 3.3: Example progression of an animation. Three moments of the execution are presented: just starting (top), almost finished (middle), and finished (bottom). The progress bar represents the time required to fully play the animation. The text depicts the action and associated emotion the agent is performing in the animation.

3.5.1 User Study

To measure the impact of the model regarding believability when compared to a more classic approach to behavior modeling, a user study was conducted in which participants were asked to fill out a questionnaire in which three videos were presented in a semi-random order[3], each containing the same scene of two agents about to cross a rope bridge using our real-time text-based representation (an illustration of the execution of this scenario can be seen in Appendix C.1). Each video shows a different algorithm for controlling the agents' behavior: *(a) Model*, the agents use the *3motion* model with action subdivision and correct timing of events, expressing coherent emotions; *(b) Classic*, the agents use the model without action subdivision and only signal the beginning and end of an action (emulating a virtual agent classic approach to behavior implementation); *(c) Misguide*, the agents use the model with action subdivision, but with incorrect timing (this is to determine if having more information shown, even if poorly timed, bias the perception of the scene as being more believable). After each video, the participants expressed their agreement with a set of statements about their perception of the agents and how the agents perceived other agents. Note that in this evaluation, the implementation did not take into account the subdivision of the follow-through, instead considering it an atomic state.

3.5.1.A Resources

A – Demographic Questionnaire: This questionnaire gathered generic demographic data, binary gender (Male and Female) and age, and specific data regarding interactions with Non-Player Characters (NPCs) (any agent or character in a virtual interactive experience) and animated environments (such as video games and animated movies). Regarding animated environments, information was gathered on how frequently one played video games and watched animated movies. Regarding NPCs, information was gathered on how preferred and important a NPC is in a video game.

[3]The ordering of the videos is discussed below and was designed to avoid bias between videos while limiting the number of questionnaires needed.

B – Believability Questionnaire: First introduced by us in [17], the questionnaire we used to assess believability is structured in two sets of statements: the *participants perception of the agents* and the *participants perception of how the agents perceive other agents*. It is important to note that in the questionnaire agents are referred to as characters, as such, we can consider them synonyms in this context. The participants were asked to rate their agreement to the following statements using a 5-point Likert scale from *1 - Strongly Disagree* to *5 - Strongly Agree*.

- From your (the participant) point of view:
  - Q1 - I understood what the characters were doing.
  - Q2 - I could predict the characters' actions.
  - Q3 - I understood what the characters were feeling.
  - Q4 - I could predict the characters' feelings.
  - Q5 - I understood the characters' intentions.
- From the virtual agents' point of view:
  - Q6 - The characters were aware of each other.
  - Q7 - The characters were aware of each other's actions.
  - Q8 - The characters could predict each others' actions.
  - Q9 - The characters were aware of each other's feelings.
  - Q10 - The characters could predict each others' feelings.
  - Q11 - The characters were aware of each other's intentions.
- Q12 - The interaction between characters in this scene was believable.

The dimensions in both sections are based on the works from [37] and [60]: awareness, behavior understandability, predictability, behavior coherency, change with experience, and social metrics. A dimension measuring the ability of an agent to perceive and interact with other agents was added. While the previous works focused on the participant's point of view, our approach also explores the participant's understanding of the virtual agents' points of view, which is one contributing factor to the creation of believable interactions (linking back to topics such as Theory of Mind and Empathy, shown in Section 2.1.4). The final question (Q12) helps us collect how the participant consciously evaluates the scene's believability, and what scene is considered the most believable.

C – Videos: We developed three videos to represent the same scene depicting two agents, Bob and Hanna, crossing a bridge (see Figure 3.4). As previously stated, what differs in each video is how the actions are subdivided and the timing used.

(a) Model video[4]: The agents use the *3motion* model with action subdivision and correct timing of events, expressing coherent emotions. This video represented the correct usage of *3motion*, thus we expected it to be the one containing more believable interactions.

[4]To view the Model video follow this link: `https://youtu.be/_ZLP-wv2yUo`

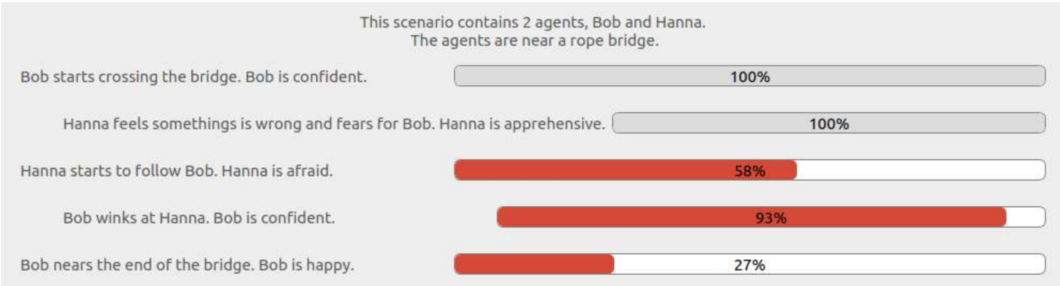


Figure 3.4: Screen capture of the *(a) Model* video depicting a moment where Hanna started following Bob while being afraid, and Bob is nearing the edge of the bridge being is happy and confident.

(b) Classic video[5]: The agents use the model without action subdivision and only signal the beginning and end of an action, emulating a virtual agent classic approach to behavior implementation that neglects anticipation. We expected this video to represent a believable scene, but to be hindered by the lack of emotion expression present.

(c) Misguide video[6]: The agents use the model with action subdivision, but with incorrect timing. This video's purpose is to determine if having more information shown (the *(b) Classic* video contains less information than the *(a) Model* video), even if poorly timed, creates more believable environments or not. We hypothesized that poor timing breaks the suspension of disbelief and thus we expected this video to be the least believable (see Section 2.1 on Believability).

To avoid bias regarding the order of the videos, three versions of the questionnaire were made, each presenting the videos in a different sequence, also assuring they do not repeat the same position in any questionnaire. The ordering follows a Latin-square distribution[7], where each row represents a version of the questionnaire and each column represents the position of a video in the order of each version. The videos were then distributed so that each video appears exactly once in each row and exactly once in each column. In table 3.1, we can see that *(a) Model*, for example, never repeats its position, favoring non-bias results.

Table 3.1: Questionnaire ordering. Rows represent a version of the questionnaire and the columns represent the order in which the videos are presented.

	1^{st}	2^{nd}	3^{rd}
Version A	a	b	c
Version B	c	a	b
Version C	b	c	a

The distribution of the questionnaires followed a linear pattern (i.e. A, B, C, A, B . . .) to have all versions have the same number of participants. Unfortunately, our implementation could not account for the case where a questionnaire was opened, but not filled, allowing inconsistencies in the number of replies in each version. Fortunately, the number of replies was almost identical in any version of the questionnaire.

[5]To view the Classic video follow this link: `https://youtu.be/0NBGq8cpQR0`
[6]To view the Misguide video follow this link: `https://youtu.be/_rQ-gHsRIGY`
[7]Latin-square distribution: `https://en.wikipedia.org/wiki/Latin_square`

3.5.1.B Procedure

Each session had a duration of 15 to 20 minutes. A participant would be provided a link to a web page that would redirect them to one of the versions of the questionnaire. They would fill out the *Demographic Questionnaire* first and then would be shown the three videos. After each video, they would fill out a *Believability Questionnaire* in the context of that video. When all videos were shown and respective questionnaires were filled out, the participant would submit the answers, closing their participation.

3.5.2 Results and Discussion

All the tests on the participants' data are presented in greater detail in [16]. Recall the three versions of the questionnaire, A, B, and C, with the videos in different orders (see Table 3.1), we will use these version naming during this section.

Testing involved 52 participants with ages ranging from 18 to 55 ($M = 25.10$, $SD = 6.06$, 14 female). 18 participants replied to version A, 18 participants replied to version B, and 16 replied to version C. Regarding their gender, 14 participants identified as female (38 remaining participants identified as male). Regarding their age, 37 participants were between 18 and 25 years old, 14 were between 26 and 40 years old, and only 1 participant was between 41 and 55 years old. Regarding the habits in interactions with NPCs and animated environments, a low percentage of participants (13.5%) do not usually play video games. Most of the participants (69.2%) have no preference for watching animated movies compared to other movies. This lack of preference also applies to their choice (53.8%) of games that show interactions with NPCs compared to other games, although many (30.8%) prefer games that show such interactions over other games. Lastly, interactions with NPCs were not considered more important than other features ($55,8\%$), for some participants (5.8%) this feature has no value.

For each video, we aggregated the responses given on the Believability questionnaire, which analysis we present henceforth. Using the Shapiro-Wilk test, we determined that none of the *data sets*, the answers given to the same statement in each video, were normally distributed and we are therefore only able to use non-parametric tests on our data. We then applied the Friedman test to each set of statements, determining whether the answers to the three videos were or were not significantly different. For fine-grain, we proceeded to use the Wilcoxon signed-rank test, which allowed us to compare how the answer to a statement in two different videos differed. An analysis of each statement is detailed in [16].

The analysis of the collected data led us to conclude that the *(a) Model* video (the video that correctly uses the model, paying attention to timing) ranked higher in *Q12* ($\chi^2(2) = 21.798, p = 0.000$) and in almost every statement (all except *Q4*), meaning that participants perceived this video to contain the most believable scene. The *(b) Classic* and *(c) Misguide* were generally indistinguishable from each other, yet, for those statements where differences arose (*Q1*-*Q4*), the *(b) Classic* was ranked higher.

The statements regarding the participant's perception of the agents (*Q1-Q5*) were where we expected to see more similarities between the *(a) Model* and the *(b) Classic* videos. The expectations were confirmed and statements, such as "Q2 - I could predict the characters' actions", were similar in value between these two videos ($Z = -0.469$, $p = 0.639$). A broken expectation was that of the statement "Q4 - I could predict the characters' feelings", where we hoped the new information given by the action subdivision would allow participants to more easily predict the agent's emotions ($Z = -1.009$, $p = 0.313$).

A weird phenomenon happen in the statement "Q5 - I understood the characters' intentions", where the three videos had no statically significant difference ($\chi^2(2) = 3.823$, $p = 0.148$). One can suppose that the intentions of the agents are easy to perceive in any of the videos or even that after watching the video the intentions were made clear. Another possible supposition is that the participants meant that they were capable of perceiving that the intentions had not changed.

The statements regarding the agent's perception of other agents and their actions, feelings, and intentions gave results that always favored the *(a) Model* video, indicating that the correct usage of the model improves believability.

3.6 Summary

In this chapter, we presented our solution to enable believable multi-party real-time interactions through the sub-division of actions and usage of non-verbal emotional communication. We first introduced the concept of action subdivision into three stages: anticipation stage, action stage, and follow-through stage. The anticipation stage can be further subdivided into interruptible and uninterruptible sub-stages, and the follow-through stage can be further subdivided into uninterruptible and interruptible sub-stages. This solution allows for a range of emotions to be communicated in the duration of a single action, which deepens the possible interactions and allows for a bigger emotional involvement of the agents. After that, we discussed how the action subdivision integrates with the agents' internal behavior loop, which is split into four steps: Perceive, React, Decide, and Perform. Delving into each step, we discussed the perception of the environment, the emotional responses, the decision-making processes, and finally the performance of actions and emotions using the action subdivision.

Following, an illustrative scenario is shown depicting how the model works and some of its novelties compared with models that do not consider action subdivision.

A text-based application was then detailed as one of the use cases of *3motion*. Text and progress bar elements were used to display unfolding scenes with agents interacting. A user study was performed where 52 participants were asked to evaluate the believability of three videos showcasing a scene of application, differentiated by their use of action subdivision and of different timings. Results suggest the video with action subdivision and correct timing, essentially depicting the correct usage of *3motion*, was rated as more believable in almost every statement compared to the other two videos.

Having discussed the model, we now understand how it works and where we can apply it. In the next Chapter 4, among other topics, we discuss an emotional model that implements the *Emotivector* model in the context of virtual coaches. In the following Chapter 5, we delve into a use case of this model, *Adfectus*, which combines *3motion* and the previously mentioned *Emotivector* emotional model.

4

A Case Study on Expression of Emotions

Contents

"Every word has consequences. Every silence, too."

Jean-Paul Sartre

This chapter discusses our work in creating a virtual coach used in Virtual Tutoring[1], a coaching mobile application to help university students in their studies. The virtual coach is relevant in the context of this work in its emotional model, which uses Emotivectors [7] to help compute the emotions expressed; in the expression of said emotions, that make use of multiple channels to communicate a single emotion (e.g. facial expressions, backgrounds); in the evaluations that point out common issues to be aware of when developing emotional synthetic characters; and in the proposal of the Triad Affect Interpretation (TAI) Method as a measure to combat these issues. We will give an overview of the architecture of the virtual coach and discuss in detail any area that is relevant to the present work. We also introduce the TAI method [22, 145], a method to better understand the user's perception of specific synthetic characters. More details on this topic were published in [19, 20]. Other works have been published that describe part of the implementation or use of these coaches [145–147].

4.1 An Introduction to Virtual Coaches in an Online Learning Environment

In recent years, there has been a widespread availability of online learning environments, offering more flexibility in a student's learning process, by allowing non-regular study schedules and access to readily available resources. This led to the development of new teaching and learning methods [148–152]. Although new methods arise, some problems remain. One such problem is the implicit existence of external factors influencing the students (e.g. responsibilities, lack of time), which lead to disinterest in the learning environment. To help mitigate these problems, we created a virtual coach that accompanies the students, adapts to their needs, helps them in their tasks, and understands (as well as responds to) their intentions, motivations, and feelings [153]. This virtual coach employs emotions along with empathy and interpersonal affect regulation strategies [46] to better motivate students, using both verbal and non-verbal communication to better express emotions.

4.2 Empathy and Interpersonal Affect Regulation in Social Interactions

An important aspect to consider when creating social virtual agents is *empathy*, an "affective response more appropriate to someone else's situation than to one's own" [51] (also discussed in Section 2.1.4). There are two main categories of arousal mechanisms known to be responsible for empathic affective responses. One category is the preverbal, automatic, and involuntary arousal mechanisms, including motor mimicry. The other is the higher-order cognitive modes, associated with conscious mental states, which include *perspective-taking* [154] defined as "putting oneself in another's shoes and imagining how he or she feels" [51], this being considered a very important component of empathy [155].

[1] `https://sites.uab.pt/tutoria-virtual/`

This leads to one other important aspect of social interactions: *interpersonal affect regulation*, defined by Niven *et al.*[46] as “deliberate attempts to regulate the feelings of others”. Based on a corpus of strategies for interpersonal affect regulation, Niven *et al.*[46] reached a classification that "primarily distinguished between strategies used to improve versus those used to worsen others’ affect, and between strategies that engaged the target in a situation or affective state versus relationship-oriented strategies" (see Figure 4.1). This research was further developed, and studies were conducted to see the impact of using these strategies, with the authors having confirmed [156] that “attempts to improve others’ affect will be associated positively, with the target reporting friendship and trust with the regulatory agent.”

	Strategies to improve affect	Strategies to worsen affect
Engagement strategies	Positive engagement	Negative engagement
	Affective engagement: Directly trying to improve the way the target feels about a situation, e.g., allowing the target to vent	Affective engagement: Directly trying to worsen the way the target feels about a situation, e.g., explaining how the target has hurt someone
	Problem-focused strategies, e.g., listening to the target’s problems	
	Target-focused strategies, e.g., pointing out the target’s positive characteristics	
	Cognitive engagement: Trying to change the way the target thinks about a situation in order to improve the target’s feelings, e.g., giving the target advice	Behavioral engagement: Trying to change the way the target behaves in relation to a situation in order to worsen the target’s feelings, e.g., complaining about the target’s behavior
Relationship-oriented strategies	Acceptance	Rejection
	Attention: Giving the target attention to communicate validation, e.g., making it clear that you care about the target	Rejecting the target’s feelings: Rejecting the target’s feelings to communicate snubbing, e.g., making it clear that you do not care how the target feels
	Valuing, e.g., making the target feel special	Confrontational strategies, e.g., being rude to the target
	Distraction, e.g., arranging an activity for the target	Nonconfrontational strategies, e.g., ignoring the target
	Humor: Being humorous towards the target to communicate validation, e.g., joking with the target	Putting one’s own feelings first: Putting one’s own feelings first to communicate snubbing, e.g., sulking around the target

Figure 4.1: Interpersonal Affect Regulation Strategies Classification from Niven *et al.*[46]. The strategies are categorized by either improving or worsening affect, and by being engagement or relationship-oriented strategies. For example, Distraction is categorized as a relationship-oriented strategy that improves affect.

Following these works, Reeck *et al.*[157] introduced the Social Regulatory Cycle (SRC), a psychological model created to better understand the processes for which one Regulator influences a Target, and the latter’s processes for dealing with said influence. In the work, four stages are defined for the Regulator: identify the target’s emotions; evaluate the need for regulation; select a strategy, and; implement the strategy. These stages are used in the application to guide the agent’s design and implementation.

4.3 Virtual Coach

In [19, 20], we created a virtual coach capable of taking into consideration the coachee’s goals and emotions, by both placing itself into the learner’s shoes, a key competence that all great coaches share [153], and motivating/engaging the user through the use of interpersonal affect regulation strategies (see Figure 4.2).

The behavior of a virtual coach is divided into four main steps (see Figure 4.3): *affective appraisal* where, by gathering user data and modeling expectations, the coach determines the affective state of the student; *strategy selection* where, using the computed affective state of all involved in the in-

Figure 4.2: Prototype of the "Virtual Tutoring" coaching application. On the left we see Maria, talking to the user under the attentive eye of her fellow coach, John.

teraction, an intervention strategy is selected to improve the user's affective state; *dialogue selection* where, having the strategy determined, a dialogue branch that best suits the strategy context is selected; and finally, *dialogue management*, where the dialogue is presented through the use of two and three-dimensional elements to create a more engaging experience. The steps are detailed in the following subsections.

4.3.1 Affective Appraisal

In this step, the virtual coach calculates an expectation for what is likely to happen within the context of the ongoing tasks. This generates expectations, to which the coach will later affectively react to, based on what is actually reported during the interactions with the coachee. Here, the coach infers what the coachee is feeling and, based on its individual personality, its own affective state as a response to that. This personality maps a mismatch between expectation and outcome, in the context of the current mood of the virtual coach, to a new emotion. As such, different coaches with different personalities will have different affective reactions to the same event.

Gathering Data

We gather both qualitative and quantitative information over time. This is done to get a better understanding of what the coachee is feeling in the current session/task context. *Qualitative* information refers to a user's personal opinion on a particular task. To get a better understanding of how we can adjust our training approach, three distinct qualitative metrics are collected by querying the user about how challenging, enjoyable (intrinsic motivation), and important (extrinsic motivation) a task is, at different points in time. *Quantitative* information is also collected and gives the coaches more objective data to understand the results of the approach over time. Again, three different quantitative metrics are recorded over time: *task performance*; *effort*, the amount of time spent on a task; and

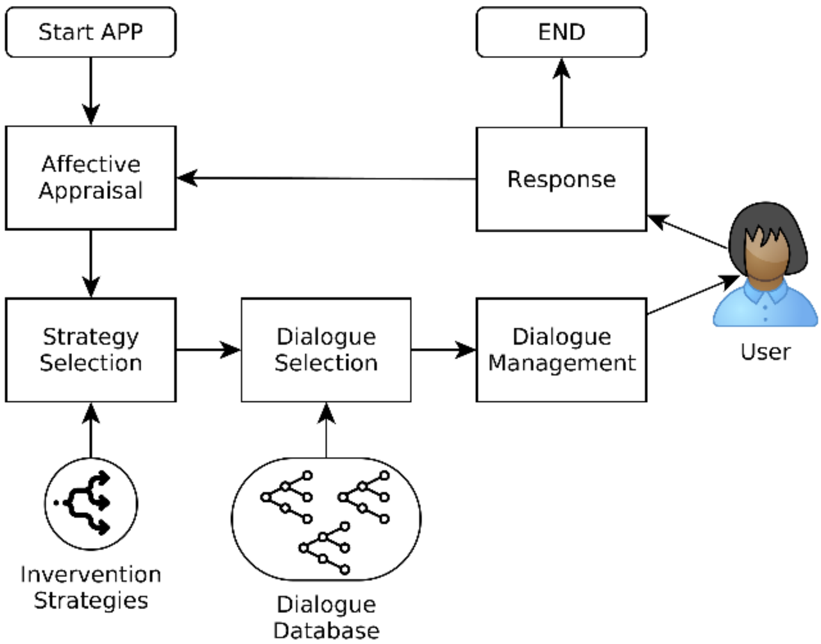


Figure 4.3: Overview of the Virtual Coach interaction loop: information is first collected and an empathic affective appraisal is made (*Affective Appraisal*), followed by the selection of an intervention strategy (*Strategy Selection*) that will dictate the dialogue that will be selected (*Dialogue Selection*). This dialogue is then interactively presented to the (*Dialogue Management*).

engagement, how frequently the coachee engages with the task. These generic task-independent metrics provide the context for understanding how the coachees are progressing over time on the multiple tasks they are involved with.

Computing Expectations

The collected data provides us with past information, such as knowing the coachee's initial performance or attitude towards a certain task; more recent information, relevant for short-term actuation; and estimates for the collected metrics based on all the collected data. Using an approach based on the *Emotivector* model [7] (see Section 2.5.2), an affective response that takes into account the previous expectations is generated. In other words, the coach calculates an estimate for all metrics of the tasks at hand and reacts to an event by comparing the mismatch between the expected metric and the actual measured metric.

In [7], they made use of metrics that were collected with high frequency and low variance, yet our application domain collects data more scarcely and with high variance. As such, we approached the analysis of the time series differently, relying on a weighted moving average of the value and its first derivative. Consider that, at time t, a metric value is $x_t \in [0, 1]$, we predict $\hat{x}_{t+1}$, its value at time $t + 1$, using the Equation 4.1 below[2].

[2] In our implementation, we used $n = 5$, and the following weights $w_0 = 0.5$, $w_1 = 0.2$, and $w_{(2,4)} = 0.1$, based on data collected from an empirical exploratory study.

$$
\hat{x}_{t+1} = \sum_{i=0}^{n-1}(x_{t-i} + \dot{x}_{t-i})w_i
$$
$$
\text{where } \sum_{i=0}^{n-1} w_i = 1, \text{ and } \dot{x}_t = (x_t - x_{t-1}) \tag{4.1}
$$

Handling Multiple Metrics

When several metrics are used simultaneously, they all compete for attention in terms of their relevance to the interaction. We implemented the concept of exogenous salience presented in [7] (ignoring endogenous salience) and used estimation error to reflect the principle that "the least expected is more likely to attract attention" [7]. In essence, we select the metric with the greatest salience computed as: $(x_t - \hat{x}_t)^2$.

From Sensation to Emotion

Directly mapping a sensation (metric mismatch) to an affective expression based on the personality of the virtual agent (e.g. mapping a greater reward than anticipated to the expression of happiness) did not produce good enough results in our exploratory study. As such, we extended the approach to take into account the current mood of the virtual coach, itself based on the previous emotion expressed by the virtual agent, to enrich the expressive palette.

We began by defining a set of possible emotions to express based on Ekman's research [64] (Happiness, Surprise, Anger, Disgust, Fear, and Sadness, with the addition of a Neutral expression), and developed an animation system supporting the expression of each emotion on a continuous scale of intensity (i.e. $0;1$). Then we defined the *mood* as three possible states (positive, neutral, or negative) based on the valence of the previous emotion expressed by the virtual agent: Happiness and Surprise were mapped to positive mood; Anger, Disgust, Fear, and Sadness were mapped to negative mood; and Neutral was mapped to neutral mood. Based on the intended personality of the virtual agent, we then paired the mood with a sensation and connect them to the expression of an emotion, e.g. being in a negative mood and receiving a punishment better than expected is appraised as surprise (see Table 4.1). This approach allowed the creation of distinct and consistent synthetic personalities (defined as the matching of mood and sensation to an emotion) while being flexible enough to create distinct behavior, such as optimistic or pessimistic perspectives.

4.3.2 Strategy Selection

Having both the user's perceived emotions as well as the coach's affective response, the virtual agent decides how to act by selecting the appropriate intervention strategy from a library that contains both task-oriented strategies and affect regulation strategies. *Task-oriented* strategies focus on the completion of a specific task, e.g. preparing for an important exam. Given their nature, these strategies are highly dependent on context, where some may depend on a particular event while others need to be presented in sequence. *Affect regulation* strategies, based on the work of Niven *et al.*[46],

Table 4.1: Example of a coach's personality that not only maps the sensation to an affective expression but also takes into account the current mood based on the valence of the previous emotion expressed. Emotions have an intensity ranging from 0 (zero) to 1 (one).

	Punishment		
Mood	Worse than Exp.	As Expected	Better than Exp.
Negative	Anger(0.5)	Sad(0.2)	Surprise(0.2)
Neutral	Sad(0.5)	Neutral	Happy(0.4)
Positive	Surprise(0.5)	Happy(0.2)	Happy(0.2)
	Reward		
Mood	Worse than Exp.	As Expected	Better than Exp.
Negative	Anger(0.2)	Neutral	Surprise(0.7)
Neutral	Sad(0.2)	Neutral	Happy(0.4)
Positive	Surprise(0.3)	Neutral	Happy(0.9)

focus on improving/worsening affect while also differentiating between engagement and relationship-oriented strategies (see Figure 4.1). With these strategies, we encourage the creation of an emotional bond with the user through interaction.

The selection of a specific strategy is made through a map between a measured metric (see section 4.3.1) and a sensation (e.g. reward worse than expected) to generate a strategy. This mapping was informed by a pedagogical team from the institution where our experiments took place. Further details are presented in [20].

4.3.3 Dialogue Selection and Management

After a strategy is selected, the dialogue best suited for tackling said strategy is chosen from a database of pre-built dialogue trees[3]. Each tree is tagged with information regarding its assigned strategy (e.g. engagement strategy), as well as the emotional states both the coaches and the users should be experiencing in the given context.

The coaches then interact with the user by traversing the dialogue tree, while expressing the emotions previously selected, which are manipulated by both dialogue and user interactions. Note that coaches are able to react to user responses, but also to a lack of response (e.g. if the user takes too long to respond), giving attention to both active and passive user behavior. This dynamic gives the coach the ability to express multiple emotions during a single action, as well as being able to react to certain situations that break the regular flow of dialogue, such as interruptions or lack of engagement from the coachee. We created two subsystems to handle the expression of emotions in this critical stage of the interaction with the user: the *Bubble System*, which controls the graphical dialogue elements; and the *Avatar System*, which controls the synthetic characters and their animations. Both systems focus on the transmission of the six primary families of emotions [64], with the varying degrees of intensity elicited during the affective appraisal (see Section 4.3.1).

[3] In our implementation, we created the dialogues for a small set of exercises with the students in our use case and reviewed by a pedagogical team. This led to the removal of certain words considered negative from the first drafts.

4.3.4 Bubble System

The Bubble System controls all 2D elements used to communicate emotion, where the main component is the balloons. The relevant aspects of the balloons are their shape, text, colors, and animations, which change accordingly to the emotion felt by the character. More details of each component are presented in Appendix B.1, and a full description of the Bubble System can be found in [146].

Balloon Shape – We created one distinct speech bubble for each emotion, plus a neutral state, based on the ones traditionally used in Comics (e.g. spiked balloons for anger) as well as using design guidelines to convey the appropriate valence for the associated emotion. To that end, we followed a user-centric iterative process until reaching a desired response to a speech balloon (e.g. curves for positive valences, and lines/obtuse angles for negative valences). For example, the balloon for anger is defined in the standard comic way as a spiked shape[4], while disgust ended up having a wave format in order to convey the feeling of nausea.

Balloon Color – We decided to follow the same approach as Pires *et al.*[158], and draw inspiration from the Disney™ movie "Inside Out"[193]. The colors chosen follow the same representations as those present in the movie, and for those emotions that were absent, a mix of other emotions/colors was used, to help communicate the appropriate valence (e.g. surprise was defined as orange by mixing the aggressiveness of red and the positive valence of yellow).

Balloon Animation – We decided to create simple balloon animations as scaling/motion over time, following the pose to pose principle (see Section 2.8). In total, we ended up with 14 distinct animations, a pair (one to show the balloon, and another one to hide it) for each of the six emotions, plus another pair for the neutral state. To distinguish between different intensities, the speed of the animation was changed accordingly (slower speeds for lower intensities, faster speeds for higher ones).

Text Animation – Regarding the text component of the speech balloons, we went with a parametric approach using animation curves (curves that define the text animation functions), since this allows us to have generalized effects and apply different curves to them, thus producing different animations (e.g. appear over time, jump, fade in/out, and shake).

4.3.5 Avatar System

The Avatar System manipulates the animated responses from the synthetic characters. The system is implemented as an animation controller with parallel layers of state machines. When presented with an emotion to express, the system updates the significant animation parameters and transitions to the required animations states. This approach of having each layer as a separate machine allows stacking animations together (i.e. blending), with the benefit of being able to use simple animations

[4]Representation of scream (anger) as seen in Spider-Man Annual #5 - A Day at the Daily Bugle: `https://imgur.com/a/1J012tw`

to achieve greater ranges of expressiveness. Our controller focuses on three main aspects: Expressions; Moods; and, Events. These can be blended together when relevant (e.g. to mix motion and expression). A full description of the Avatar System can be found in [147].

Expressions – The Expression Module handles brief expression requests, to be interwoven to create varied behavior. The main design goal of this subsystem is to support a variable range of emotional behavior, only having access to a limited number of assets. To achieve this, we implemented a system of blending trees for each emotion. This allowed us to tune the resulting intensity by adjusting the weight of the clips assigned as the blend tree's nodes.

Moods – The Mood Module coordinates the idle animations that communicate a character's emotional state. These dominate most of the facial regions and had to be attenuated when composed, otherwise, the output of the expression module would go unnoticed. Additionally, we also needed to animate the full range of back-channeling modalities, such as the various eye and head movements. This involved adding another set of overlapping layers with these animations, to play alongside the already existing idle animations.

Events – The Event Module comprises all animations not directly linked to the expression of emotion (i.e. attention-driven gaze, and speech-driven mouth movements). For these event animations, we were once again required to add a damping mechanism, so that they could play smoothly with those of the previous modules[5]. Additionally, to ensure that these motions adapted to the emotional state of the character, we added a slight variation to the speed and frequency those animations play at[6].

4.3.6 System Summary

The behavior model of the virtual coach focuses on four main steps: *affective appraisal*, *strategy selection*, *dialogue selection*, and *dialogue management* (see Figure 4.3). In these steps, we analyze and gather both qualitative and quantitative information, so that the coach can deduce the emotional state of the user, as well as its own synthetic affective response, which may differ based on the implemented personality. This is achieved by contrasting each tracked metric with its predicted value. Once the strategy (be it task-oriented or affect regulating) is decided, the appropriate dialogue tree is selected based not only on said strategy but also on the emotions and expectations of the coach. Lastly, the appropriate emotional responses are expressed by the coach through the use of 2D and 3D visual feedback while, at the same time, gathering information based on the responses given by the user during the interactive dialogue.

[5]Difference in talking animations with and without damping: `https://goo.gl/a8wb5F`
[6]Speed/frequency difference between the moods of happiness and sadness: `https://goo.gl/Yx9dHo`

4.4 Virtual Tutoring Use Case and Evaluation

Returning to the previously mentioned problem in the rise of disinterest of students in online learning environments, a preliminary evaluation was performed to better understand and appraise the approach described, where the virtual coach tutors in a real-world online distance learning environment scenario to better understand their impact.

A prototype was developed with two 3D animated virtual coaches (see Figure 4.2), that interacted with and assisted students, during a 10-day period, while preparing for the final exam of an online course. During these 10 days, a sample of 13 students (6 female) with ages ranging from 29 to 66 years old interacted with the virtual coaches on their mobile devices, directly providing them with both quantitative and qualitative data on how their study was progressing, and receiving comments and suggestions from them. Quantitative data was based on the amount of daily study time reported by the students, as well as the results from three formative assessments, while subjective data included perception of challenge, as well as endogenous and exogenous motivation. At the beginning and end of the evaluation, participants were asked to fill out a questionnaire on their beliefs regarding virtual agents. Additionally, a System Usability Scale (SUS) [159] questionnaire was given at the end of the evaluation to assess the general usability of the system and inform of eventual interaction problems that may have occurred.

Answers to the questionnaires show no changes were detected concerning the beliefs of the participants, in regards to the expression of emotion in virtual agents, during the conducted pre- and post-tests. An interesting fact, however, is that some students continued to interact with the application even after the trial period was over, possibly showing interest in continuing to use the features provided by the application. The median SUS score is 85 ($min = 57.5$, $Q_1 = 75$, $Q_3 = 95$, $max = 100.0$), which indicates an above-average score and near excellent rating [160]. These early results indicate a potential in this application, where users want to use it and seem to improve their studies, but improvements are needed in order to resolve inconsistencies and better the emotional and empathic impact of the tutors.

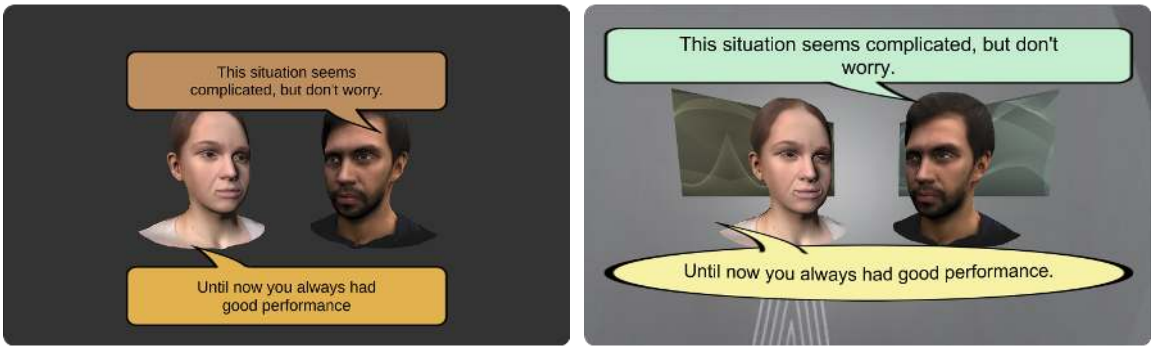
4.5 Evaluating Emotional Expression and Recognition

During development, concerns over the need for the clear expression of affective states led us to carefully consider the evaluation methods supporting our decisions. To that end, two versions of the application were built to perform comparative studies:

First version – This version features a visual expression mechanism using the characters and base animations provided by Didimo[7]. Dialogue-wise, the text is presented via speech balloons featuring a brown hue, a rectangular shape, and two short fade animations (one to show the balloon, and another to hide it). The environment around the characters was comprised solely of a gray static background (see Figure 4.4a).

[7]Didimo, Inc. `https://www.didimo.co/`

Second version – We created this version with the intent of creating more natural expressions and, at the same time, improving emotion expressions through 2D elements. It uses smooth transitions and varying intensities, allowing the characters to express emotions while talking and reacting to events. The speech balloons feature distinct colors, shapes, and animations for different emotions/intensities. The background environments also assist in conveying the mood of the characters, and in the portrayal of the conversation topic (see Figure 4.4b). Essentially, this version encapsulates the previously described components that interface with the user, the *Dialogue Management* and its subsystems (see Section 4.3.3).



(a) *First* version **(b)** *Second* version

Figure 4.4: The two versions of the coaching application depicting Maria (left), talking to the user under the attentive eye of her fellow coach, John. The rendition of text shown here was replaced in the experiments.

Having the two versions, we conducted two separate experiments. The *first experiment* was aimed at understanding whether the changes added by our modules had improved the expression of emotion in our system. We enlisted the help of two separate population samples by having them compare, side-by-side, the first version of the application with the second version of the application. One sample was comprised of *experts*, namely, Ph.D. students and professors from the fields of intelligent virtual and robotic agents, while the other contained only non-experts. The intent was to analyze whether the data collected from the two samples would present any significant disparities.

The *second experiment* focused on emotion recognition, given the importance of understanding whether the participants were capable of recognizing the emotions being communicated by the application. We gathered a population sample mixing both experts and non-experts, with similar demographics to the first experiment, and uniformly divided it into two distinct groups. One group would experience only the first version of the application, while the other would experience only the second version. This would allow us to investigate the positive/negative impact our changes had on the users' ability to recognize the conveyed emotions.

During these experiments, all the dialogue text was replaced with sentences composed of a single letter (e.g. *"OOO O OOOOO"*) to avoid biased information elicited from interpretations of the dialogues.

These experiments and their results were published in [21].

4.5.1 First Experiment: Emotion Expression Comparison

Thirteen videos, featuring distinct expressions of emotion from each of the two systems, were recorded and randomly presented, side by side, in an online questionnaire[8]. Participants were asked to evaluate the success of each animation in communicating the specified emotions, for both systems, on a scale from 1 to 7, with 1 being the lowest rank. A total of 10 responses for the non-expert sample were collected, with 5 belonging to female participants. The average age was 27 years old, ranging from 23 to 40 years of age. The expert sample had the same number of participants and a similar distribution of sex and age, averaging 29 years old, on a range spanning from 22 to 36 years of age. Altogether, we collected a total of 260 comparisons between the two versions of the application.

Table 4.2: Scores given by participants (experts, non-experts, total), for both intensities and their union, divided into quartiles for the First and Second versions.

Scores given by participants		**Low Intensity**			**High Intensity**			**Total**		
		Q1	Q2	Q3	Q1	Q2	Q3	Q1	Q2	Q3
First	Non-Experts	2	3	4	2	2.5	4	2	3	4
	Experts	1	2	3	1	2	3	1	2	3
	Total	2	2	3.25	2	2	4	2	2	4
Second	Non-Experts	3	5	6	4	6	7	3.75	5	6
	Experts	2	4	5	4	5	6	2.75	5	6
	Total	2	4.5	6	4	5	6	3	5	6

Looking at the median scores (Q2) from the non-expert sample, depicted in Table 4.2, we can observe that participants found the second version to be better at conveying the specified emotions. A Wilcoxon signed-rank test shows that this difference had statistical significance $(Z = -6.771, p < 0.0005)$. On a 1 to 7 scale, we have a median value of 3 for the first version (regardless of emotion intensity), while the second version has a median value of 5. If we separate between low and high intensities, the disparity in median values remains, with the gap between high intensities increasing by a small margin. These median scores were found to be of statistical relevance, for both low $(Z = -4.484, p < 0.0005)$ and high $(Z = -5.109, p < 0.0005)$ intensities.

For the expert sample, we could observe a wider gap, with a median value of 2 for the first version (when ignoring emotion intensity), while the second version features an identical median value of 5. A Wilcoxon signed-rank test shows that this difference also had statistical significance $(Z = -6.779, p < 0.0005)$. Taking into account emotional intensity, the tendency observed in the first condition remains, with the gap between the medians of high-intensity emotions showing a slight accentuation. These scores were also found to be statistically relevant, for both low $(Z = -4.032, p < 0.0005)$ and high $(Z = -5.276, p < 0.0005)$ intensities.

In summary, the second version was seen by experts and non-experts to be better at conveying emotions than the first version, independent of the intensity.

[8] A version of the questionnaire presented to the participants of the first experiment can be found at `https://goo.gl/PmUztE`

4.5.2 Second Experiment: Emotion Recognition

Twenty-six videos (13 for each version) featuring different emotional expressions (6 primary emotions x 2 intensities + neutral), were recorded and randomly presented in an online questionnaire[9]. For each video, participants were asked to identify which emotion they believed was being expressed, its intensity, and what aspects (characters/balloons/environment) influenced their decision. The questionnaire was based on the Geneva Emotion Wheel [161], with the participants selecting first the recognized valence, followed by the respective emotion (from a set of 20 possible). A total of 38 respondents participated in our survey, with a total of 170 videos watched for the second version and 156 for the first version.

Comparing the new data with the results from the previous experiment, it was possible to identify emotions that were perceived as better animated during the first experiment but suffered from poor emotional recognition in the second experiment. An example is the emotion of sadness, specifically, its low-intensity expression (that we will refer to as sadness-low) had poor accuracy in both versions, with the second version having 0% of correct answers. Surprise-low was also more accurately perceived in the first version, despite having previously obtained favorable results in the comparative experiment. Nevertheless, we also have emotions that coincide with the favorable results obtained in the comparative experiment, that point to them as being better animated, namely anger-low (see Figure 4.5), fear-high, and surprise-high. The remaining emotions did not show significant signs of being better or worse in terms of emotional recognition, even though they were pointed as being better expressed in the second version during the first, side-by-side, experiment.

If we crosscheck our results with the findings of Bassili [99] on emotional confusion (Table 4.3), we can observe that most recognition mistakes falling outside said findings were corrected in the second version, and the remaining cases exhibit only singular occurrences. The exception here is fear-low, which began being mistakenly recognized as anger.

Table 4.3: Overlap (blue background) of Bassili's findings and the emotional recognition mistakes obtained during testing. Each number represents the difference in mistakes between the First and Second versions, with positive values representing more mistakes in the Second version (e.g. high-intensity disgust is confused with anger in the first version, but is confused with fear in the second version of the application).

. . . could have been confused with . . .

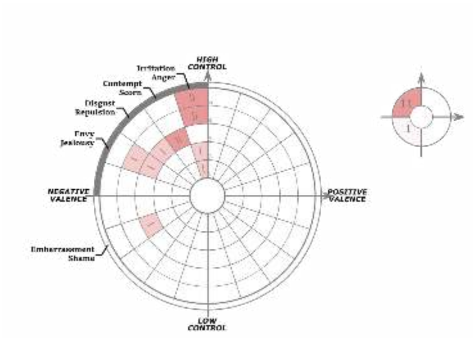
Low Intensity

	happiness	surprise	sadness	fear	disgust	anger
happiness		1		1		
surprise				1		1
sadness					-1	1
Fear	-2	3			-2	3
disgust		1		1		-3
anger						

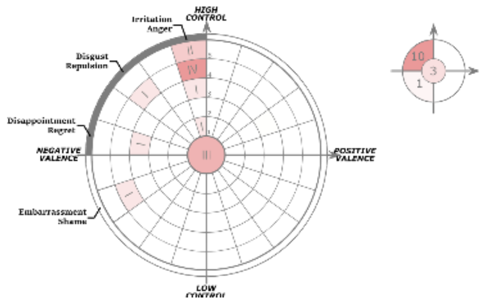
High Intensity

	happiness	surprise	sadness	fear	disgust	anger
happiness						
surprise						1
sadness	1			-1		
Fear		-2	2		-1	-1
disgust				2		-4
anger			1	2	-2	

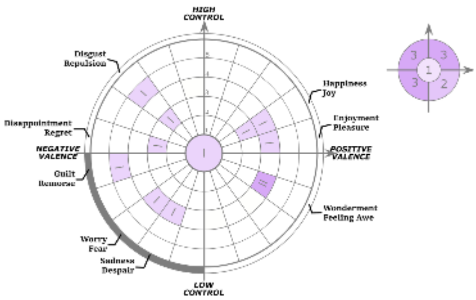
[9]A version of the questionnaire presented to the participants of the second experiment can be found at `https://goo.gl/8ZaZT6`



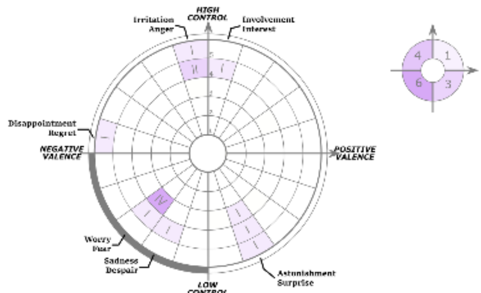
(a) Heatmap for the *First* version of Anger (low intensity)



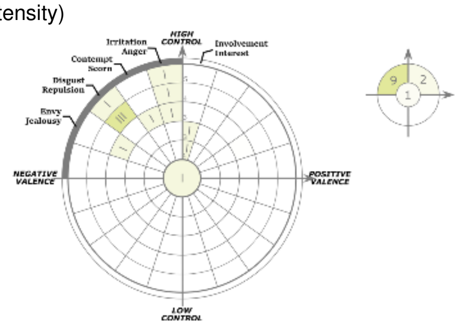
(b) Heatmap for the *Second* version of Anger (low intensity)



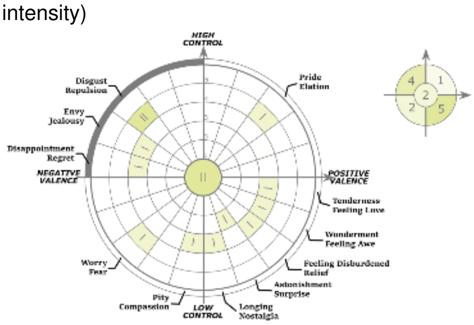
(c) Heatmap for the *First* version of Fear (low intensity)



(d) Heatmap for the *Second* version of Fear (low intensity)



(e) Heatmap for the *First* version of Disgust (low intensity)



(f) Heatmap for the *Second* version of Disgust (low intensity)

Figure 4.5: Heatmap for the emotions of anger, fear, and disgust (low intensity), for each application version. In the bigger wheel, the roman numerals indicate the number of times an emotion was recognized, and the Arabic numerals (one for each ring) indicate the intensity said emotion was recognized at. The middle circle indicates how many times users did not recognize any emotion. The smaller wheel indicates the total number of emotions recognized in each quadrant. The heatmaps for anger and fear exemplify emotions that were better recognized in the second version of the application (although fear is more often confused with other emotions). The heatmaps for disgust exemplify an emotion that was better recognized in the first version.

Regarding the dimensions of expression reported by the participants as influencing recognition, the animated characters were reported as being 1.45 times more influential than the balloons, and the balloons 3.55 times more influential than the environment. The more prominent reported features were facial expressions (29.02%), followed by the shape of the speech balloons (18.13%). It is important to note that the relative importance of each feature changes from one emotion to the next. For example, fear-high and anger-low were better recognized in the second version, but the character features were reported as more relevant for fear while the balloon features were pointed as more relevant for anger. This information is crucial to focus development on the features that are the most relevant for effective emotion expression.

4.5.3 Discussion

The two experiments offer conflicting results regarding emotion recognition. The first suggests that the second version of the prototype is better than the first at expressing any emotion. While the second study suggests that the second version isn't as good as it might have appeared.

A possible factor contributing to these differences is that side-by-side comparisons may lead to feature dominance, meaning that users can report preferring a specific version (e.g. for being more 'polished' than the other), despite being worse at communicating the specific emotion (e.g. for having contradictory cues, confusing users, etc.). In our prototypes, although both used the same models and animations, the second version had blending animations, animated backgrounds, richer text, and bubble animations, which may lead to feature dominance.

Furthermore, in the first experiment users were asked to rate the expressions for a given emotion (e.g. "How would you rate A_i and A_j, when comparing the two in the context of expressing a specific emotion E?") and in the second experiment users were asked to select which emotion best suits the animation (e.g. "What emotion did A_k expressed? E_1; E_2; E_3; ..."). Another possible factor may arise in which the first version leads to *forced-choice* – "an unintentional cheat sheet for the test subjects"[66] – which guides participants to an answer. In the first experiment, if the question mentioned *sadness*, the participant would only have to compare the animations to their concepts of sadness. In the second experiment, the selection pool was wider, thus the participant was not guided in their answer and had to try and recognize the expressed emotion, which resulted in less accurate emotion recognition. Although the experiments measured different metrics, we must be aware of forced-choice and its impact on the participants.

Because the experiments had a limited number of participants and because several factors were changed from the first to the second version of the system supporting the expression of emotion, it is impossible to pinpoint a specific change that could explain the obtained results. A more interesting question this experiment raises is *how much information can incremental evaluation provide during the development of virtual agents able to express emotions?*

4.6 Understanding the User Perception of the Emotional Expression

Given the conflicting results of the previous study, in an effort to better understand how participants perceive the characters, we designed the Triad Affect Interpretation (TAI) method [22, 145], which is based on the repertory grid [162], and applied it to our characters. The TAI method helps build a model of how users perceive different expressions of emotions of a specific synthetic character. The model aims at detecting problems and guiding the development of future animations, helping to improve the way emotions are communicated so that they are better recognized by the users.

4.6.1 The Repertory Grid

The repertory grid technique [162] is a method for eliciting personal constructs and is normally used to explore an interviewee's views on a particular topic with the absence of researcher bias. This technique was first developed for use in psychology but started being used in other areas as well.

There are two important aspects to take into account when talking about the repertory grid: the *elements* and the *constructs*. The latter can be defined as the *"rules"* we instinctively develop and by which we view or categorize situations, while an element is defined as *"the things or events which are abstracted by a construct"*[163]. Constructs can also be viewed as *"a way in which two or more things are alike and at the same time different from one or more things"*.

When using this technique, the first design decision is the selection of elements. Elements should be within the range of convenience of the constructs used. Constructs apply only to a limited number of people, events, or things, depending on the subject at hand. Each element is then written manually on a card and different triads (a set of three elements) are presented to the interviewee until all combinations have been covered, or the interview is terminated. Five or more elements are needed to produce a sufficient number of triads so that construct elicitation can be repeated.

For eliciting constructs the question *"in what way are two of these alike and at the same time different from the third?"* is asked of each interviewee when showing each triad. The question can be adapted depending on the study at hand, but it is always important to have in mind observer bias, in other words, it is important that the question does not guide the observer to a biased answer (e.g. asking "in what way are two of these better than the third?" guides an answer to be done under a value judgment).

A rating process is then performed to correlate each element of the study with each construct. For each element, each construct is normally presented using a five or seven-point scale (forming a matrix), and the participant is then tasked with rating where the element fits in the construct scale (e.g. if an element is a square and a construct is how straight or curved a line is, a participant would say that the square has straight lines). After the rating, this method outputs a model of how participants view each element based on the most commonly selected constructs.

4.6.2 The TAI Method

Our goal is to model how people perceive expressions of emotions in an animation so that we can improve it based on similarities with other expressions. To that end, we propose the Triad Affect Interpretation (TAI) method [22, 145], which consists in using the previously discussed triad analysis [162] in the context of emotion expression.

We first choose the *elements* for our repertory grid, in our case, animations of facial expressions, and select the best way to present them. Secondly, we elicit the *constructs*, personal rules by which one views or categorizes situations, in our case facial features, from each participant. Finally, participants are asked to rate each *element* against a set of selected *constructs*, allowing us to detect correlations between an animation and a construct. With this method, we expect to identify the most important *constructs*, or facial features, in each animation for each participant. With the ratings, we can better understand how the animations are perceived by all the participants.

Our study involved two different questionnaires (see Figure 4.6), one for determining the constructs through content analysis which is commonly used to analyze qualitative data in repertory grid [162, 164, 165] and the other for the rating of the animations (our elements) against each construct. The ratings are then analyzed using principal component analysis and cluster analysis, common types of analysis used in repertory grid [162, 164, 165]. Data gathering from both questionnaires was mainly aimed at national participants through convenience sampling.

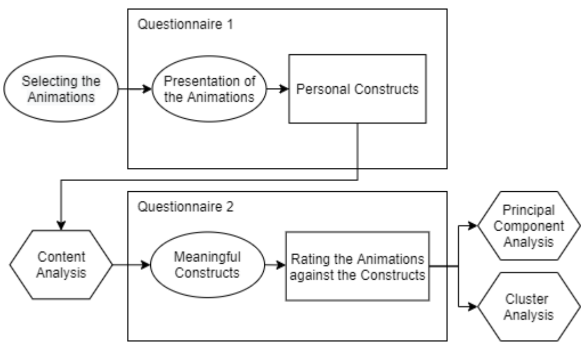


Figure 4.6: Overview of the TAI method's execution. The round shapes represent steps performed by the experimenter, while the rectangular shapes represent the contribution of the participants. The hexagonal shapes refer to the data analysis steps.

Figure 4.7: Snapshot of the six animations of João using the emotions' highest intensity. The emotions expressed are the following from top left to bottom right: *Anger*, *Disgust*, *Fear*, *Happiness*, *Sadness*, and *Surprise*.

4.6.2.A Selection of the Animations of Emotions

The first step is choosing the *elements* that will be presented, in our case, we want to compare different expressions of emotions in synthetic characters. We decided on six animations, each one representing one of Ekman's "basic" emotions [67]: *Anger*, *Disgust*, *Fear*, *Happiness*, *Sadness*, and

Surprise. We made use of the base animations of both synthetic characters as provided by Didimo portraying the emotion in its highest intensity. Figure 4.7 presents a snapshot taken of João portraying each of the six animations using the emotion's highest intensity: *Anger*[10], *Disgust*, *Fear*, *Happiness*, *Sadness*, and *Surprise*. We chose a limited number of expressions (one for each emotion), blends (always from neutral to an emotion), and movement (no head movement) to better understand how users perceived just the changes provided by the facial expressions, thus reducing the number of contributing factors in the perception of the expression.

4.6.2.B Presentation of the Animations

The second step is selecting how to present the elements to the participants. A solution is to show the participant three different animations, in random combinations, until all combinations had been covered, or no more constructs were elicited.

To balance the amount of work required by each participant, we decided to display only a fixed number of combinations. Since we have six animations, we would need a total of 20 combinations of 3 animations ($C_3^6 = 20$). To maximize the number of different combinations seen by the participants, we decided to display six combinations per participant to allow for the elicitation of an adequate number of constructs while not tiring the participants, based on a small pilot study.

To create the distribution that would cover as much variation as possible per participant, we considered 10 different lists, each containing the 6 animations with different orders as seen in Table 4.4, where each letter corresponds to a certain animation of emotion, six in total: A, B, C, D, E, and F. If we divide each list (a row in the table) in two we have the 20 combinations of 3 animations, each list containing a combination pair (e.g. in Table 4.4, the combination pair number 3 holds *ABF* and *DEC*).

Table 4.4: All 20 combinations of 3 animations distributed between 10 pairs.

Combination Pairs	Combination 1	Combination 2
0	A B C	D E F
1	A B D	C E F
2	A B E	D C F
3	A B F	D E C
4	A D C	B E F
5	A E C	D B F
6	A F C	D E B
7	A D E	B C F
8	A D F	B E C
9	A E F	D B C

The next step was to distribute them evenly among the participants so that we can get the maximum amount of combinations. Since we wanted to show 6 combinations per person, we take 3 combination pairs of the 10 available (e.g. pairs 0, 1, and 2 from Table 4.4). For each participant, we provide a different set of combination pairs. More specifically, if we consider n to be the number of the participant, the chosen pairs will be $n_{mod}10$, $n_{mod}10 + 1$ and $n_{mod}10 + 2$ (e.g. if $n = 13$, then the pairs chosen would be 3, 4, and 5).

[10]A video portraying João expressing *Anger* can be seen here `https://drive.google.com/file/d/1gx_dtNPgTG9KJ46Butjifp8Dzfxiri5q/view?usp=sharing`

Furthermore, we also decided to represent both synthetic characters, João and Maria, so each of the six combinations would have animations from the different avatars. The first combination would be from João and the second from Maria, then João, and so on. This allows us to check if a different avatar would have an impact on the perception of the animations, although they both used the same animations. Yet, during the evaluations, no significant differences were found.

4.6.2.C Personal Constructs

In this step, we collect relevant *constructs*, in our case facial features, through a questionnaire (Questionnaire 1 in Figure 4.6), and then sort them into themes by using content analysis. The questionnaire[11] consists of six sections, each presenting a combination of three animations. Three of the sections present João's animations and the other half presents Maria's animations. Each section had four questions:

Q1: *"By comparing the three animations presented above, identify the two that are alike."*

Q2: *"How are two of them alike and at the same time different from the third."*

Q3: *"Provide us with one characteristic that you found was alike in the two animations."*

Q4: *"Provide us with the opposite characteristic from the one mentioned above, describing the different animation."*

We introduce these sections by giving an example of how a section can be filled out. This example is given in another context with no connection to emotions so that the participant would not be biased. It consisted in showing three different shapes – a circle, a square, and a triangle – and exemplified a possible answer.

Based on issues detected during the pilot, some notes were added to help the creation of meaningful constructs, namely: *"do not repeat characteristics between sections"* and *"do not use names of emotions to justify the choice of Q1"*. Both **Q3** and **Q4** helped better understand the characteristics that the participants had in mind.

After gathering all responses, we used content analysis to find meaningful constructs by sorting them into themes and analyzing the frequency with which they appeared. *Content analysis* is a subjective qualitative analysis by which elements and/or construct labels are placed into common categories or main issues and interpreted for meaning [165].

4.6.2.D Rating Animations against the Constructs

After selecting the most elicited constructs from the first questionnaire, the final step was to ask the participants to rate the animations with the chosen constructs using a second questionnaire[12]

[11] A version of Questionnaire 1 presented to the participants can be found at `http://web.tecnico.ulisboa.pt/ist186514/Experiments/SC1-anonymized.html`

[12] A version of Questionnaire 2 presented to the participants can be found at `https://forms.gle/k1XgM4HdcH6Pttom6`

(Questionnaire 2 in Figure 4.6). Each animation was shown in its individual section and the participants were asked to rate the animation through a 7-point bipolar scale between the two opposing words defining each of the selected constructs (e.g. “mouth opens wide” versus “mouth closes”).

To avoid any order bias, we used a *Latin square* design to order the animations, creating an $n \times n$ array filled with n different symbols, each associated with one emotion and each occurring exactly once in each row and exactly once in each column. As such, our 6 animations were presented in different orders using 6 different versions of the questionnaire in a way that no animation occurs in the same spot in more than one version, reducing the potential bias mentioned before.

After gathering all responses, we create a model of how participants perceive the characters by performing a Principal Component Analysis (PCA). Afterward, using the created model, we ran a Cluster Analysis to identify prototypes of the more distinct emotions.

Both questionnaires collected demographic data, such as age, gender, maternal language, and interaction habits with synthetic characters.

4.6.3 Results

To create a model of how participants perceive our characters, we ran a first evaluation using the TAI Method. We present our results by dividing the presentation between the *elicitation* of the constructs and their *rating*.

4.6.3.A Elicitation of the Constructs

In this first stage, we used Questionnaire 1 (see Section 4.6.2.C) to gather constructs that the participants and the surrounding community understand. Our sample ($N = 21$) was aged between 21 and 56 years ($M = 26$, $SD = 9.540$). 48% were female and 52% were male. Regarding their experience with synthetic characters, 85.7% had interacted with NPCs in video games and 90.5% had interacted with virtual agents in other contexts. 28.6% reported having interacted with social robots in the past and 1 had worked with/developed synthetic characters and/or virtual agents.

Each participant identified 6 different constructs for a total of 126. We discarded 25 constructs due to the following reasons:

- the participant used the same construct more than once (4 constructs had this issue);
- the participant used emotions to describe the different constructs (14 constructs had this issue);
- the construct was too broad or unspecific, e.g. “eyebrows” (7 constructs had this issue).

We then took the remaining 101 constructs and organized them around common themes (e.g. the constructs of “eyebrows go up/down” and “eyebrows move/do not move” were merged into a single construct “eyebrows go up/down”). As a result of the content analysis, we identified 32 distinct constructs. We additionally concluded that a frequency below 3 occurrences was not meaningful enough to be considered. Given that, a total of 9 final meaningful constructs were selected. They are listed in Table 4.5.

Table 4.5: Selected meaningful constructs, defined by associated word pairs and sorted by frequency.

Construct		
Pole 1	**Pole 2**	**Freq.**
mouth opens wide	mouth barely opens/closes	15
eyebrows go down	eyebrows go up	12
eyes open	eyes close	9
forehead goes down	forehead goes up	7
teeth visible	teeth not visible	7
face contracts	face expands	6
cheeks go down	cheeks go up	4
round shaped mouth	curved shaped mouth/smile	4
teeth close together	teeth wide apart	3

4.6.3.B Rating the Expression of Emotions

Making use of the collected constructs, we can now rate the animations expressing emotions through Questionnaire 2 (see Section 4.6.2.D). In this stage, our sample ($N = 40$) was aged between 15 and 56 years old ($M = 25$, $SD = 7.808$). 60% were female and 40% were male. 62.5% had not answered the first questionnaire. Regarding the experience with synthetic characters, 80% had interacted with NPCs in video games, and 77.5% had interacted with virtual agents in other contexts. 27.5% reported having previously interacted with social robots and 2 had worked with or developed synthetic characters and/or virtual agents.

Using the participants' ratings, we built our model for *the perception of animations of emotions* using both PCA and cluster analysis.

A – Principal Component Analysis To group correlated constructs, we performed a Principal Component Analysis (PCA), a variable-reduction technique that aims to reduce a larger set of variables into a smaller set. We ran multiple dimension reduction analyses with different numbers of factors to find what the optimal number was. More than 5 factors did not account for more gain in variance, so we stopped considering those further. Considering only 2 factors only yielded 57% of variance, so it was also discarded. We decided on using 3 factors that accounted for 69.9% of the variance. Using this number of factors we created the rotated component matrix depicted in Table 4.6. Note that two of the constructs are cross-loaded on more than one component. Items with a loading less than 0.4 were removed as recommended and standard process [166].

From this PCA, we created our *model for the perception of animations of emotions*. Note that these results are specific to the 3D model and animations in our use case and may or may not generalize to other synthetic characters. Having each component subdividing the facial expression into its more relevant features for emotional expression:

Component C1: reveals that the forehead, eyebrows, eyes, mouth movement, teeth visibility, and face usually move in a similar direction; for example, the animations for *Surprise* and *Fear* would be at one extremity and the animations for *Anger* and *Disgust* would be at the other.

Component C2: tells us that the mouth area is highly correlated with the cheeks and how close the teeth are together; for example, if our synthetic character is smiling, the cheeks would go up and

Table 4.6: Rotated Component Matrix for 3 components C1, C2, and C3 with an absolute value above 0.4.

Construct		C1	C2	C3
forehead goes down	... goes up	0.810		
eyebrows go down	... go up	0.791		
eyes close	... open	0.601		
cheeks move down	... move up		0.815	
mouth remains closed	... opens widely	0.420		0.795
round shaped mouth	curved shaped mouth/smile		0.863	
teeth not visible	... totally visible			0.907
teeth close together	... wide apart	0.526	-0.442	
face contracts	... expands	0.846		

the teeth would be closer. Looking at the animation of *Happiness*, we can verify these points.

Component C3: reveals that the more the mouth is opened the more the teeth are visible, which is not surprising.

Figure 4.8 shows the distribution of cases labeled with the emotion they express, each axis representing a different component in 3D space. Each case is represented using its factor score. Since a factor is, by nature, unobserved, we used the regression method to generate a plausible factor score for each case.

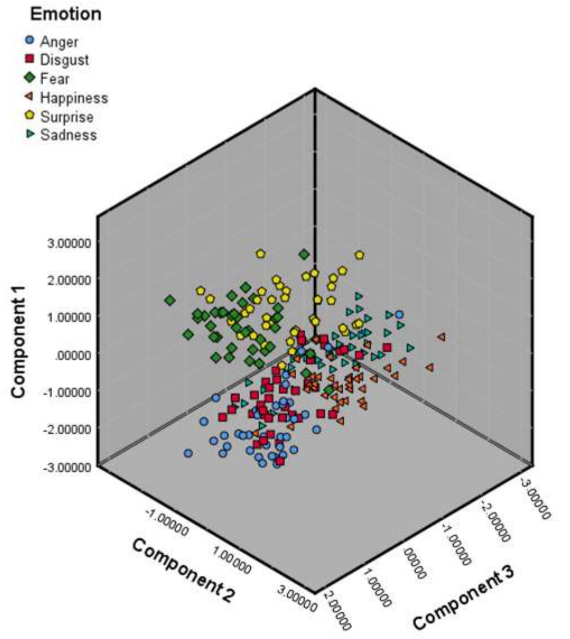


Figure 4.8: Distribution of the cases labeled with the emotion it represents, using the calculated factor scores for each case.

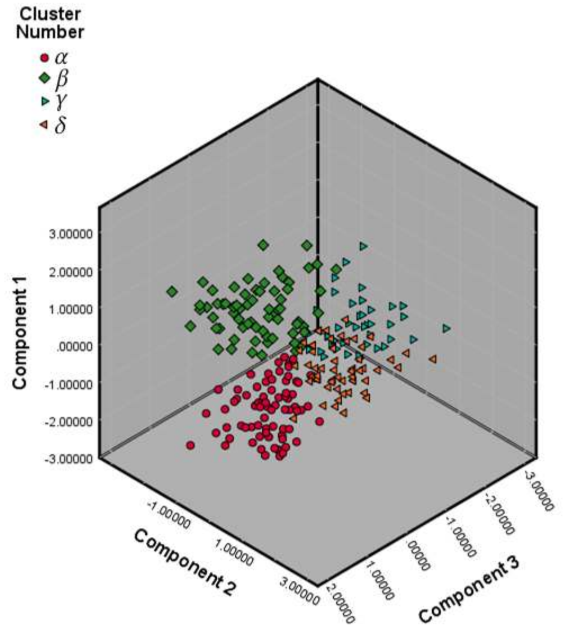


Figure 4.9: Distribution of the cases labeled with the cluster it belongs, using the calculated factor scores for each case.

B – Cluster Analysis Some emotions are easily confused with others [99]. Using our model created from the PCA, we grouped emotions together based on similarities, thus allowing us to identify prototypes of the more distinct emotions. Cluster Analysis tries to divide a certain number of observations into groups with common characteristics. With a total of 240 cases, i.e. the 6 different animations of emotions multiplied by the number of participants (40), we verified if the different emotions would be grouped in the same cluster.

We used k-means clustering with the squared Euclidean distance method to minimize within-cluster variances. We considered the factor scores from the 3 components model for the variables. Since k-means needs a value for k beforehand, we had to understand the optimal number of clusters for our data. We concluded that 4 clusters were ideal for our study by the use and comparison of the *elbow and silhouette methods*. Table 4.7 shows the number of cases of each emotion in each cluster when using the factor scores for each of the 3 components as variables and Figure 4.9 shows the distribution of cases for each cluster.

Table 4.7: Number of cases of each emotion in each cluster. The cases were compared with the squared Euclidean distance between their respective factor scores.

Cluster	Anger	Disgust	Fear	Surprise	Sadness	Happiness	# of cases
α	36	27	0	1	9	1	74
β	0	2	36	29	1	0	68
γ	2	3	0	9	27	1	42
δ	2	8	4	1	3	38	56
Total				**40**			**240**

Table 4.8: Cluster centers for each component in the model for the perception of animations of emotions.

	Cluster			
Component	α	β	γ	δ
1	-1.0	1.0	-0.3	0.4
2	-0.1	-0.8	-0.3	1.3
3	0.7	0.3	-1.7	0.0

We can summarize each cluster as follows:

Cluster α is represented by *Anger* and *Disgust* and the main feature to consider is the contraction of the face, in the upper area involving the eyes and forehead (Component 1).

Cluster β is represented by *Fear* and *Surprise* and the main features to consider are the face expanding and the mouth more rounded with the teeth wide apart (Component 1 and 2).

Cluster γ is represented by *Sadness* and the main feature to consider is the mouth more closed with almost no teeth visible (Component 3).

Cluster δ is represented by *Happiness* and the main feature to consider is the smile (Component 2).

4.6.4 Discussion

Wegrzyn *et al.*[167] mapped how individual face parts contribute to successful emotion recognition. Their findings relate to an evaluation with photos of real actors portraying Ekman's "basic" emotions [67]. Part of their findings, relating to confusion between recognition of emotions, correlate with what was previously seen in other studies [21, 99, 168, 169] and reinforced in the current work, namely that *Fear* and *Surprise*, and *Anger* and *Disgust* tend to be confused with each other. But, similarly to this work, they explore what features are more relevant to the correct recognition of emotion. They found that "observers were mostly relying on the eye and mouth regions when successfully recognizing an emotion"[167], furthermore, the expression of *Sadness* and *Fear* has a high reliance on the eyes, while *Disgust* and *Happiness* focus on the mouth. An analysis of the upper and lower face halves also shows that "sad, fearful and angry faces are most often recognized by upper facial features, while happiness and disgust are recognized by information in the lower face half".

We can compare the results in [167] with our own. In both studies, *Happiness* and *Disgust* were identified similarly by a focus on the mouth and its shape, the same applies to *Anger* with a focus on the upper face. *Fear* and *Sadness* differed, being identified by the mouth area in the current work and the eyes or the upper face area in [167]. This small comparison may suggest that improvements are needed in the animations that differed (*Fear* and *Sadness*) if we wish to have them be more human-like. Not only that, but it also informs us that to improve we should focus on the upper area of the face.

A feature that is present only in this work is the contraction/expansion of the face, characterizing anger, disgust, and fear. It is not clear how this feature correlates with the results in [167]. Perhaps due to the usage of only still images of the actors in the study in [167], the feature did not make itself present. Nevertheless, it is a feature that should be further explored in the future.

4.7 Summary

In this chapter, we presented an architecture that allows for the creation of intelligent virtual agents acting as coaches in distance learning environments. Based on (1) the objective and subjective metrics provided by the users, as well as the learning environment, (2) the affective state of the users, inferred from the evolution of these metrics over time, how they create expectations, and how these expectations are actually met, and, (3) the synthetic affective response of the virtual coach based on its synthetic personality, the system is able to select from a repertoire of dialogue trees gathered from a human pedagogical team, the one that better conveys the interpersonal regulation strategy appropriate for that situation. We then provided an overview of all components built in a first implementation of the approach and described a real-world case study featuring two virtual coaches, running on a mobile device during a 10-day evaluation period, whose results support the interest of students in the continued interaction with the empathic virtual coaches, in the context of online distance learning education.

To better understand the impact on emotion expression/recognition of our prototype and the changes made to make said expression more natural, we conducted two sets of experiments. The first one directly compared the first version to the second version (side-by-side) while expressing specific emotions, to both expert and non-expert participants. Results clearly supported that the second version was better rated at communicating each individual emotion. The follow-up experiment asked participants to recognize individual emotions expressed by each version. The results showed that emotion recognition of the second version had not improved in all cases, as was expected based on the first experiment. We noticed that, although the expression for some emotions was rated as being better in the first experiment, the second version had a lower recognition rate than the first version, as we could confirm in the second experiment. We also noticed that particular elements were reported as influencing emotion recognition in the different emotion expressions, pointing to separate needs for each family of emotions.

This discussion exemplifies why, when choosing the system that will better support the expression of emotions for intelligent virtual agents, in the sense that the portrayed emotions will be easier to recognize, asking the users to compare two different versions of the system, side by side, may not be the most adequate way to proceed. This study, therefore, mirrors the incremental formative evaluations that often underlie virtual agent development, and offers a cautionary tale in evaluating aspects of a system during its development and its evolving context of use.

In an effort to better understand the user's perception of our characters and their animations, we created the TAI method, which allows for the creation of a model of how people perceive different emotions in specific synthetic characters. The generated model provides insight into the different emotions, their similarities, and what features lead to their misinterpretation.

We started by selecting six animations from our characters, João and Maria, each representing a different emotion. Then, we collected meaningful constructs by asking participants to fill out a questionnaire containing triad comparison tasks, where three videos portraying animations of emotions were presented side-by-side, and asking participants to identify the two that are more alike and different from the third. Using the collected constructs, we designed a questionnaire where each animation was to be rated against each construct using a bipolar scale. Using PCA and cluster analysis, we then created our model for the perception of animations of emotions. PCA reduced the number of constructs to 3 correlated components: *(C1)* corresponds to how people viewed the face in general (if it was more contracted or expanded); *(C2)* relates to the movement of the mouth and its relation to the cheeks movement; and *(C3)* relates to the mouth opening and the visibility of the teeth. To further understand the similarities between the animations associated with the expression of emotions, we performed a cluster analysis that grouped the animations into 4 clusters: *Anger* and *Disgust* were grouped in Cluster α; *Fear* and *Surprise* in Cluster β; *Sadness* in Cluster γ; and *Happiness* in Cluster δ.

With the TAI method, it is possible to create a model of how people perceive different emotions that can provide insight into the different emotions that may be confused with each other due to specific features of the animation system. We can visualize which features are more prominent in each different cluster and thus how the different animations are more alike. *Anger* and *Disgust* are characterized mainly by a contraction of the face (Component 1); *Fear* and *Surprise* by the expansion of the face and the shape of the mouth being rounded (Component 1 and 2); *Happiness* by its smile, a curved shaped mouth (Component 2); and finally *Sadness* by the closed mouth with no teeth visible (Component 3). These features could be exaggerated or modulated to achieve the desired recognition level.

This chapter helps inform the development of specific empathic synthetic characters; we can extrapolate this knowledge to other contexts as it shares many common features of synthetic characters' development. The use of *Emotivectors* presented here will be used as the basis of the emotional model to be implemented within *3motion*, as it provides anticipatory capabilities to agents. Furthermore, the lessons learned from the comparative studies will help guide future evaluations of the main work, not only as cautionary tales but also as comparative data when considering emotion recognition tasks. Finally, the TAI method will also be used in the context of full-bodied animated characters, showcasing its utility with different synthetic characters.

5

Designing Believable Interactions in *Adfectus*

Contents

"Oh yes, the past can hurt. But from the way I see it,
you can either run from it or... learn from it."
Rafiki from The Lion King

In this chapter, we will discuss how we designed a use case for 3motion *(discussed in Chapter 3) that mimics common game development settings and assets. The use case is a video game named* Adfectus *developed using modern technologies and set in a 3D environment the player can interact with. First, we detail the game and its components, focusing on the different characters' emotional expression capabilities. We then detail how* 3motion *was implemented in the context of this use case. Finally, we detail multiple evaluations performed to assess the believability of the characters, interactions, and overall experience of the use case, and also an evaluation of how participants perceive the animations of the characters.*

5.1 Adfectus

Figure 5.1: Showcase of Adfectus' characters and arena. Two playable characters can be seen in the foreground holding weapons and ready to face each other on the ground floor of the arena.

Adfectus is an arena game in which two characters battle one another (see Figure 5.1). To win the game, a character must better his opponents. This game is being developed to be a testbed for *3motion*, showcasing the mechanics that are often used in action games in a 3D environment and, at the same time, offering new mechanics that can be improved further through the use of *3motion*, such as interruptible and uninterruptible stages of attacks [1].

Visually, the game uses a low-poly aesthetic[2], this allows us to make use of exaggerated expressions, such as scaling parts of the body, without compromising the believability of the characters (see Section 2.1). The setting is reminiscing of the medieval fantasy genre[3], a genre of fiction set in a fictional universe that specifically pertains to or is influenced by the medieval period of European history.

[1] A gameplay video of Adfectus can be seen at `https://youtu.be/hZmKEcBiEe4`.
[2] Low Poly Aesthetic - `https://en.wikipedia.org/wiki/Low_poly`
[3] Medieval Fantasy Genre - `https://aesthetics.fandom.com/wiki/Medieval_Fantasy`

The game was implemented using Unreal Engine 5[4], in which we used Blueprints (a tool for visual programming within the engine) and the C++ programming language.

In it, an arena is the main place portrayed, it is composed of a ground floor and an upper round balcony that surrounds the floor. On the floor are the *duelists* that will face each other to win the battle and are controlled by players. The players fight in a series of rounds, where the best player out of 5 rounds wins. Each round, *duelists* deal blows to each other until either runs out of health. Each player has a separate viewport, which is placed side-by-side as shown in Figure 5.4.

5.1.1 The Duelist - Player Character

The duelist is the main character of Adfectus and the main focus of our work. It is the character that battles another duelist and determines the winner or loser of the battle. To battle, the duelist will use melee-type weapons, a **sword and shield** or an **axe**, along with armor (see Figure 5.1). The character has an associated health, which is a value that ranges from an arbitrary positive value[5] to 0 (i.e. death), decreasing every time the character takes a hit. During the battle, the duelist can perform attacks with their weapon to damage the opponent. We implemented a **light attack**, which is quicker but does less damage, and a **heavy attack**, which is slower but does more damage on hit. It can also **block** incoming attacks with its weapon and shield, and **move** around the floor. Using a sword allows for quicker but weaker strikes, while the axe offers strong but slower attacks. To clearly distinguish the two duelists, different armor is worn by each one[6]. The armor set is chosen mainly due to a big color hue difference or a unique piece that differentiates itself (see Figure 5.1).

The character is controlled with a gamepad and is viewed from an over-the-shoulder third-person camera perspective. Furthermore, this camera automatically focuses on the opponent at a close range, to make sure that the players focus on the opponent and are able to clearly notice the intricacies of their behavior (such as their emotions).

Attacks can be interrupted. This mimics the anticipation stage of *3motion*, where an action is interruptible until a certain threshold is reached[7]. This introduces the ability to perform **feints**, whereby, by canceling an attack before the anticipation becomes uninterruptible, the duelist can start performing a new action sooner than expected. Mechanically, the player clicks and holds the attack button on the gamepad, then, if the player chooses to keep the button pressed (and the threshold is reached), the attack continues, or the button is released, and the attack is canceled. After the anticipation stage, the action enters a follow-through stage, where it no longer deals damage but has not yet finished. This stage has a minimum duration before it can be interrupted, thus creating uninterruptible and interruptible substages, retaining character responsiveness while also limiting control [30, 31].

Block is a short uninterruptable action. When a player decides to block, the character will lift their shield to deflect an incoming attack for a short duration (around 1 second). During this time, any

[4] Unreal Engine 5 (https://www.unrealengine.com/) [July 3rd, 2023]

[5] The current implementation of health used a fixed value of 100.

[6] In the current implementation, the armor does not provide any mechanical feature (i.e. a set of armor does not provide any additional protection against attacks).

[7] In the current implementation, the interruptibility threshold is defined through animation events that determine which state an action is in.

attack damage is nullified and the character is safe. Note however that, after a block, the character is locked for a short duration (around 200 milliseconds), meaning that the character cannot start another action (attack or block) during that time, this allows for the feints to be impactful and allow for more interesting gameplay scenarios.

While a human player controls the duelists, their emotional expressions are independent and showcase how the duelist is experiencing combat. The implementation of this behavior, both of the computation of the emotions and of their expression, evolved over several iterations. For its computation, we made use of different emotion models within *3motion*, making use of anticipatory behaviors; in the end, we used a solution *Emotivectors* [7]. As for the expressions, by the end, it featured body language, facial expressions, colored outlines, and background iconography, which reinforces the emotion expressions and their recognition. Each emotion model and expression modality are detailed in future sections.

5.1.2 The heads-up display (HUD)

During a fight, the characters' health and rounds won are displayed in the Heads-up Display (HUD) (see figure 5.2), on the top left and bottom right corners of the screen. The red bar indicates how much health is left, and a blue square indicates a round won.

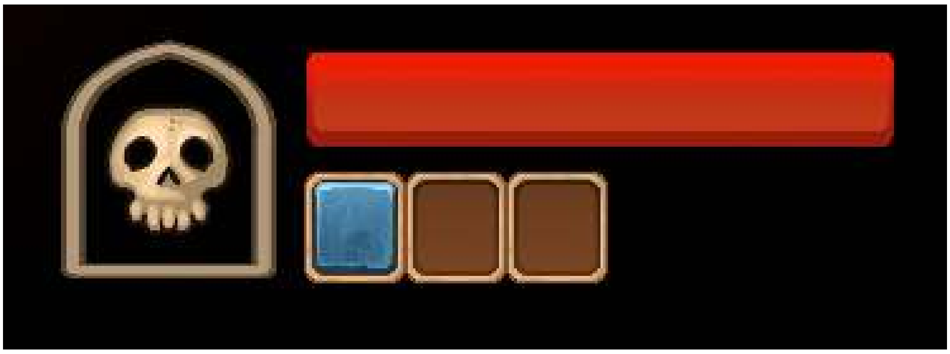


Figure 5.2: Health and rounds won HUD.

5.1.3 Visual effects (VFX)

Early on in development, we noticed that it was hard to notice when the characters hit each other. To fix this problem, we added particles that are expelled when an attack damages the opponent or is blocked by the opponent. Such particles differ depending on the type of interaction (e.g. red particles on a successful strike; white particles on a blocked strike).

In addition to particle effects, we also added a red vignette, a red taint at the edges of the player viewport (see Figure 5.4) to indicate the physical condition of the character. At the start of a fight, the vignette is invisible, but as characters receive damage, the vignette becomes more visible and vibrant. This type of vignette is very commonly found in video games.

5.1.4 3Motion in Adfectus

3motion was introduced into the game as a character component or something that can be attached to the character, each agent step (i.e. Perceive, React, Decide, and Perform) was implemented as different models that could be added to the component and were updated in a loop, running on each frame of the game. While the first implementation of *3motion* was engine-agnostic, we decided to integrate our implementation with the Unreal Engine. This decision allowed for faster and better development, where part of our model could be edited and expanded within the editor (by using the user interface or through visual scripting with the Blueprints). We have already portrayed the use of *3motion* in an illustrative scenario with two duelists, which helps showcase the interactions possible (see Section 3.3).

The *perceive* step, where the character perceives the world around it, was implemented as an active perceiver. On each update, the model perceives some changes in the world and stores them as percepts. Essentially, we keep a log of events about the world. As in the first implementation of the model, we added a delayed perception to allow for more believable behavior. The information collected was divided into two categories: character information and combat events. Character information collects data on the actions, emotions, health, and position of a character, not only from the opponent but also from the character itself. Combat events collect data on what is happening in combat, including strikes that hit, miss, or are blocked. In the illustrative scenario, the *perception* step can be seen, for example, in **(2)**, when Rua perceives and becomes aware of Gorm's attack.

The *react* step, where emotions are computed, had different implementations depending on the evaluation, but in the end, it was implemented in the BALTAS model[8], which incorporated multiple *Emotivectors* [7] that computed emotions based on the anticipatory interpretation of the available percepts. Different implementations are discussed when relevant. In the illustrative scenario, this step can be seen, for example, in **(6)**, where after Gorm blocks Rua's attack, they feel relieved and irritated, respectively, due to the action's outcome; but also in **(2)**, when Rua becomes confident in anticipation to Gorm's attack.

The *decide* step, where actions are computed, was implemented in the GENA model[9] that would internally mimic the character's actions, which were taken by the player. It essentially would interpret the player input and animations' states to determine what action was underway (i.e. light attack, heavy attack, block, etc.) and in which state (i.e. uninterruptible anticipation, interruptible anticipation, etc.). This model helps inform the remaining models of the character's actions state. In the illustrative scenario, this step can be seen any time a character attacks or blocks, for example, in **(1)** and **(3)**, where Gorm and Rua decide to attack, but also in **(4)**, where Gorm decides to cancel his attack and block instead.

The *perform* step, where the actions and emotions are executed, was implemented as an output of information. This model did not actually make changes to the character's internal behavior, instead, it made available what actions and emotions the character was performing (as pure information)

[8]The model was codenamed BALTAS (Belief and Anticipation Laced Tenacious Affect System) as a reference to the author's father.

[9]The model was codenamed GENA (Genial Emotional iNformed Actions) as a reference to the author's mother.

and the other systems would visit this information and change accordingly. This information would also be visible to the other agents' perception models. In the illustrative scenario, this step is seen everywhere, as it represents the actual execution of actions and expressions, for example, in **(2)**, where Rua becomes confident, while the computation of the emotion is done by the *react* step, the actual expression is performed by this step.

Action subdivision was implemented as animation events (see Figure 5.3), which are events that inform the underlying software that an animation reached a certain step in its execution. This information would then be added to the component and fed into the perception models.

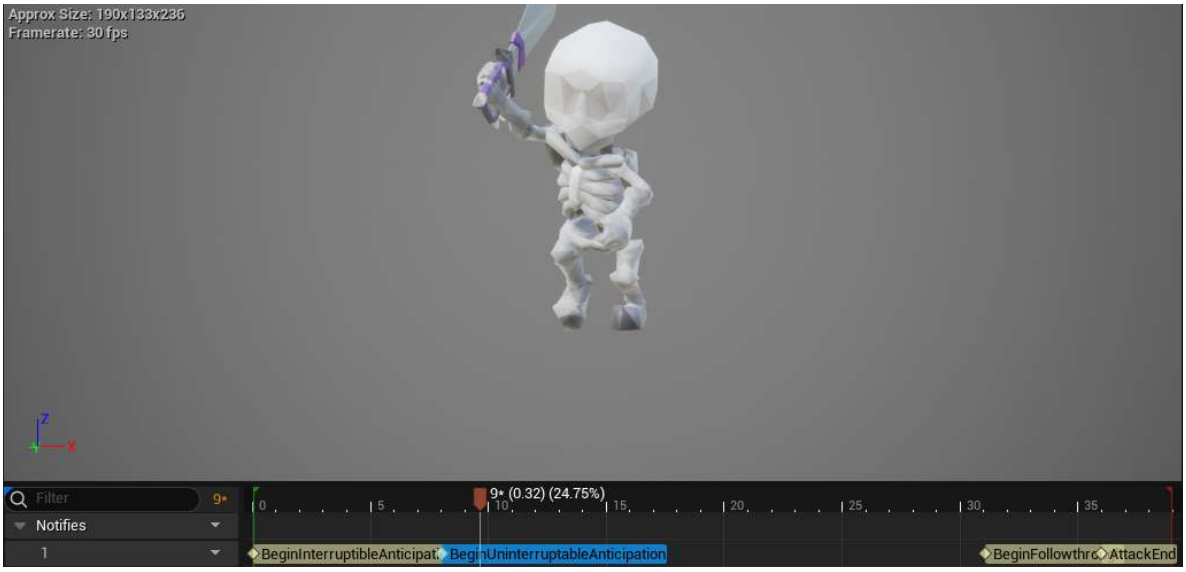


Figure 5.3: Showcase of Animation Events in the sword light attack animation. The character holds its sword about to strike forward. A small indicator on the track at the bottom shows the playing animation as being between frames 9 and 10. Four animation events, each representative of the Anticipation and Follow-through stages proposed in *3motion*, are displayed on the track. The event in blue represents the latest stage change, Uninterruptible Anticipation, meaning that between one event and the next the character is in that action stage.

5.2 First Evaluation on Believability

Similarly to the previous use case using a Text-based Application (see Section 3.5), to validate our hypothesis as presented in Section 1.3, we want to measure the impact on believability of introducing anticipatory emotional reactions in *Adfectus*. In [23], we proposed an emotional model that, while not using *3motion*, makes use of anticipatory behavior and reactions. In this section, we explore the emotional model developed in [23] and showcase the emotion expressions created.

For this study, we created two disparate characters, each with a different appearance, weapons, and set of animations. Character A is a skeleton that uses an enormous two-handed axe to fight his opponents. His attacks are slower but deal more damage. Character B is a goblin with a one-handed sword and shield whose attacks are faster, but deal less damage. After each round, the player characters swap, so that the participants have the chance to try both characters.

To focus on the interactions between only the characters that fight, we propose a scenario with the environment stripped of content (see Figure 5.4). To allow for depth perception and clear positioning

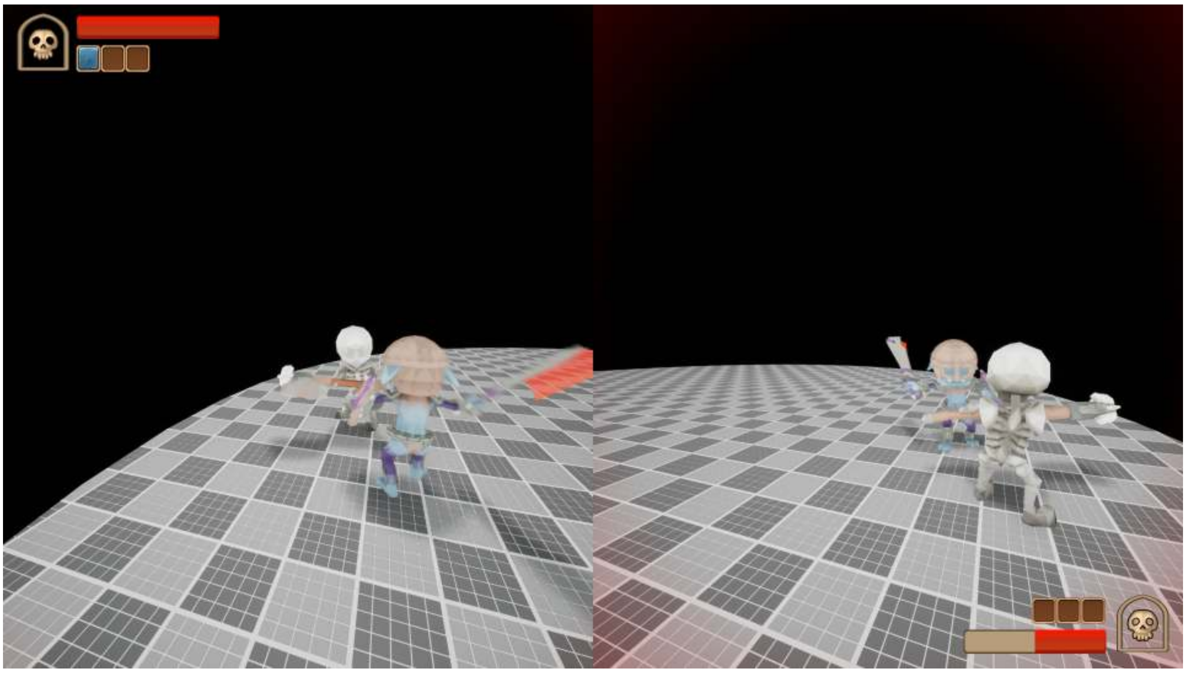


Figure 5.4: Screenshot of Adfectus showcasing a moment where one character is attacking the other. Note that the environment is stripped of content, only presenting a grid-like pattern and the characters.

in the environment, the floors and walls are shown with a grid-like pattern and a light will cast shadows from the characters onto the environment. By placing the characters in such environments, we hope that players and spectators avoid connecting the believability of the interactions with the environment and focus on the interactions between characters.

5.2.1 Emotion Expression

At the basis of our conceptual model are emotions. To evaluate our approach and given the combat orientation of our case study, we choose a subset of relevant affective states to portray: *Confidence*, *Defensiveness*, *Rage*, and *Fear*.

To convey such affective states, we treated the character's body as a set of resources (i.e. face, head, arms, and others), some more available than others depending on the base animation. We searched for the resources that were more available (i.e. not being directly used by the base animation for long periods of time) and applied various transformations (i.e. rotations or translations) to them with the goal of conveying our set of emotions. To design these new animations we took inspiration from the prototypical postures to display the emotions explored in Section 2.6 (see Table 2.2 and Figure 2.11). The following subsections describe how we expressed each emotion[10].

5.2.1.A Fear [11]

We used trembling as the core mechanism to portray fear as it is an effective and clear way of portraying such emotion. The implementation of trembling is done in two steps, the torso, and the

[10] As a point of reference, a video portraying a character from Adfectus in its neutral state can be seen at `https://youtu.be/AZLpaZ512fs`.

[12] A video portraying a character expressing Fear can be seen at `https://youtu.be/d-nwvuUL1AY`.

legs.

To implement torso trembling we rotate the spine bone, which is the parent of all bones in the upper body of the character, over the characters' local up and forward axis using coherent pseudo-random rotations. Such rotations are calculated using a continuous noise function. It is crucial that this function is continuous otherwise the character will jump between rotations and as a result look jittery. The domain and range of this function are also scaled depending on the rotation axis, meaning that each axis of rotation will have a separate trembling frequency and amplitude.

The noise function is also used to generate pseudo-random translations that are applied for each knee bone, causing the legs to tremble. These translations were also made symmetric over the characters' local up axis so that one leg trembles in the opposite direction of the other. This solution caused the feet to also tremble because the knee bone is parent to the ankle bone. To fix this issue we applied the inverse tremble translation to the ankle bones.

5.2.1.B Rage [12]

To portray Rage we chose to create a new set of attack animations following three steps. (1) Exaggerate both the duration and movement during the anticipation, eventually even making the character tremble in order to communicate a large amount of energy the character is placing in the upcoming attack. (2) Decrease the duration of the action execution stage which increases the speed of the attack. (3) Exaggerate both the duration and movement during follow-through to more clearly show the effort required to perform this enraged attack.

5.2.1.C Defensive [13]

To make a character appear defensive we created variations of the idle and walk animations following two steps. (1) Make the character hide behind its weapons whenever possible, as if always prepared to defend against the next incoming attack. (2) Move the whole upper body slightly back, to portray the idea that the character is moving away from the opponent to avoid getting hit.

5.2.1.D Confidence [14]

In order to make the characters appear confident we made them more relaxed and not too worried about the combat. Luckily we found an animation asset on Mixamo[15] portraying a laid-back character. This asset was not made for our characters as it did not account for the weapons that our characters were holding. As a result, we adapted it by first blending the transformations on each finger bone with the finger bone transformations from our original animation and by making the characters place their main weapon on the shoulder, suggesting that the character does not need to be prepared for combat, for example, because he has almost defeated the opponent.

[13] A video portraying a character expressing Rage can be seen at `https://youtu.be/gNpAG2x9o8Q`.
[14] A video portraying a character expressing a Defensive posture can be seen at `https://youtu.be/Km4_ulMcrB8`.
[15] A video portraying a character expressing Confidence can be seen at `https://youtu.be/vJErjy9NB60`.
[15] Mixamo is an online library of free animations. (https://www.mixamo.com/) [July 3rd, 2023]

5.2.2 Emotional Models

Following the *3motion*, an action is split into stages: interruptible anticipation, uninterruptible anticipation, uninterruptible follow-through, and interruptible follow-through. Each stage results in a separate percept meaning that for a single action, a character may receive up to four percepts depending on whether it was finished or canceled. This percept carries information about the action, its type, stage, and information about who performed it such as the character's health. This information is then used by the emotion selection model to choose which emotion the characters should express.

5.2.2.A Health Ratio

To create a simple representation of how well a character is doing in the context of a battle, we used a health ratio. This ratio is, at any time frame, a value that represents the belief of likely victory or loss in the fight and the value we use to inform the selection of a character's emotion. To calculate this value, we considered that it should be closely related to the character's well-being, for example, if the character is close to death then his belief of likely victory or loss will likely tend towards loss. However, if the opponent is also damaged, that should contribute to the aforementioned belief. This means that the health ratio should account not only for the character's well-being but also for the opponent's. More formally we defined the health ratio by the following equation:

$$w = (h_{player} - h_{opponent}) \times 0.5 + 0.5 \tag{5.1}$$

Where h_{player} is the player's normalized health in the interval $[0, 1]$ and $h_{opponent}$ is the opponent's normalized health in the interval $[0, 1]$. As such, when a character's normalized health is equal to the opponent's then the character will believe it is as likely to win as it is to lose (e.g. $h_{player} = h_{opponent} = 0.7 \implies w = 0.5$), however, if its normalized health is lower than the opponent's then it will believe that it is more likely to lose (e.g. $h_{player} = 0.3, h_{opponent} = 0.7 \implies w = 0.3$). As a result, the character should express more negative emotions if w is lower and more positive emotions if w is higher.

5.2.2.B Personality

How we chose an emotion based on the current health ratio is determined by the character's personality. In this context, a personality is defined as a set of health ratio intervals that dictate when the character will express a particular affective state. Whenever the health ratio reaches one interval then the affective state becomes active and the character expresses it as long as the health ratio remains within that interval. Furthermore, these intervals may overlap which will result in multiple affective states being expressed at the same time. In our case, overlapping emotions affect separate body resources or are applied on top of the current transformations applied to the character's body. Table 5.1 is an example of a personality:

While being a simplified model of personality, this approach allowed us to easily create different personalities by simply altering these intervals. For instance, a very confident character could have his *Confidence* interval in the $[0.2, 1]$ interval which could indicate, that almost no matter the odds he

Table 5.1: Example of a personality. Note how Defensiveness overlaps with Fear and Rage.

Affective State	Health ratio interval
Fear	$w \in [0, 0.15]$
Rage	$w \in [0.16, 0.3]$
Defensiveness	$w \in [0, 0.4]$
Confidence	$w \in [0.7, 1]$

still displays confidence. Additionally, with this solution, one can easily add new affective states by creating additional intervals.

5.2.2.C Reactive Model

Considering the above concepts, if we compute the health ratio every time the character receives a percept then the character will express an emotion as soon as either character's health changes. In other words, the character will express emotions in reaction to actions that cause either character's health to change. This behavior was implemented in a *reactive model* and will serve as the base model for comparison as it is a common solution found in games.

5.2.2.D Predictive Model

In this model, the characters express emotions in anticipation of what the opponent might do. Therefore, the model moves emotion expression solely to the anticipation stage.

The characters anticipate by creating a prediction of the outcome of an action and that prediction dictates what emotion they will express. For example, if the opponent initiates an attack and the character predicts that the attack will cause damage then it should react accordingly, perhaps by expressing fear or changing to a defensive posture. In our context, these ideas translate to creating a health ratio prediction during the anticipation stage and using that prediction to express emotions (as opposed to the reactive model which reactions are based on the current health ratio).

We approached the computation of a prediction by relying on a weighted moving average of the value, in our case a perceived health ratio, and its first and second derivatives. Considering $\hat{w}$ the anticipated health ratio, we compute it using the Equation 5.2 below.

$$\hat{w} = WMA_w + WMA_{\dot{w}} + WMA_{\ddot{w}} \tag{5.2}$$

where WMA_w, $WMA_{\dot{w}}$ and $WMA_{\ddot{w}}$ are weighted moving averages of the last k health ratios, their first derivatives, and second derivatives, respectively. Considering w_i is the value of an element at the index i in the history of values, the derivatives are computed as the differences between values of the sequence:

$$\begin{aligned} \dot{w}_i &= (w_i - w_{i-1}) \quad \text{and} \\ \ddot{w}_i &= (\dot{w}_i - \dot{w}_{i-1}) \end{aligned} \tag{5.3}$$

The first derivative offers information on the tendencies of the previous values (e.g. if they tend downward or upward). While the second derivative offers information on the acceleration of such tendencies (e.g. if they are going up/down slowly or quickly). This solution was inspired by the computation of expectations seen in Section 4.3.1.

Having our lists of values, we compute the weighted moving average:

$$WMA_x = \frac{2}{n(n+1)} \sum_{i=n-k}^{n} x_i[i-(n-k)] \tag{5.4}$$

where n is the number of elements in the history of values and x_i is the value of an element (in our case, a perceived health ratio). k refers to the window of values used to compute the average (for our use case, it had a value of 15). This equation linearly decreases the weight of an element in the history, meaning that the more recent an element is the more it will contribute to the prediction. Other predictors were tested, such as the simple moving average or the exponential moving average, but from empirical testing, the weighted moving average was the one that yielded the best results.

Additionally, instead of having a single prediction for the entire anticipation stage, we create separate predictions for each of the substages of anticipation: interruptible and uninterruptible. These predictions are created from the health ratio history of the last action stages that were reached by either stage. This means that each of these stages may have a different history and therefore a different prediction. This approach naturally separates emotional reactions to feints (i.e. attacks that are intentionally canceled in the interruptible anticipation stage) from emotional reactions to attacks that are likely to connect. For example, imagine a fight where Character A performed six attacks. Out of those six, five were feints, and only one actually dealt damage to Character B. In other words, five attacks moved to the interruptible anticipation and then were canceled and only one attack moved through all stages and hit. This means that during the next opponent's attack, the prediction made during interruptible anticipation will anticipate that the attack will be a feint and so the character will not react, however, if the opponent's action moves to the uninterruptible anticipation stage then the character will likely predict that the opponent's attack will hit and as a result become defensive, for example.

5.2.2.E Predictive with Assertion Model

This model adds reactivity to the predictive model by incorporating surprise. Like the predictive model, during anticipation, the character predicts the health ratio and that dictates the emotion it will express during that stage, however, during the follow-through stage, the character can change the currently expressed emotion depending on the mismatch between the outcome and the prediction (i.e. the character can be surprised). An example mismatch would be a situation where the character expected the health ratio to increase but instead it decreased.

If there is a mismatch, we compute a health ratio boost that is added to the current health ratio (i.e. the one predicted during follow-through). This boost is computed as the difference between the real health ratio during follow-through and the predicted health ratio:

$$Boost_w = (w_{follow-through} - \hat{w}) \times d_t \ , \quad d_t \in [0, 1] \tag{5.5}$$

Since surprise is not permanent as we are only surprised for a short period after it happens, we introduce a decay d_t that decreases this boost linearly over time until eventually reaching 0. Additionally, a threshold was used to exclude cases where the values between the real value and the prediction are close to one another.

The addition of this assertion step created really interesting behavior. For example, let us say that a character was hit by the opponent a few times with a light attack. Then the opponent initiates another attack, and as a result, the character will predict a certain decrease in health ratio, taking a defensive posture as a result. This decrease is directly determined by the damage of the opponent's previous light attacks as those are the only type of attacks that the opponent has performed up to this point. However, this time the opponent's attack is heavy which causes a larger decrease in the health ratio than what the character had predicted, resulting in a negative boost. Applying this boost will cause a further decrease in the computed health ratio, resulting in the character, who was only defensive, to start trembling in fear. In other words, the character was expecting to be hit with a light attack but instead was surprised with a heavy attack causing a stronger negative emotional reaction than the one the character had anticipated.

5.2.3 Emotion Expression Validation

We conducted a pilot to validate the transformations we applied to our animations to allow for emotion expressions (see Section 5.2.1). This manipulation check on the emotion expressions is an important step in assuring that the characters in the comparative evaluation we perform next with the emotional models are not being misinterpreted due to issues in the perception of the expressions.

We used a questionnaire in which participants were asked to describe what the characters were feeling in a set of videos. Each video was short and sequentially showed the animation from three perspectives, the front, the side, and the back of the character. After each video, the participant first had to describe what the character was feeling throughout the animation and then answer a 10-element randomly ordered Positive and Negative Affect Schedule (PANAS) questionnaire [117]. This combination of using an open question, allowing for a free-form description of events, and then the use of PANAS, following a structured format, was chosen to better understand the participants' perception of the character.

The emotion validation sample consisted of 5 participants carefully chosen to be representative of the diversity of our population (men and women from different gaming backgrounds). Overall, the participants were able to correctly identify the emotions showcased in the videos. As such, we integrated the emotion expressions in our final experiment without any major changes.

5.2.4 Interaction Believability

To assert which of our emotion models (see Section 5.2.2) create more believable interactions, we prepared an evaluation making use of a questionnaire (see it in full in Appendix D.1) and a version of

Adfectus that contains the three created emotion models. Given the competitive nature of *Adfectus*, the evaluations were always performed with pairs of participants. Each participant would fill out their own questionnaire.

To start the evaluation, participants were asked to answer a series of demographic questions, including their age, gender, the frequency at which they play video games, whether they were familiar with and liked fighting games, whether they interacted regularly with synthetic characters, and how they valued emotion expression in games. Afterward, to become used to the controls and gameplay, the participants played a tutorial version of the game, which does not make use of any emotional model (i.e. the characters did not express any emotion).

Once the participants were comfortable with the tutorial version, they would move to the next part of the experiment where we would start by asking them explicitly, to pay attention to the characters' animations, namely when and how they change. Such a confirmation step is required, because, as shown in other similar works [170] and as we noticed in preliminary evaluations, animation transformations that happen throughout gameplay can easily go unnoticed by players.

In this step of the experiment, we also presented the set of questions related to the believability of the interactions the participants would have to answer after play, once again to make sure that participants looked at exactly what we wanted so that they could answer the questions appropriately. This initial presentation of the main questionnaire also included examples of what we were referring to in the statements, to make sure that they clearly understood what we were referring to. These examples were removed from the actual main questionnaires to avoid biasing the answers.

After this step, participants were asked to play three separate versions of the game. Each with an associated emotion model: *(R) reactive model*, *(P) predictive model*, and *(A) predictive with assertion model*. To reduce the bias from playing the versions in a certain order, each pair of participants played each version in an order based on the possible combinations of versions, meaning that the first pair of participants would play versions $[R, P, A]$, the second would play $[R, A, P]$, the third $[P, R, A]$, and so on until all combinations were met, afterward we looped back to the first sequence.

After playing a version, the participants were asked to rate the following set of statements relating to the believability of the interactions experienced during the gameplay:

Q1 - I understood what the characters were doing.

Q2 - I understood what the characters were feeling.

Q3 - I understood the characters' intentions.

Q4 - The characters expressed emotions in a coherent manner.

Q5 - The characters expressed emotions in reaction to what happens.

Q6 - The characters expressed emotions in anticipation of what could happen.

Q7 - The characters expressed emotions in reaction to what happened considering what was anticipated.

Q8 - The characters' interactions were believable (i.e. coherent given the context in which they happen).

Each statement was rated on a 9-point Likert scale, ranging from 1 - *Strongly Disagree* to 9 - *Strongly Agree*. The first four statements (Q1-Q4) are connected to behavior understandability, emotional expressiveness, predictability, and behavior coherency, all important metrics that contribute to character believability as mentioned in the work of Gomes *et al.*[37]. The next three statements (Q5-Q7) focus on understanding whether the participants perceived the models as we intended. For example, if the participants noticed that in the *(R) reactive model* the characters expressed emotions in reaction to what happens, or if in the *(P) predictive model* the characters were expressing emotions in anticipation of the opponent's attacks. The last statement (Q8) helps us identify, according to the participants, which version had the most believable interactions.

5.2.5 Results

Our group of participants ($N = 32$) was between 19 and 30 years old ($M = 24$, $SD = 2.609$; see Figure 5.5). 10 were female and 22 were male. As for the participants' habits with video games, 50.0% reported making time to play video games, while the remaining played rarely or when the opportunity presented itself. Regarding action/fighting games, most of the participants (68.75%) do not enjoy or are not familiar with this genre of games, furthermore, 59.38% rarely play these types of games. As for experience with synthetic characters, 53.13% interact regularly with synthetic characters (2 participants work with these characters; see Figure 5.6). Additionally, regarding the expression of emotions by characters, most of the participants (93.75%) prefer characters that express emotions, but only 31.25% consider it crucial for their enjoyment.

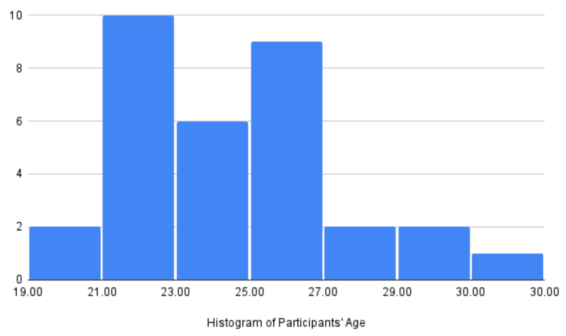


Figure 5.5: Histogram that shows the distribution of the participant's ages.

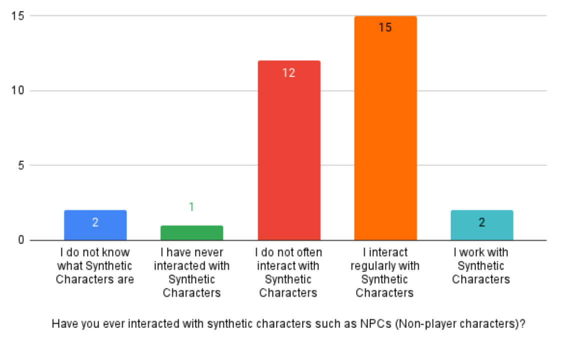


Figure 5.6: Distribution of each case for the question "Have you ever interacted with synthetic characters such as NPCs (Non-player characters)?"

A Shapiro-Wilk test determined that none of the data sets of the statements were normally distributed. To understand if there was a significant difference between versions, for each statement we performed a Friedman test, furthermore, where appropriate a post hoc analysis with Wilcoxon signed-rank tests was conducted with Bonferroni correction applied, resulting in a significance level set at $p < 0.017$.

We concluded that neither the *(P) predictive model* nor the *(A) predictive with assertion model* seemed to improve the believability of character interactions when compared to the more traditional purely *(R) reactive model*, $\chi^2(2) = 4.169$, $p = 0.124$. Nevertheless, there seemed to be positive improvements (although with no statistically significant difference) in the *(P)* and *(A)* models across multiple dimensions regarding the believability of the interactions *(Q3, Q4)* when compared with *(R)*. Furthermore, most participants rated the interactions as believable regardless of version *(Q8)*, suggesting a positive experience even with only the reactive model. Note that, although not significant, participants rated version *(A) predictive with assertion model* ($M = 7.12$, $SD = 1.641$) higher than the other models, *(R) reactive model* ($M = 6.47$, $SD = 1.934$) and *(P) predictive model* ($M = 7.06$, $SD = 1.684$); with more participants, this difference might be significant.

We also found that the participants were not able to correctly identify the emotional behavioral patterns we had intended for each model, i.e. expressing emotions in reaction to what happens when in the *(R)* version, in anticipation of what could happen when in the *(P)* version, or in reaction to what happens considering what was anticipated when in the *(A)* version. We found that participants rated the *(P) predictive* version higher when compared to the *(R) reactive* version in terms of how much the characters were expressing emotions in reaction to what happens *(Q5)*, which is the opposite of what we had intended for that version. This may indicate that some emotions may not be clear enough, resulting in some participants not noticing them. Another likely possibility is that the participants may be noticing the emotions only for a small interval during their expression, perhaps because they are focused on combat. A small delay in perception might be enough to perceive the predictive model as if it were reactive. There is also the possibility that the duration of these interactions needs to be larger in order for the participants to notice the emotional behavior or maybe the differences between the three models are just too subtle and therefore participants will always have a hard time correctly identifying their patterns.

In an effort to better understand the results, we performed a subgroup analysis. We found that for a subgroup of *participants who are more familiar with synthetic characters* ($N = 17$) there were statistically significant improvements in two dimensions that refer to the believability of interactions *(Q3, Q4)*. Starting with the characters' intentions *(Q3)*, in this subgroup, a Friedman test showed a statistically significant difference between the three versions ($\chi^2(2) = 7.960$, $p = 0.019$), furthermore, a Wilcoxon signed-rank test showed that this subgroup was able to better understand the characters' intentions in the version *(A) predictive with assertion model* when compared with version *(R) reactive model* ($Z = -2.658, p = 0.008$). Surprisingly, however, the comparison between the *(R) reactive model* and the *(P) predictive model* did not yield the same result ($Z = -1.565, p = 0.118$). One would expect that having characters express emotions in anticipation of what could happen would be enough to improve how the participants understand the characters' intentions, however, that only seemed to be the case when a reactive emotional response was also present, i.e. the *(A) predictive with assertion model*; reinforcing the importance of the follow-through stage in the model.

This subgroup also showed statistically significant differences in referring to the degree to which characters expressed emotions coherently *(Q4)*, $\chi^2(2) = 10.051$, $p = 0.007$. Post-hoc analysis with a

Wilcoxon signed-rank test showed a statistically significant improvement from the *(R) reactive model* to both the *(P) predictive model* ($Z = -2.393, p = 0.017$) and the *(A) predictive with assertion model* ($Z = -2.680, p = 0.007$), which reinforces the importance of considering anticipation to coherently express emotions.

In addition to the results from these analyses, we also reached interesting conclusions from the comments of the participants during user testing. To our surprise, multiple participants appeared to be very focused on the characters' facial expressions when trying to notice the emotional behavior, even after being informed to look at the characters' animations and being given examples reinforcing that message. This could mean that if our characters had facial emotional expressions, the models might have had larger differences.

We also noticed that, even though we had previously validated the emotion expression of rage, participants were misinterpreting it in the context of combat. Whenever it was expressed, participants would simply describe the character as becoming "slow out of nowhere", while justifying this behavior as the character feeling tired or scared. We believe this is due to the fact that rage is the only emotion that affected gameplay by increasing the duration of the anticipation and, as a result, the attack became slower. Nonetheless, the rest of the emotion expressions seemed to be clear whenever they were noticed.

Another important problem identified by the participants was the high damage of the attacks from the character that held a two-handed axe. While it can be a balanced match if the players have some experience with combat games, many participants were not experienced which led to them quickly losing against the character with the two-handed axe and as a result barely having any time to notice any emotional behavior.

5.2.6 Discussion

In this section, we started by looking into ways of expressing emotions that would make sense in the context of our case study *Adfectus*. This resulted in the adaptation of the existing neutral animations into a new set of animations capable of expressing *fear*, *rage*, *confidence*, and *defensiveness*. Preliminary evaluations validated these new emotion expressions as able to portray the expected emotion to participants. Additional minor tweaks were added to the animations to improve their clarity.

With this set of emotions in place, we proceeded to the implementation of the emotion models, responsible for choosing which emotions and when to express them. Three variations were implemented. The *reactive* model had characters express emotions in an immediate reaction to what happens and served as a comparison base as it is the most common solution found in games. The *predictive* model had characters express emotions in anticipation of what could happen. And lastly, the *predictive with assertion* model had characters express emotions in reaction to what happens considering what was anticipated.

To ascertain this work's hypothesis, we resorted to user testing following a within-subjects design. As such, participants got the chance to experience all three models in various orders and answered a set of questions referring to interaction believability and to emotional behavioral patterns.

We then analyzed the results and found that while for the entire sample, there are no significant differences between models, for players who are more familiar with synthetic characters, both the *predictive* and *predictive with assertion* models result in more coherent emotion expression and better expression of characters' intentions when compared to the *reactive* model, which indicates that, at least for this particular demographic, these models result in more believable interactions.

5.3 Understanding the Perception of Adfectus' Characters Expressions

In the previous evaluation, some participants reported that the character was "tired out of nowhere" or afraid when trying to attack. After close inspection, what the participants were referring to was the expression of *Rage*. At the time, this emotion was expressed through an exaggerated and prolonged swing of the weapon. During the preliminary expression validation, this emotion was perceived correctly, yet during gameplay it was perceived differently. To try and fix this misinterpretation, we improved the expression of *Rage*, making it present in the walking and idle stances, and further refined all other expressions by introducing other emotion expression channels (e.g. facial expressions) that we will discuss below.

To validate and understand whether our improvements were being perceived correctly, as in our previous work on Virtual Coaches (see Section 4), we used the Triad Affect Interpretation (TAI) method [22, 145] to build a model of how users perceive different expressions of emotions of *Adfectus* characters. Refer to Section 4.6 for an in-depth description of the method. It is important to note that using this method on these characters also helps to demonstrate the flexibility of this method and how it can be applied to different characters and body parts (since these characters are full-bodied with a cartoon look, whereas characters described earlier were only visible from the top of the shoulders in a realistic look). In this section, we will focus on the changes made to the character's expressions and the adaptation of the method to our use case, as well as our findings on applying this method to these characters.

5.3.1 Emotion Expression Improvements in Adfectus' Characters

As stated previously, after the first evaluation, we found that a particular emotion, *Rage*, was being misinterpreted. Furthermore, due to the third-person perspective used to control the duelists, certain expressions may be obstructed, therefore the use of secondary channels is needed to best showcase the character's expressions throughout the gameplay. Additionally, the nature of the game must also be considered. Adfectus is a fighting game, hence players will very likely be immersed in combat, focused on defeating the opponent and so any subtle expression of emotion, will probably be missed. In an effort to fix the emotion misinterpretation and expand on the emotion expression capabilities of the character, while aware of the works on emotion expression explored in 2.6, here follows our implemented changes.

5.3.1.A Rage

Figure 5.7: Updated expression of Rage in Adfectus. The expression now is reflected in the idle and walk animations.

We improved the expression of *Rage*, making it present in the walking and idle stances, while removing the old animation employed during the attack actions. This makes the expression more in line with the remaining expressions and removes any gameplay differences between them. The animation was developed using Blender[16] (version 3.5). We created 4 animations for the combination of being in the idle[17] or walking[18] state and either carrying a sword or an axe. We made use of the stock animations (the ones we use for the neutral expressions) and additively made manipulations to the posture and movement to create the new animations.

We took inspiration from the expression of *Anger* in Table 2.2. And so, the expression of *Rage* can be described as a forward-leaning body, open arms, forward-facing head, palms heading downwards, and more frantic movement (see Figure 5.7).

5.3.1.B Facial Expressions

We added *Facial Expressions* to the characters. We choose to use an iconic representation (similar to those in Figures 2.7 and 2.5), which are simplified but recognizable faces, which ease development and increase recognizability. In Figure 5.8, we showcase each facial expression.

Here is a description of each facial expression.

Neutral has open round eyes and a straight line for the mouth. During empirical testing, we tried removing the mouth, but the appearance would be uncanny in some low-light cases.

Confidence has open eyes with eyebrows lightly covering the eyes and a smile. This representation took inspiration from the representation of *Cruelty* seen in [84].

[16] Blender (`https://www.blender.org/`) [June 20th, 2023]

[17] A video portraying the expression of Rage in the idle state wielding a Sword can be seen here `https://gitlab.com/RicardoEPRodrigues/tai-evaluation/-/blob/v1.0.1/videos/TAI-%20sword%20idle%20rage.mp4`

[18] A video portraying the expression of Rage in the walking state wielding a Sword can be seen here `https://gitlab.com/RicardoEPRodrigues/tai-evaluation/-/blob/v1.0.1/videos/TAI-%20sword%20walk%20rage.mp4`

(a) *Neutral* **(b)** *Confidence* **(c)** *Defensiveness* **(d)** *Fear* **(e)** *Rage*

Figure 5.8: Showcase of the facial expressions used in Adfectus. The emotions expressed are the following from left to right: *Neutral*, *Confidence*, *Defensiveness*, *Fear*, and *Rage*.

Defensiveness is perhaps the most complex one. It is also the only one that is not symmetrical. This expression uses one open eye and one closed. The closed eye is represented by a curved line arching upwards, as to be the opposite of the smile in confidence. Furthermore, the mouth here is represented by a squiggly line.

Fear is represented by a big open mouth and wide open eyes. This representation took inspiration from the representation of *Terror* seen in [84].

Rage is mostly represented with strong angular lines. The eyes are partially obstructed by the eyebrows that "collide" into an angle. The mouth is also portrayed as a line with a sharp bent pointing upwards.

5.3.1.C Colored Outlines

We implemented *Colored Outlines*. This secondary channel reinforces the emotions felt by making use of colors that outline the body of the character, each depicting an emotion. When a character is in a neutral expression (no emotion being expressed), no outline is shown. We take inspiration from the mapping of colors to emotions seen in Section 2.6.1. Table 5.2 presents each emotion, its associated color, and the corresponding hexadecimal color value. A description of each outline is shown below.

Confidence is represented by a bright yellow, similar to what is used for *Joy* in [92] and in the film *Inside Out* [193].

Defensiveness is represented by a soft cyan color, similar to what is used for *Surprise* in [92] and is also used to represent *Sadness* in the film *Inside Out* [193].

Fear is represented by a darker purple, similar to what is used for *Grief* in [92] and is also used to represent *Fear* in the film *Inside Out* [193].

Rage is represented by a bright red, similar to what is used for *Anger* in [92] and in the film *Inside Out* [193].

Table 5.2: Color selection for the outline for each emotion.

Emotion	Color	Hex
Confidence		#F8F495
Defensiveness		#94FDDB
Fear		#D79AF6
Rage		#FF0000

5.3.1.D Background iconography

We implemented *Background iconography*, a secondary channel reinforcing the emotions felt, in the form of colored halos shown over the heads of the characters, such as the ones seen in the game *Darkest Dungeon* [189] (see Figure 2.9 in Section 2.6.1).

(a) *Confidence* **(b)** *Defensiveness* **(c)** *Fear* **(d)** *Rage*

Figure 5.9: Showcase of the background iconography used in Adfectus. These images were displayed above the character's head. The emotions expressed are the following from left to right: *Confidence*, *Defensiveness*, *Fear*, and *Rage*.

We created halos using the base design seen in *Darkest Dungeon* [189] and applying upscaling, color shifts (to match the *Colored Outlines*), and adding additional elements when needed (see Figure 5.9). The halos are only displayed when the character switches from one emotion to the other. The halo appears over the character's head, moving upward and then slowing down before fading away. A description of each halo is shown below.

Confidence is represented by a white halo with irregular forms resembling a paint brush stroke. This shape tries to invoke calm and resolve.

Defensiveness is represented by a cyan halo with fading lines, crossed with strikes resembling thorns pointing inward, and a shield at the top. The shield was added to reinforce the defensive nature of the feeling, and the symbology was inspired by the *Guardian* class from the game *Guild Wars 2* [178], which is focused on protecting itself and others.

Fear is represented by a purple halo with regular lines with small dents (as if something almost broke it), and at the top the shape finishes with lines pointing upwards and outwards of the character, suggesting a desire to escape.

Rage is represented by a bright red halo with strong lines and crossed strikes pointing inward (similar to Defensiveness but more well-defined). The overall shape has hard edges as if touching it would cut.

While prototyping the feature, we found that the halo of the duelist closest to the camera (the one being controlled by a player in their respective viewport) would obstruct most of the view. Therefore, we decided to limit the range in which a halo is visible. Making it invisible when too close to the camera. In practice, this means the halo is only shown on the opposing duelist. We address this limitation later.

5.3.1.E All Together Now

Having all the new features, and joining them with the previous ones, we have all new expression capabilities in our characters. Figure 5.10 presents a snapshot taken of our character portraying each of the four animations[19]: *Confidence*, *Defensiveness*, *Fear*, and *Rage*. In the figures, we can see that each new feature is expressed independently but in conjunction with the remaining capabilities.

(a) *Confidence* **(b)** *Defensiveness* **(c)** *Fear* **(d)** *Rage*

Figure 5.10: Snapshot of the four emotion animations in Adfectus in the idle stance and holding an axe. The emotions expressed are the following from left to right: *Confidence*, *Defensiveness*, *Fear*, and *Rage*.

5.3.2 Using the TAI method for Manipulation Check of Emotion Expression in Adfectus

Having discussed the improvements in the characters' expressions, we need now to validate them and understand how they are actually perceived by applying the TAI method (thus performing a manipulation check). Looking at the definition of the method, we can summarize it in these steps: select and present the *elements*, elicit the *constructs*, rate the *elements* against the *constructs*, and create the model using multiple statistical analysis methods. Therefore, what follows are adaptations we made to apply this method to our characters.

5.3.2.A Selection of the Animations of Emotions

One main change is, of course, the *elements*, which are no longer only facial expressions, but the full body of the character, including facial expressions, body motion, colored outline, and background iconography. Figure 5.10 (in Section 5.3.1.E) presents a snapshot taken of our character portraying each of the four animations: *Confidence*, *Defensiveness*, *Fear*, and *Rage*. Given the possible different animations available during gameplay, we selected a subset of those animations, namely we blended the expression of the four emotions with either an idle or walking state, and with either carrying a sword and shield or an axe. Therefore, we ended up with a total of 16 animations, as a combination of walk state ($N = 2$), weapon choice ($N = 2$), and emotion ($N = 4$).

[19]Videos portraying the different emotion expressions can be seen here `https://gitlab.com/RicardoEPRodrigues/tai-evaluation/-/tree/v1.0.1/videos`

5.3.2.B Presentation of the Animations

The second step is to select how to present the elements to the participants. Having a total of 16 animations, if we were to show all possible combinations of 3 to all participants ($C_3^{20} = 560$), it would be impossible to complete in a reasonable time.

To balance the amount of work required by each participant, we decided to display only a fixed number of combinations. We wanted the participant to experience all possible combinations of emotions ($C_3^{4} = 4$). For one of these combinations, we match it with a combination of walk state and weapon choice ($C_3^{2+2} = 4$), so that each emotion is expressed with the same walk type and weapon. With this solution, we have a total of 16 possible combinations to display to the user ($4 * 4 = 16$). To further shorten the work required by each participant, we decided to show only 4 combinations; this allows for the elicitation of an adequate number of constructs while not tiring the participants.

To avoid any order bias, we used a *Latin square* design to order the animations, guaranteeing that each participant always saw all possible combinations of emotions, but varying the combinations of walk state and weapon choice shown. Additionally, the order of the videos was randomized.

5.3.2.C Personal Constructs

In this step, we collect relevant *constructs*, in our case the mentioned full-bodied animations, and then sort them into themes using content analysis. In compliance with the defined method, we created a questionnaire to gather said constructs (see it in full in Appendix D.2.1) and consists of four sections, each presenting a combination of three animations. The questionnaire was built using the jsPsych framework[20] and the JATOS platform[21]. The source code and anonymized data is freely available[22] Each section had four questions (equal to the ones in the Virtual Coach evaluation):

Q1: *"By comparing the three animations presented above, identify the two that are alike."*

Q2: *"How are two of them alike and at the same time different from the third."*

Q3: *"Provide us with one characteristic that you found was alike in the two animations."*

Q4: *"Provide us with the opposite characteristic from the one mentioned above, describing the different animation."*

Based on issues detected during the pilot, some notes were added to help the creation of meaningful constructs, namely: *"please do not repeat characteristics from previous sections"* and *"please describe the observed behavior supporting your answers and not your interpretation of the behavior. For instance, do not use references to emotions (e.g., anger) but refer to the underlying behavior supporting your interpretation (e.g., grumbling).".* Two attention-check questions were added to filter out unattentive participants, failing any attention-check would terminate the experience in a failure state.

After gathering all responses, we used content analysis [165] to find meaningful constructs by sorting them into themes and analyzing the frequency with which they appeared.

[20]jsPsych JavaScript Framework (`https://www.jspsych.org/`) [May 25th, 2023]
[21]JATOS Platform - "Just Another Tool for Online Studies" (`https://www.jatos.org/`) [May 25th, 2023]
[22]Access the questionnaire source code at `https://gitlab.com/RicardoEPRodrigues/tai-evaluation/-/tree/v1.0.1/`

5.3.2.D Rating Animations against the Constructs

After selecting the most elicited constructs, the final step was to ask participants to rate the animations with the chosen constructs. We created a second questionnaire for this purpose (see it in full in Appendix D.2.2) using the Qualtrics platform[23]. A template of the Qualtrics questionnaire is freely available[24] Each animation was shown in its individual section and participants were asked to rate the animation using a 7-point bipolar scale between the two opposing words that define each of the selected constructs (e.g. "Grounded Posture" versus "Shaky Posture").

To avoid any order bias, the order of the animations was randomized and every participant would rate all 16 animations. Three attention-check questions were added to filter out unattentive participants, failing at least two attention-checks would terminate the experience in a failure state.

After gathering all responses, we create a model of how participants perceive the characters by performing a PCA. Afterward, using the created model, we ran a Cluster Analysis to identify prototypes of the more distinct emotions.

Both questionnaires collected demographic data, such as age, gender, maternal language, and interaction habits with synthetic characters.

5.3.3 Results

Having updated the method for our use case, we want to apply it to our characters. We present our results by dividing the presentation between the *elicitation* of the constructs and their *rating*.

5.3.3.A Elicitation of the Constructs

In this first stage, we used the questionnaire to collect constructs that the participants and the surrounding community understand. Participants for this evaluation were recruited using Prolific[25]. We restricted participation to people originating from European countries, since the perception of emotions differs from culture to culture, as discussed in Section 2.3. We selected European countries for this evaluation, as they share the same (or similar) culture as the participants in the previous study (see Section 4.6).

Our study included a total of 39 participants, with 10 samples being rejected due to failed attention checks (1 sample) and failure to follow given instructions (9 samples), namely, the participant used emotion words when explicitly asked not to, and the participant repeated the same characteristics when asked not to.

The remaining sample ($N = 29$) was between 19 and 54 years old ($M = 24$, $SD = 7,317$). 51% were women and 49% were men. Regarding their experience with synthetic characters, 96.5% had interacted with NPCs in video games, and 75.8% had interacted with virtual agents in other contexts. 20.6% reported having interacted with social robots in the past and 2 had worked with/developed synthetic characters and/or virtual agents.

[23] Qualtrics (`https://www.qualtrics.com/`) [May 25th, 2023]

[24] Access the template of the animation rating questionnaire at `https://gitlab.com/RicardoEPRodrigues/tai-evaluation/-/blob/v1.0.1/data-stage2/Ricardo_-_TAI_Method_-_Adfectus_-_Part_2.qsf`

[25] Prolific (`www.prolific.co`) [May 3rd, 2023]

Each participant identified four different pairs of constructs for a total of 116 pairs. We discarded 15 constructs for the following reasons:

- the participant used emotions to describe the different constructs (12 constructs had this issue);
- the participant used the same construct more than once (2 constructs had this issue);
- the construct was too broad or unspecific, e.g. “Casual” (1 construct had this issue).

We then took the remaining 101 constructs and performed a content analysis, in which we organized the constructs around common themes (e.g. the construct “Smoothness/Twitching” was merged with the construct “Grounded Posture/Shaky Posture”). As a result of the content analysis, we identified 25 distinct constructs. We additionally decided that a frequency below 3 occurrences was not meaningful enough to be considered. Given that, a total of 13 final meaningful constructs were selected. They are listed in Table 5.3.

Table 5.3: Selected meaningful constructs, defined by associated word pairs and sorted by frequency.

Construct		
Pole 1	**Pole 2**	**Freq.**
Grounded Posture	Shaky Posture	17
Natural Expression	Unnatural Expression	13
Weapon Held High	Weapon Held Low	10
Grimace Facial Expression	Cheerful Facial Expression	8
Ready to React	Relaxed	7
Prepared	Unprepared	7
Fast Walk Speed	Slow Walk Speed	7
Weapon Held in Front	Weapon Held Back	7
Leaning Back/Upright	Leaning Forward	6
Close Body Posture	Open Body Posture	4
Head Down/Low	Head Up/High	3
Offensive Stance	Defensive Stance	3

5.3.3.B Rating the Expression of Emotions

Using the collected constructs, we can now rate the animations that express emotions through our second questionnaire. Participants for this evaluation were recruited using convenience sampling from the Clickworker[26] platform and national pool. As in the previous step, we restricted participation to people originating from European countries.

Our study included a total of 98 participants, with 13 samples rejected due to failed attention checks (5 samples) or completion of the questionnaire being too quick (less than 8 minutes; 8 samples).

Our remaining sample ($N = 85$) was between 20 and 76 years old ($M = 38$, $SD = 10,61$). 31,4% were female, 67,4% were male, and 1 participant identified as “non-binary / third gender”. Regarding the experience with synthetic characters, 70.5% had interacted with NPCs in video games, and 89.4% had interacted with virtual agents in other contexts. 31.7% reported having previously interacted with social robots and 11,8% had worked with or developed synthetic characters and/or virtual agents.

[26]Clickworker (`https://www.clickworker.com/surveys/`) [May 3rd, 2023]

Using the ratings of the participants, with a total of 1360 cases, i.e., the 16 different animations of emotions multiplied by the number of participants (85), we built our model for *the perception of emotions animations* for the *Adfectus* characters using both PCA and the cluster analysis.

A – Principal Component Analysis – To group correlated constructs, we perform a Principal Component Analysis (PCA) with orthogonal rotation (varimax). The Kaiser-Meyer-Olkin measure verified the sampling adequacy for the analysis, KMO = .791 ('good' according to [171]). Bartlett's test of sphericity $\chi^2(55) = 5367.31$, $p < 0.001$, indicated that correlations between items were sufficiently large for PCA. An initial analysis was run to obtain eigenvalues for each component in the data. We decided to use 3 components that in combination accounted for 60.4% of the variance and had eigenvalues over Kaiser's criterion [171] of 1 (see the complete analysis in Appendix D.2.3). Using this number of components, we create the rotated component matrix depicted in Table 5.4. Items with a loading less than 0.4 were removed as recommended and as a standard process [166].

Table 5.4: Rotated Component Matrix for 3 components C1, C2, and C3 with an absolute value above 0.4.

Construct		C1	C2	C3
Ready to React	Relaxed	0.882		
Grimace Facial Expression	Cheerful Facial Expression	0.857		
Weapon Held in Front	Weapon Held Back	0.740		
Prepared	Unprepared	0.704		
Close Body Posture	Open Body Posture	0.697		
Head Down/Low	Head Up/High	0.592		
Leaning Back/Upright	Leaning Forward	-0.490		
Weapon Held High	Weapon Held Low		0.776	
Grounded Posture	Shaky Posture		0.651	
Offensive Stance	Defensive Stance			0.731
Fast Walk Speed	Slow Walk Speed			0.724

From this PCA, we created our *model for the perception of animations of emotions*. Having each component subdividing the expression into its more relevant features for emotional expression:

Component C1: represents a more relaxed stance, where a cheerful facial expression is shown, the weapon is held back, and the character is perceived as being unprepared with the head up while leaning back. It often represents the expression of *Confidence*. Its negative captures the expression of *Defensiveness* and *Rage* affective states.

Component C2: is expressed by a shaky posture and the weapon being held low. It often represents the expression of *Fear*.

Component C3: is defined by a defensive stance and a slow walking speed. It captures the expression of *Defensiveness*. However, its negative expression expresses *Rage* (with an offensive posture and fast walk speed).

Figure 5.11 shows the distribution of cases labeled with the emotion they express, each axis representing a different component in 3D space. Each case is represented using its factor score. Since a factor is, by nature, unobserved, we used the regression method to generate a plausible factor score for each case.

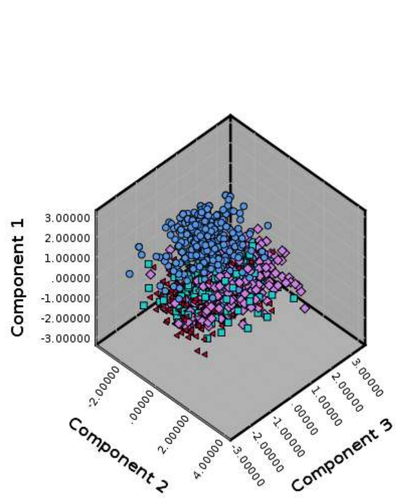


Figure 5.11: Distribution of the cases labeled with the emotion it represents, using the calculated factor scores for each case.

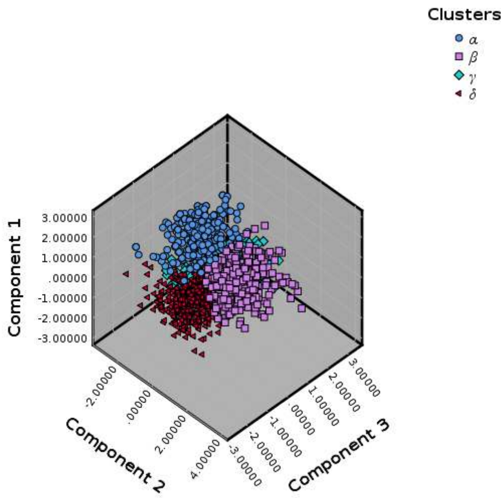


Figure 5.12: Distribution of the cases labeled with the cluster it belongs, using the calculated factor scores for each case.

B – Cluster Analysis – Using our model created from the PCA, we performed a Cluster Analysis to group emotions together based on similarities, thus allowing us to verify if the different emotions would be grouped in the same cluster. We used k-means clustering with the squared Euclidean distance method to minimize within-cluster variances. We considered the factor scores from the 3 components model for the variables. Since this clustering method requires an a priori number of clusters, we had to understand the optimal number of clusters for our data. We concluded that 4 clusters were ideal for our study (see the complete analysis in Appendix D.2.4).

Figure 5.12 shows the distribution of cases for each cluster (the center of each cluster can be seen in Table 5.5). Figure 5.13 shows the means of each component per cluster. This helps us understand how each cluster is expressed.

Table 5.5: Cluster centers for each component in the model for the perception of animations of emotions.

	Cluster			
Component	α	β	γ	δ
1	1.5	0.2	-0.8	-0.5
2	-0.8	1.1	-0.5	-0.1
3	0.0	0.2	0.9	-1.0

To verify whether emotions are being perceived correctly and not being confused with each other, we created Table 5.6 that shows the number of cases of each emotion in each cluster. The percentage distribution of cases per emotion can be seen in Figure 5.14. With this information, we can not only understand how each cluster is expressed, but also understand how emotions are distributed in these clusters. We can summarize each cluster as follows:

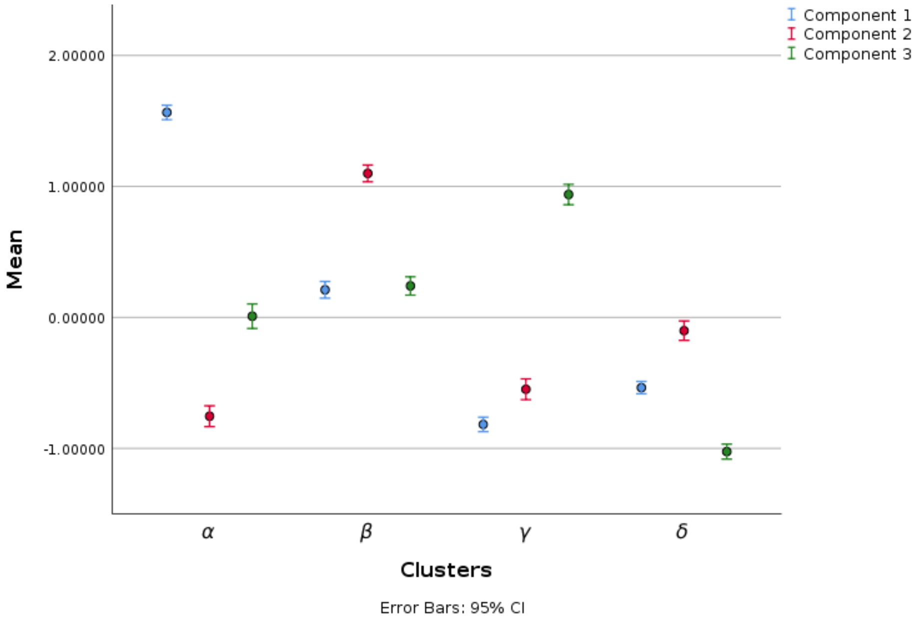


Figure 5.13: Chart of the clustered means of PCA components by cluster. Values close to 0 seem to not impact the perception of the cluster expression. For example, Cluster α is defined by a positive presence of Component 1 and a negative presence of Component 2, while Component 3's presence is negligible.

Cluster α is mainly represented by the positive value of Component 1, with a relaxed posture and cheerful face, and a negative value of Component 2, with a grounded posture and the weapon held high. It captures the expression of *Confidence*.

Cluster β is mainly represented by the positive value of Component 2, with the weapon held low and a shaky posture. It captures most of the expression of *Fear*, but we also see the presence of *Defensiveness* here.

Cluster γ is represented by a negative value of Component 1, with the weapon held in front, a prepared stance, and closed body posture, and Component 2, with a grounded posture and the weapon held high, and a positive representation of Component 3, with a defensive stance and slow walk speed. It captures the expression of *Defensiveness*, with a smaller presence of *Rage* and *Fear*.

Cluster δ is represented by the negative values of Component 1, with a grimace facial expression, weapon held in front, leaning forward, and prepared, and Component 3, with an offensive stance and fast walking speed. It captures the expression of *Rage*, while also having some presence of *Defensiveness*.

Each cluster has a predominance of a specific emotion, having a one-to-one mapping to each emotion. Yet, there is some confusion in the perception of some emotions, namely those of negative valence. *Defensiveness* is the emotion that is less powerful in its presence, but it is still different enough to be distinguished as a separate emotion. Overall, we can say that each emotion is perceived correctly and independently from the others, where the most identifiable is *Confidence* (in Cluster α), and the least identifiable is *Defensiveness* (in Cluster γ).

Table 5.6: Number of cases of each emotion in each cluster. The cases were compared with the squared Euclidean distance between their respective factor scores. The values in bold show the largest sample of emotion in each cluster.

Cluster	Confidence	Defensiveness	Fear	Rage	# of cases
α	**253**	1	3	0	257
β	58	97	**179**	44	378
γ	8	**138**	84	101	331
δ	21	104	74	**195**	394
Total		**340**			**1360**

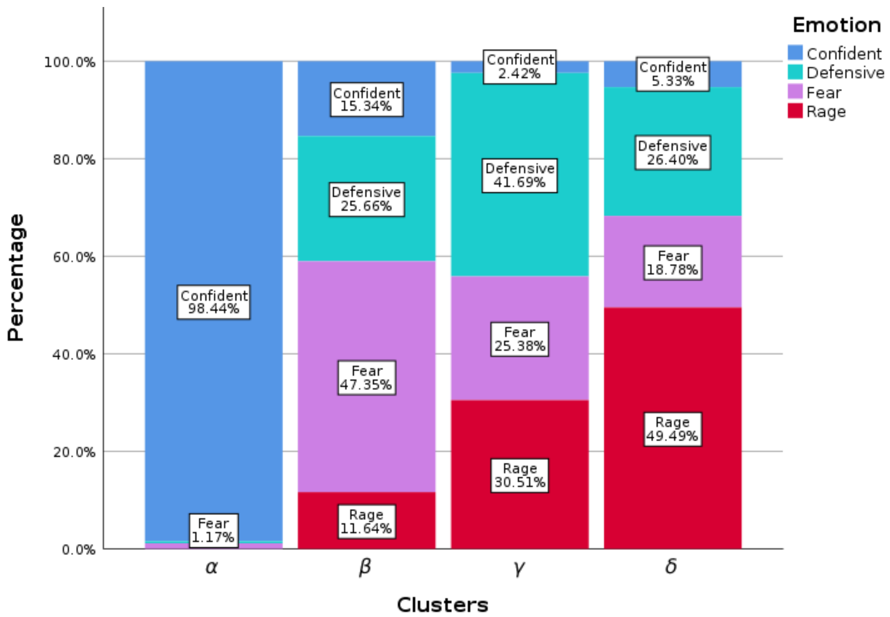


Figure 5.14: Percentage of cases of each emotion in each cluster.

5.3.4 Emotions, Components, and Clusters

At this point, we have explored how the expression is defined in each Component, how different Clusters are expressed through those Components, and what emotions they hold. Is there still information to be extracted? We looked back at the Components from the PCA and looked at their linkage to the emotions to better understand how each emotion is perceived. We devised Figure 5.15 that shows the clustered means of each component per emotion.

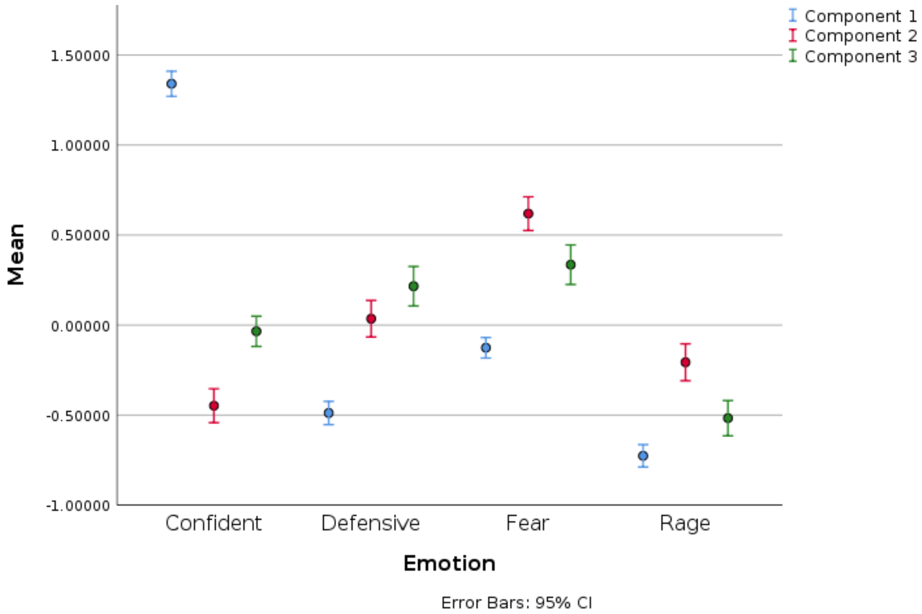


Figure 5.15: Chart of the clustered means of PCA components by emotion.

Note that using the clustered means of the PCA components per emotion could produce misleading results. If the emotions are often confused with one another, then the difference in means is not significant and cannot be used to differentiate between emotions correctly. Fortunately, as we saw in Cluster Analysis, we can assume that all emotions are perceived as separate entities.

If we look back at Figure 5.13 and compare the distributions with these results, it is no surprise that there is a similar mapping between the two. The Cluster α is almost identical to *Confidence*. The Cluster β is similar to *Fear*, having a small difference in the value of Component 1. The Cluster γ holds the most significant difference compared to *Defensiveness*, which is to expect since the cluster also has a lot of *Rage* and *Fear* cases. And the Cluster δ that is similar to *Rage*, where the difference is shown in the intensity of Component 1.

Using this analysis, we can summarize each emotion as follows:

Confidence is mainly represented by the positive value of Component 1, with a relaxed posture and a cheerful face, and a negative value of Component 2, with a grounded posture and the weapon held high.

Defensiveness is represented by a negative value of Component 1, with the weapon held in front, a prepared stance, and closed body posture, and a positive representation of Component 3, with a defensive stance and slow walk speed.

Fear is mainly represented by the positive value of Component 2, with the weapon held low and a shaky posture, and of Component 3, with a defensive stance and slow walk speed.

Rage is represented by the negative values of all three components. Component 1, with a grimace facial expression, weapon held in front, leaning forward, and prepared, Component 2, with the weapon held high and a grounded posture, and Component 3, with an offensive stance and fast walking speed.

5.3.5 Discussion

In this section, we sought to understand if we could improve our characters' expressions. We knew the participants had issues identifying the expression of *Rage* and, due to the combat nature of the game, we wished to expand on the emotion expression capabilities of the characters.

We introduced changes and new features to the characters to make this possible. We reworked the expression of *Rage* to be more in line with the other emotions. Added *facial expressions*, through the use of emoji-like figures. Added *colored outlines* with colors inspired by research and media. And added *background iconography* in the shape of halos to reinforce the emotion expressed.

With all these changes, we needed to validate and understand if the improvements to the characters' animations and expressions were being perceived correctly. More than that, we wanted to understand how people perceived these expressions (not only if they were correctly perceived). To that end, we employed the *TAI method* to help us create a model of the perception of the characters.

After *construct* elicitation, we collected 13 meaningful constructs that describe how the participants ($N = 29$) view the characters. The constructs, while varied, focused on descriptions of body movement (e.g. *Grounded/Shaky Posture*, *Head Down/Up*, *Offensive/Defensive Stance*). Only one constructed discussed the face, *Grimace/Cheerful Facial Expression*, which separates positive expressions from negative expressions. Additionally, no meaningful constructs refer to background iconography or color outlines (while some users mentioned these features, it was not noticed by enough participants to be considered relevant). This does not mean that these secondary channels are not important to the emotion expression, conversely, it might signify that, while they are important, other features are more important in the setting of the experiment.

Then we asked participants ($N = 85$) to rate each animation against the constructs. Afterward, we performed a PCA and a Cluster Analysis to create our model. Our PCA yielded 3 components: **(C1)** represents a more relaxed stance and the expression of *Confidence* (its negative captures the expression of *Defensiveness* and *Rage*); **(C2)** represents a shaky posture and the expression of *Fear*; and, **(C3)** represents a defensive stance and the expression of *Defensiveness* (its negative captures the expression of *Rage*).

In the Cluster Analysis, we concluded that 4 clusters were ideal for our study. These clusters have a one-to-one mapping to each emotion. However, this mapping is not perfect. Negative emotions are more easily scattered in different clusters, where *Defensiveness* is the emotion that is less powerful in its presence. Overall, we can say that each emotion is perceived correctly and independently from the others. This confirms that each emotion is being perceived correctly.

Having the latter confirmation, we looked back at the PCA and tried to understand how each emotion is perceived, through the use of a clustered means of each component per emotion. We found many similarities with the Cluster Analysis, but also some differences, namely on *Defensiveness*, which is the most confused emotion of the four. With all these analyses, here follows a description of each emotion by how they are perceived:

Confidence has an open, relaxed, and unprepared posture, but is also grounded. The character has a cheerful facial expression. The head is tilted up and the character is leaning upright. The weapon is held high and back. (Positive Component 1; Negative Component 2; Cluster α)

Defensiveness has a closed, ready-to-react, and prepared posture. The character has a grimacing facial expression. The head is tilted down and the character is leaning forward. The weapon is held in front. The character takes a defensive stance and a slower walk speed[27]. (Negative Component 1; Positive Component 3; Cluster γ)

Fear has shaky posture. The weapon is held low. The character takes a defensive stance and a slower walking speed. (Positive Component 2; Positive Component 3; Cluster β)

Rage has a closed, ready-to-react, and prepared posture. The character has a grimacing facial expression. The head is tilted down and the character is leaning forward. The weapon is held in front. The character takes an offensive stance and a faster walking speed. (Negative Component 1; Negative Component 3; Cluster δ)

With these results, we have a clearer understanding of user perception and how each emotion is perceived. Furthermore, while improvements can be made, namely to the *Defensiveness* expression, generally the different emotion expressions are being correctly and independently perceived by the participants. We decided to move forward with the current expressions of emotion as previously presented.

5.4 Second Evaluation on Believability

Having our improvements to the emotion expression of the characters (validated in the previous Section 5.3), we looked into improving the computation of the emotions. We previously introduced an emotional model with anticipatory capabilities (see Section 5.2), now, given our prior knowledge of the use of *Emotivector* [7] (see Section 2.5.2 and Section 4.3.1), we bring a new system to compute emotions, the *BALTAS Model*.

Similar to the previous evaluation on Believability (see Section 5.2), to understand the impact of the expression improvements and of the new emotional model on the believability of the experience, we performed a new evaluation. This evaluation will help us to understand whether our solution validates our hypothesis, that is, that *by subdividing actions according to the traditional principles of animation* (implemented in *3motion*) and *allowing for emotional reactions and interruption of behavior during their expression* (implemented in the BALTAS system), *the overall real-time interaction will be perceived as more believable*.

[27] Note that the only animations with a different walk speed are those expressing *Rage*. The remaining animations follow the same speed cycle.

To perform this evaluation, we have three versions of the game, two versions from the first evaluation, the *Reactive Model* and the *Predictive with Assertion Model*, and a final version where we introduce the new emotion system, the *BALTAS Model*. The first version, as in the previous evaluation, gives us a system similar to those used in game development with more reactive behavior. The second version helps us to understand how it performs when compared with the last version, but also to understand if changes in the emotion expression system, introduced in the user perception study, have an impact on believability. The third version proposes a new way of computing emotions in this environment, which we hope will improve believability further. We made use of the same scenario setting with the content-free environment (see Figure 5.4). To have participants rate each version, we employ the Believability Questionnaire, which was used previously to help us measure believability in *3motion* (see Section 3.5.1.A).

5.4.1 More Expressive HUD

Other than the mentioned changes, we decided to improve the Heads-up Display (HUD)[28] displaying the health of the characters. The previous design provided only information on health and was not symmetrical, in that one health bar was displayed in the top left corner and the other in the bottom right corner. Furthermore, due to the third-person perspective used to control the duelists, which obstructs the player's character's facial expressions, and the lack of a method to display the *background iconography* of the player's character (this issue is described in Section 5.3.1.D), we decided to rework the HUD to have elements that help showcase the character's expressions during gameplay.

Figure 5.16: Screenshot of the redesigned HUD displaying the character on the left with more health and showing *Confidence*, and the character on the right with less health and showing *Defensiveness*. The design is symmetric and has a small banner between the two health bars for visual flair.

In Section 2.6.2, we discussed how some games make use of HUD elements to showcase character's expressions (see Figure 2.12). Taking these references as inspiration, we decided to use the previously implemented solutions seen in Section 5.3.1 and reuse them into the HUD. We created a frame where we display a still image of the character. We added a background to the frame that changed colors based on emotion. The colors are the ones used in the *Colored Outlines*. Over the still image of the character, we display the facial expression using the same images as previously defined. Similarly, the background iconography is also portrayed above the head of the character. Figure 5.16 displays the new HUD with each character portraying a different emotion. We also made the design symmetric.

[28]In video games, the heads-up display is the method by which information is visually relayed to the player as part of a game's user interface.

5.4.2 The BALTAS Model

Codenamed *BALTAS* (Belief and Anticipation Laced Tenacious Affect System), this react model for *3motion* (see Section 3.2.2) was implemented to make use of the previously described *Emotivectors* [7] (see Section 2.5.2) with the lessons learned from our previous implementation on the Virtual Coach application (see Section 4.3.1). Therefore, this module integrates the *Emotivector* concept into the *3motion* model.

This model updates frequently (every frame) and checks if its *Emotivectors* (can be updated every frame or by callbacks) have new salient sensations (the mismatch between perceived values and expected values). These *Sensations* are then updated, and their lifetime (computed in seconds) is deduced. When a Sensation's lifetime is zero (0) it no longer has an impact on the character and can be discarded. For each active Sensation, an emotion is computed based on the character's personality. From this set of emotions and sensations, a single emotion is selected (depending on the selection method).

5.4.2.A Generic Emotivector Implementation

Looking back (see Section 2.5.2), we know that an emotivector is an "anticipatory mechanism attached to a sensor". In essence, it computes predictions of future events based on past history and creates sensations by comparing received events with its expectation.

Similarly to the solution implemented in the first evaluation on Believability (see Section 5.2), we created predictors based on averages and their derivatives. For this use case, we decided to make use of exponential moving averages (instead of weighted moving averages), computed as follows:

$$EMA_i = (x_i * k) + (EMA_{i-1} * (1 - k)) \quad \text{where} \quad k = s/n + 1 \tag{5.6}$$

EMA_i denotes the calculation of the average for the element i in the history of values. x_i is the value of element i. n is the number of elements in the history of values. We use a smoothing factor, s, of 2. Also, consider $EMA_1 = x_1$. This equation exponentially decreases the weight of an element in the history depending on the smoothing factor, meaning that the more recent an element is, the more it will contribute to the prediction. From empirical testing, this solution was the one that yielded the best results.

Unlike the previous study, but as suggested in the original work [7], we implemented a Meta predictor that, while also using an exponential moving average, measures the error in the prediction of values (the mismatch between the expected value and the received value). This allows us to filter results that are not impactful enough (if the prediction error is within the expected, then the sensation is not impactful).

The Emotivector outputs a Sensation. This sensation describes the mismatch between perceived values and expected values. Therefore, this mismatch can be a *Reward* or a *Punishment* and be *greater / worse than expected* (since the "as expected" values are filtered out by the Meta predictor). The sensations also store the computed salience that can be only the exogenous value or also the endogenous value if a "desired" value is defined (both values are computed as in [7]), along with a lifetime in seconds to determine how long it should be active (this value defaults to 5 seconds, can be programmatically changed, and is not influenced by the salience).

5.4.2.B Mapping an Emotion to a Sensation

How then do we compute an emotion from a sensation? Given the variety of sensations with different saliences, we decided on a direct mapping between a Sensation of an Emotivector and an Emotion. We define this mapping as a Personality that can change between characters (this is similar to what is seen in the previous implementation with the mapping between Health Ratio and Emotion, see Section 5.2). For example, defensiveness is assigned to a sensation of *Punishment Better than Expected* with a presence greater than 0.2. See an example of a personality in Table 5.7.

We experimented with having the salience of each sensation decay over time, but empirical trials favored having a static value of salience but with a limited lifetime.

Table 5.7: Example of a BALTAS personality. Note how multiple emotions can share the same sensation but have different salience intervals.

Affective State	Sensation	Salience interval
Fear	Punishment Worse than Expected	$s \in [0, 1]$
Rage	Reward Worse than Expected	$s \in [0.3, 1]$
Defensiveness	Punishment Better than Expected	$s \in [0.2, 1]$
Confidence	Reward Better than Expected	$s \in [0, 1]$

5.4.2.C Choosing the Emotion

Having multiple pairs of Sensations and Emotions, we need to select a final emotion to express. We implemented multiple selection methods:

- **Most Salient** – A Winner-takes-all approach. The emotion of the most salient sensation is selected.
- **Most Common** – A democratic approach. The emotion most frequently seen among all sensations is selected. This solution discards the salience values.
- **Most Common then Most Salient** – An encapsulation of the two previous ones. The emotion most frequently seen among all sensations is selected. If in a tie, the most salient is selected.
- **Most Powerful** – Selects the emotion with the highest accumulated salience. Multiple sensations that compute the same emotion will accumulate their salience in the emotion. This solution offered a good balance in fairness, making use of the salience value while being democratic by design, and was the solution we choose for the evaluations.

5.4.2.D Game-specific Emotivectors

Based on the generic implementation of the Emotivectors, we created two to compute the reactions of the characters to the world around them, in this case, the combat experience. Both focus on the health of the characters, looking at either one's own health or that of the opponent[29]. Note that the range of values was normalized to a range between 0 and 1, and we made it so that values near 1 are "good" values (meaning that they can be interpreted as positive to the character) and values near 0 are "bad" values.

A – Own Health Emotivector – This emotivector collects the changes in the health of the character (that is, an assessment of the self). Values range from full health at 1 to low health near 0. When certain combat-related events occur, the current value of health is added to the emotivector and a prediction is computed. These events include: being hit by the opponent; a miss by the opponent; and, a successful block of an opponent's attack. This solution prevents a continuous collection of health data that does not change significantly. *This allows the character to anticipate if they are close to defeat or still have a lot to give.*

To better show the evolution of values over time in this emotivector, we plotted a graph in Figure 5.17 that shows the perceived values and the predictions performed. In this example (collected from an actual match) the character is hit 3 times (at time steps 2, 7, and 8); coincidentally, these strikes break expectations, as the perceived values are outside of the prediction error (the error bars size is based on the predictions from the Meta predictor), and thus generate a salient sensation. Note also how between time steps 3 and 6 the perceived values are within the prediction error; therefore, no sensation is generated.

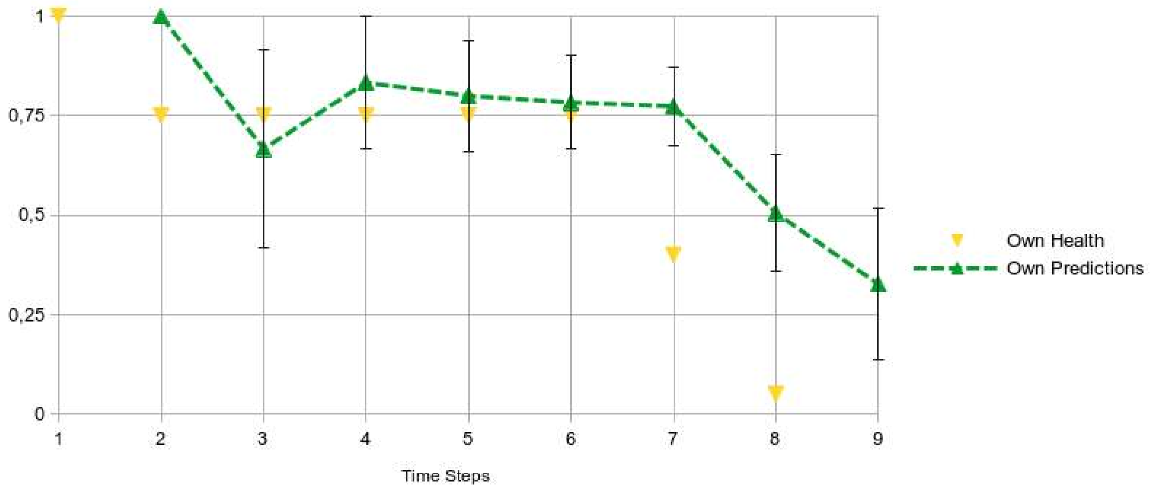


Figure 5.17: Example of Own Health Emotivector values and predictions over time. The error bars size is based on the prediction error expectations from the Meta predictor.

Until this point, the generated sensations, while anticipatory in nature, will happen in reaction to the events. To add more anticipatory behavior (i.e., reactions during the anticipatory stages of an

[29]We explored other solutions focusing on the attack success rate, but these were discarded in favor of the ones shown.

action), we added rule-based behaviors, as seen in many video games, but we devised the rules to be based on the anticipatory data in the emotivector.

We do this by checking if the opponent is starting an attack (in the interruptible anticipation stage), and if the character *believes* that the attack will hit or miss, then a new sensation is computed. How then do we determine that a character "believes" something? We determine this belief by looking at the latest prediction and value in the *Emotivector*. Given the range of the prediction error (from the Meta predictor), we check if the latest value is outside of that range. If so, it means that the character believes that a change is going to happen.

As an example, in Figure 5.17, in time step 8, if the opponent starts an attack, the latest value (seen at time step 7) is within the range of the prediction and its prediction error, so no sensation is computed. However, in time step 9, the latest value (seen in time step 8) is outside of the range of the prediction and its prediction error, and the prediction value is higher than the latest value, meaning that the character believes that the next attack will fail (showing an optimistic view). Therefore, at the start of the opponent's attack, a sensation is generated.

We devised a small set of rules to create more varied reactions to anticipatory behavior. Here follows a description of each rule.

The character believes that the opponent will be able to hit it. – This rule is active when the predicted value is lower than the previous value. It creates a sensation with *Punishment Worse than Expected*. The salience of this sensation is the exogenous value between the prediction and the previous value.

The character believes that the opponent will not be able to hit it (miss or block). – This rule is active when the predicted value is greater than the previous value. It creates a sensation with *Reward Better than Expected*. The salience of this sensation is the exogenous value between the prediction and the previous value.

The character believes that the next attack will kill it. – This rule is active when the predicted value minus the prediction error is lower than 0. It creates a sensation with *Punishment Worse than Expected*. The salience of this sensation is the exogenous value between the prediction and the previous value.

Furthermore, we allow the character to emotionally react to the opponent's expressed emotion. This additional **Opponent's Emotion Reaction** rule is also unlocked during the anticipation stage of an attack and if the opponent is feeling a positive emotion (i.e., *Confidence*) the character reacts negatively, creating a sensation with *Punishment Worse than Expected*. The salience of this sensation is the exogenous value between the prediction and the previous value.

Each of these rules has an associated importance value that can be manipulated to diversify the personality of each character. Sensations generated have their salience multiplied by this importance. We expect this addition to improve the expression of the characters.

B – Opponent's Health Emotivector – Similar to the previous one, this emotivector collects the changes in the health of the opposing character. The collected values are inverted, meaning that if the opponent is in full health, the recorded value is a 0, while if the health is low, it will be closer to 1. When certain combat-related events occur, the current value of health is added to the emotivector and a prediction is computed. These events include: a hit on the opponent; a miss on the opponent; and, a block by the opponent. *This allows the character to anticipate if they are close to a victory or not.*

To better show the evolution of values over time in this emotivector, we plotted a graph in Figure 5.18 that shows the perceived values and the predictions performed. Contrary to the previous example, in this example (collected from the same match as the previous one), the opponent is hit more regularly (eventually losing); therefore, the predictions tend to accompany this movement, as seen in time steps 3 and 4. There are, however, still moments of surprise, for example, time step 5, where the prediction expected a miss, but the opponent was hit.

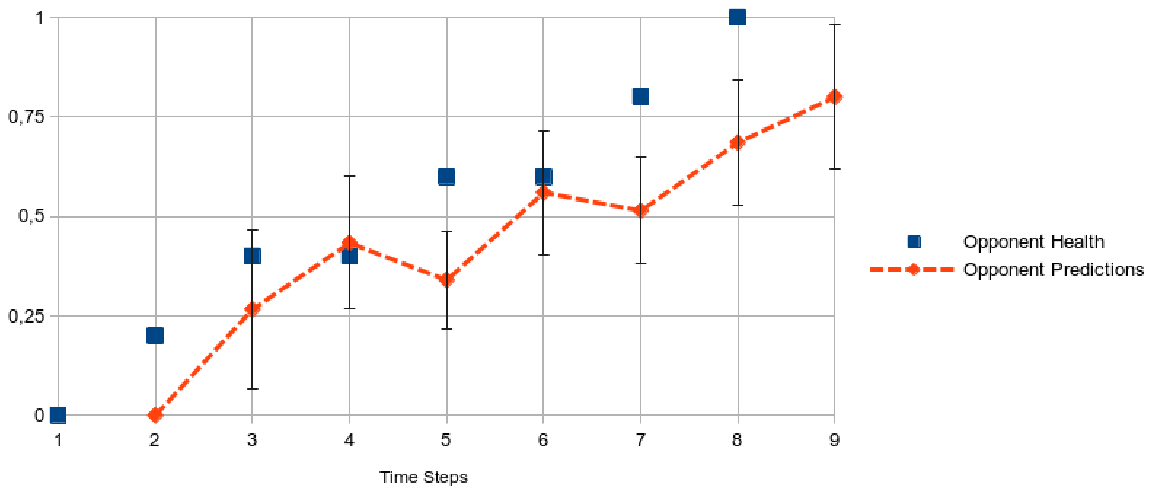


Figure 5.18: Example of Opponent's Health Emotivector values and predictions over time. The error bars size is based on the prediction error expectations from the Meta predictor.

Keeping the similarities with the previous emotivector, we also added rule-based behaviors. Here, we focus on the ability of the character to hit its opponent. Therefore, we check if the character is starting an attack (in the interruptible anticipation stage) and if it *believes* that the attack will hit or miss, to then compute a new sensation. The belief computation is the same as previously described.

As an example for this specific *Emotivector*, in Figure 5.18, at time step 8, if the character starts an attack, the latest value (seen at time step 7) is within the range of the prediction and its prediction error, so no sensation is computed. However, in time step 9, the latest value (seen in time step 8) is outside of the range of the prediction and its prediction error, and the prediction value is lower than that of the latest value, meaning that the character believes that its next attack will miss (showing a pessimist view). Therefore, at the beginning of its attack, a sensation is generated.

We devised a small set of rules to create more varied reactions to anticipatory behavior. Here follows a description of each rule.

The character believes that it will be able to hit the opponent. – This rule is active when the predicted value exceeds the previously seen value. It creates a sensation with *Reward Better than*

Expected. The salience of this sensation is the exogenous value between the prediction and the previous value.

The character believes that it will not be able to hit the opponent (miss or block). – This rule is active when the predicted value is lower than the previous value. It creates a sensation with *Punishment Worse than Expected*. The salience of this sensation is the exogenous value between the prediction and the previous value.

The character believes that the next attack will kill the opponent. – This rule is active when the predicted value plus the prediction error exceeds 1. It creates a sensation with *Reward Better than Expected*. The salience of this sensation is the exogenous value between the prediction and the previous value.

Here, we also allow the character to emotionally react to the opponent's expressed emotion. This **Opponent's Emotion Reaction** rule is active if the opponent is feeling a negative emotion (i.e., *Rage*, *Fear*, or *Defensiveness*) and the character reacts positively, creating a sensation with *Reward Better than Expected*. The salience of this sensation is the exogenous value between the prediction and the previous value.

These rules mimic many of what was described in the previous *Emotivector*, but with a different perspective, which we hope will further favor the believability of the interaction.

C – More than the Sum of its Parts – As a way of better understanding the defined *Emotivectors*, we created Figure 5.19. This figure showcases the evolution of both emotivectors over the duration of a combat from the perspective of one character (the data was collected from actual gameplay and these values are the combination of both Figures 5.17 and 5.18). The time steps mark the different events throughout the interaction; for example, between time steps 4 and 6, there was an attack by the opponent that either missed or was blocked, thus the health did not change, but in time step 5, the character also finished an attack.

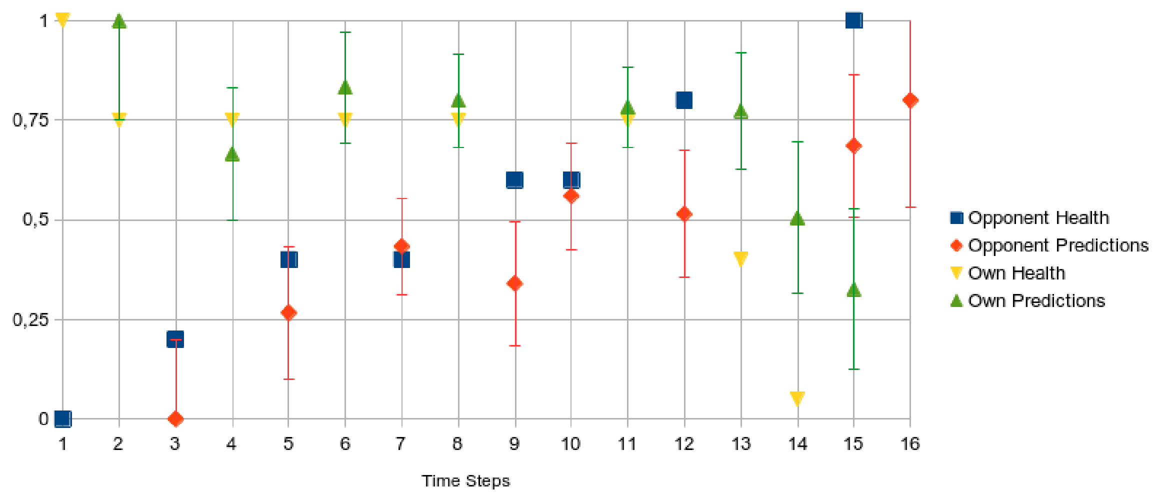


Figure 5.19: Showcase of the evolution of values and predictions of both health-related emotivectors over time. The error bars size is based on the prediction error expectations from the Meta predictor. The values in yellow refer to the values in the *Own Health Emotivector* with their prediction in green. The values in blue refer to the values in the *Opponent's Health Emotivector* with their prediction in red.

In the figure, we can see that the first computed sensation was at time step 9, where the character expected to miss an attack, but there was a hit. Thus a *Reward Better than Expected* sensation is created. Then, at time step 12, a similar sensation is created. However, in the following time step (13), the character is struck by the opponent, and its expectations are broken. A *Reward Worse than Expected* sensation is generated. A similar sensation is computed at time step 14. Finally, at time step 15, a *Reward Better than Expected* sensation is computed after a successful hit on the opponent. Note that, at that time step, when the character started an attack, the rule *The character believes that the opponent will not be able to hit it* was active, and thus an additional sensation *Reward Better than Expected* is computed.

Although this example does not explore all possible emotions, it shows the expressive potential of the system.

5.4.2.E Personalities in the Evaluation

Before moving on to the evaluation, it is important to discuss the personalities created for the characters in it. The personality is defined by two sets of parameters, the mapping of sensation to affective state, and the importance of each rule and emotivector. We created a more aggressive personality for the character wielding an axe, and a more defensive personality for the character wielding a sword and shield.

In the following, we showcase the modifications done to implement the personalities. Not all rules are displayed. Assume that the importance values default to 1 when not mentioned.

A – Axe Character – To implement an aggressive personality, we promoted the use of the *Rage* expression and focused our attention on the opponent's health. The former was done by adding the expression of *Rage* in the character's personality when a sensation of *Reward Worse than Expected* occurs, but also when a sensation of *Punishment Better than Expected* occurs. Table 5.8 shows the personality used for this character.

Table 5.8: BALTAS personality used on the character wielding an axe.

Affective State	Sensation	Salience interval
Fear	Punishment Worse than Expected	$s \in [0, 1]$
Rage	Reward Worse than Expected	$s \in [0, 1]$
Rage	Punishment Better than Expected	$s \in [0.1, 1]$
Defensiveness	Punishment Better than Expected	$s \in [0, 0.1]$
Confidence	Reward Better than Expected	$s \in [0, 1]$

As for the *Emotivectors* importance, we wanted to focus on the opponent's health more than the character's own. So, instead of boosting the *Opponent's Health Emotivector*'s importance and its rules, we decided to decrease the importance of *Own Health Emotivector*. Here are the updated values for this personality:

– Importance of *The character believes that the opponent will be able to hit it*: **0.5**

- Importance of *The character believes that the opponent will not be able to hit it*: **0.5**
- Importance of *The character believes the next attack will kill it*: **0.8**
- Importance of *Own Health Emotivector*: **0.9**

B – Sword and Shield Character – To implement a defensive personality, we promoted the use of the *Defensiveness* expression and focused our attention on the character's health. The former was done by adding the expression of *Defensiveness* in the character's personality when a sensation of *Punishment Better than Expected* occurs, but also when a sensation of *Reward Worse than Expected* occurs. Table 5.8 shows the personality used for this character.

Table 5.9: BALTAS personality used on the character wielding a sword.

Affective State	Sensation	Salience interval
Fear	Punishment Worse than Expected	$s \in [0, 1]$
Defensiveness	Reward Worse than Expected	$s \in [0, 0.2]$
Rage	Reward Worse than Expected	$s \in [0.2, 1]$
Defensiveness	Punishment Better than Expected	$s \in [0, 1]$
Confidence	Reward Better than Expected	$s \in [0, 1]$

As for the *Emotivectors* importance, we wanted to focus on the character's own health more than the opponent's. So, instead of boosting the *Own Health Emotivector*'s importance and its rules, we decided to decrease the importance of *Opponent's Health Emotivector*. Here are the updated values for this personality:

- Importance of *The character believes that it will be able to hit the opponent*: **0.5**
- Importance of *The character believes that it will not be able to hit the opponent*: **0.5**
- Importance of *The character believes the next attack will kill the opponent*: **0.8**
- Importance of *Opponent's Health Emotivector*: **0.9**

These changes still allow the expression of all the emotions, but promote the use of specific ones. The selection of these configurations was based on empirical testing and offers an example of character personalization.

5.4.2.F Previous Models Personalities

Previously, in the *Reactive Model* and *Predictive with Assertion Model*, we defined personality as a mapping between the affective state and the health ratio interval (current or anticipated). Because we are reusing these models and comparing their personalities with those of BALTAS, it is important to showcase the personalities defined for the use case. For consistency, we used the previously defined personalities, which are the opposite of those of BALTAS. The Axe is more defensive, while the Sword and Shield character is more aggressive.

A – Axe Character – For this personality, we decided to promote the *Defensiveness* expression, raising its health ratio interval, while reducing the *Rage* expression, making it close to a low value of the health ratio and shortening the interval. Table 5.10 shows the personality used for this character.

Table 5.10: Previous models personality for the Axe Character.

Affective State	Health ratio interval
Fear	$w \in [0, 0.1]$
Rage	$w \in [0.1, 0.2]$
Defensiveness	$w \in [0.2, 0.3]$
Confidence	$w \in [0.7, 1]$

B – Sword and Shield Character – For this personality, opposite to the previous, we decided to promote the *Rage* expression, expanding and raising its health ratio interval, while reducing the *Defensiveness* expression, making it close to a low value of the health ratio and shortening the interval. Table 5.11 shows the personality used for this character. Recall that the health ratio starts at 0.5, as characters start with full health. Therefore, an interval reaching 0.45 will be more active with small variations of the health ratio.

Table 5.11: Previous models personality for the Sword and Shield Character.

Affective State	Health ratio interval
Fear	$w \in [0, 0.15]$
Rage	$w \in [0.3, 0.45]$
Defensiveness	$w \in [0.15, 0.3]$
Confidence	$w \in [0.7, 1]$

5.4.3 Experiment

Having discussed our latest implementation, we want to assert which of our emotion models, the new *BALTAS model*, the *Reactive Model*, or the *Predictive with Assertion Model* (see Section 5.2.2), create more believable interactions. To do this, we prepared an evaluation using the Believability Questionnaire (see it in full in Appendix D.3), previously used in *3motion* (see Section 3.5.1.A) and a version of *Adfectus* containing the three emotion models created. Given the competitive nature of *Adfectus*, the evaluations were always carried out with pairs of participants. Each participant would complete their own questionnaire.

To start the evaluation, participants were asked to answer a series of demographic questions, including their age, gender, the frequency at which they play video games, if they were familiar with and liked fighting games, if they regularly interacted with synthetic characters, and how they valued emotion expression in games. Afterward, to become used to the controls and gameplay, the participants played a version of the game that did not make use of any emotional model (i.e. the characters did not express any emotion).

Once participants were comfortable with the game, they would fill out two questions about their liking of the game so far and how easy the characters were to control. Afterward, they would move to the next part of the experiment where we ask them explicitly to pay attention to the characters' animations, namely when and how they change. Such a confirmation step is required, because, as shown in other similar works [170] and as we noticed in preliminary evaluations, some animation transformations that occur throughout the game can easily go unnoticed by players.

In this step of the experiment, we also presented the set of questions related to the believability of the interactions the participants would have to answer after play, once again to make sure that participants looked at exactly what we wanted so that they could answer the questions appropriately. This initial presentation of the main questionnaire also included examples of what we were referring to in the statements to make sure that they clearly understood what we were referring to. These examples were removed from the actual main questionnaires to avoid biasing the answers.

After this step, participants were asked to play three separate versions of the game. Each with an associated emotion model: *(A) reactive model*, *(B) predictive with assertion model*, and *(C) BALTAS model*. To avoid any order bias, we used a *Latin square* design to order the versions, creating an $n \times n$ array filled with n different symbols, each associated with one version and each occurring exactly once in each row and exactly once in each column. As such, our 3 versions were presented in different orders in a way that no version occurs in the same spot in more than one combination, reducing the potential bias mentioned before.

After playing a version, the participants were asked to rate the following set of statements relating to the believability of the interactions experienced during the gameplay:

Q1 - I understood what the characters were doing.

Q2 - I could predict the characters' actions.

Q3 - I understood what the characters were feeling.

Q4 - I could predict the characters' feelings.

Q5 - I understood the characters' intentions.

Q6 - The characters were aware of each other.

Q7 - The characters were aware of each other's actions.

Q8 - The characters could predict each others' actions.

Q9 - The characters were aware of each other's feelings.

Q10 - The characters could predict each others' feelings.

Q11 - The characters were aware of each other's intentions.

Q12 - The interaction between characters in this scene was believable.

Q13 - The characters displayed different personalities.

Each statement was rated on a 7-point Likert scale, ranging from 1 - *Strongly disagree* to 7 - *Strongly agree*. As described in Section 3.5.1.A, from **Q1** to **Q5**, the questions relate to the interpretation of character behavior from the perspective of the participant; from **Q6** to **Q11**, the questions relate to the interpretation of character behavior from the perspective of the characters; **Q12** relates to the general believability of the interaction; and, **Q13** is a new addition to better understand if the participants could distinguish personalities between the characters. After this question, two free-form questions asked participants to provide one or two adjectives to describe the personality of each character (if they found one).

Additionally, for us to check if the participants were actively aware of both characters' expressions, before the aforementioned questions, we added a checkbox question where participants were asked to check if they were aware of the expressions of the character they were controlling and the one their opponent was controlling.

At the end of the experiment, participants could also leave free-form comments on the experiment. Furthermore, a small informal talk was held with most participants, where they were asked what version they preferred and why.

5.4.4 Results

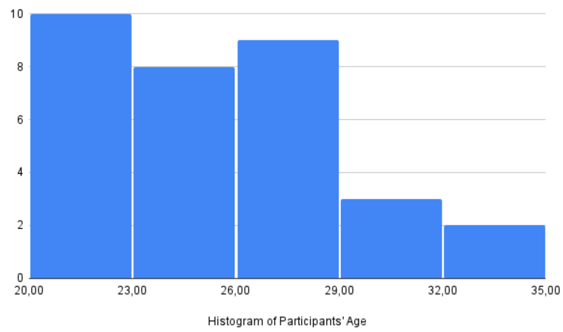


Figure 5.20: Histogram that shows the distribution of the participant's ages.

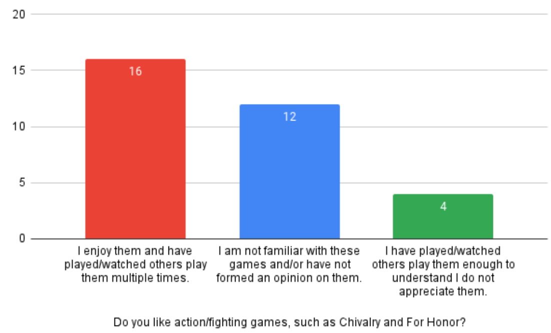


Figure 5.21: Distribution of each case for the question "Do you like action/fighting games, such as Chivalry and For Honor?"

Our group of participants ($N = 32$), collected by convenience sampling, was between 20 and 33 years old ($M = 24$, $SD = 3.542$; see Figure 5.20). Eight were women, 22 were men, and two preferred not to identify their gender. Regarding the participants' habits with video games, 50.0% reported making time to play video games, while 37.0% played when the opportunity presented itself, and the rest rarely played. Regarding action/fighting games, 50.0% of the participants enjoy playing or watching this genre of games, 37.5% was not familiar with this genre, and the remaining 12.5% did not enjoy this genre of games (see Figure 5.21). Furthermore, 65.6% of the participants rarely or never played these types of games, and only 2 participants reported actively scheduling time to play these types of games. Regarding the experience with synthetic characters, 75% interact regularly with synthetic characters (4 of which work with these types of characters). Furthermore, regarding the expression of emotions by characters, most participants (90.6%) prefer characters that express emotions, but only 21.9% consider it crucial for their enjoyment.

A Shapiro-Wilk test determined that only a few of the data sets of the statements were normally distributed, thus we decided to use non-parametric tests. To understand whether there was a significant difference between versions, for each statement we performed a Friedman test; furthermore, where appropriate, a post hoc analysis with Wilcoxon signed-rank tests was performed with a Bonferroni correction applied, resulting in a significance level set at $p < 0.017$. A detailed view of each statement is presented in Appendix D.3.2. Figure 5.22 shows the averages and standard deviations of each statement for each version.

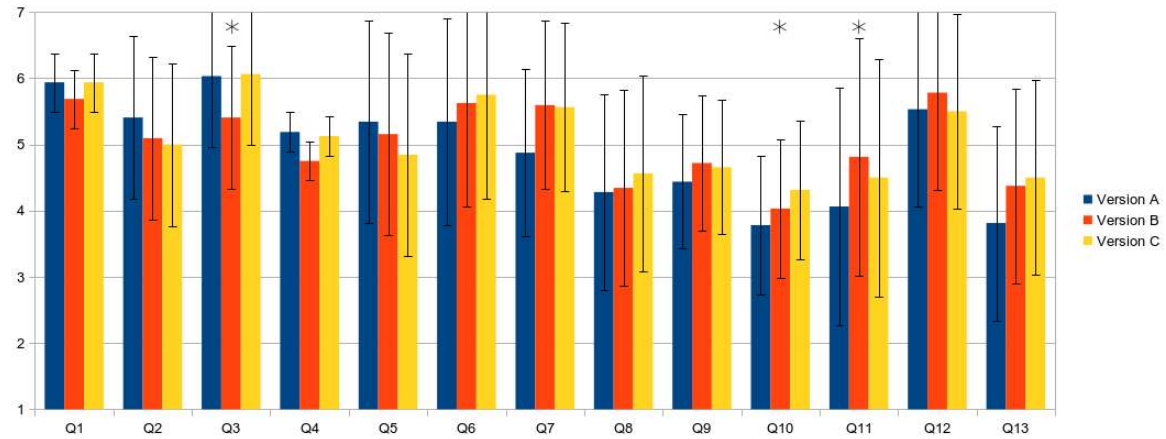


Figure 5.22: Averages and associated standard deviations of each statement for each version. An asterisk (*) marks questions in which we found statistically significant differences in the Friedman test.

We concluded that neither the *(B) predictive with assertion model* nor the *(C) BALTAS model* appeared to improve the believability of character interactions *(Q12)* compared to the more traditional purely *(A) reactive model*, $\chi^2(2) = 1.724$, $p = 0.422$. Where version *(C)* seemed to score the lowest among the three (5.50, $SD = 1.566$), followed by *(A)* (5.53, $SD = 1.319$), and version *(B)* the highest (5.78, $SD = 1.338$). Nevertheless, most participants rated the interactions as believable regardless of version, suggesting a positive experience even with only the reactive model. We believe that these high values are due to the emotion expression channels implemented. A quick comparison of the believability ratings between this study and the first believability evaluation shows that the values of version *(A)* increased from $M = 5.10$ ($SD = 1.701$; recomputed by linear conversion from a 9-point to a 7-point Likert scale) to 5.53 ($SD = 1.319$), and the values of version *(B)* increased from $M = 5.59$ ($SD = 1.480$) to 5.78 ($SD = 1.338$). This seems to suggest an overall improvement in believability, regardless of the emotion model used.

We did not find statistically significant results on most questions, the exceptions being in Q3, Q10, and Q11, which we discuss further below. Yet, most of the questions pointed toward a good perception of the character's actions and feelings, either from the point of view of the participant (Q1 - Q5) or the point of view of the characters (Q6 - Q11). However, in questions Q8 and Q9, in general, participants appeared to be unable to distinguish whether the characters could predict each other's actions (Q8) or if they were aware of each other's feelings (Q9), there was, however, a small positive trend regardless of the version. For Q8, version A had a mean of 4.28 ($SD = 1.631$); version B had a mean of 4.34 ($SD = 1.578$); and version C had a mean of 4.56 ($SD = 1.564$). For Q9, version A had a mean of 4.44 ($SD = 1.605$); version B had a mean of 4.72 ($SD = 1.591$); and version C had a mean of 4.66 ($SD = 1.473$).

Let us now focus on where we found differences.

Q3 - I understood what the characters were feeling. For this statement, version A had a mean of 6.03 ($SD = 1.470$); version B had a mean of 5.41 ($SD = 1.478$); and version C had a mean of 6.06 ($SD = 1.014$). We found statistically significant differences between the versions ($\chi^2(2) = 6.179$, $p = 0.046$). Thus, we performed the post hoc analysis. No significant differences were found between any pair: between versions A and B, $Z = -2.156, p = 0.031$; between versions C and B, $Z = -2.244, p = 0.025$; and between versions A and C, $Z = -0.053, p = 0.958$.

So, while a statistical difference was observed in the Friedman test, we found no major differences between pairs. We can still see that Version C was higher in mean value, followed by Version A, and Version B was the lowest in rank. Versions A and C are also very similar in values, as seen by the low difference in the Wilcoxon test ($Z = -0.053, p = 0.958$). If we wish to take meaning from these values, then in all versions the characters' feelings are understood, and in versions A and C the characters' feelings are slightly better understood than in version B.

Q10 - The characters could predict each others' feelings. For this statement, Version A had a mean of 3.78 ($SD = 1.699$); Version B had a mean of 4.03 ($SD = 1.534$); and Version C had a mean of 4.31 ($SD = 1.378$). We found statistically significant differences between the versions ($\chi^2(2) = 6.660$, $p = 0.036$). Thus, we performed the post hoc analysis. No significant differences were found between any pair: between versions A and B, $Z = -0.947, p = 0.344$; between versions C and B, $Z = -1.554, p = 0.120$; and between versions A and C, $Z = -1.434, p = 0.152$.

So, while a statistical difference was observed in the Friedman test, we found no major differences between pairs. We can still see that Version C was higher in mean value, followed by Version B, and Version C was the lowest in rank. If we wish to take meaning from these values, then in all versions participants seem to not be able to agree or disagree if characters are able to predict each other's feelings. Yet, for version A, there is a tendency for characters to be unable to perform a prediction, while in version C there is the opposite tendency, where characters seem more capable of predicting the opponent's feelings.

Q11 - The characters were aware of each other's intentions. For this statement, version A had a mean of 4.06 ($SD = 1.625$); version B had a mean of 4.81 ($SD = 1.230$); and version C had a mean of 4.50 ($SD = 1.437$). We found no statistically significant differences between the versions ($\chi^2(2) = 7.814$, $p = 0.020$). We found statistically significant differences between the versions ($\chi^2(2) = 6.660$, $p = 0.036$). Thus, we performed the post hoc analysis. No significant differences were found between versions C and B ($Z = -1.201, p = 0.230$) and between versions A and C ($Z = -1.496, p = 0.135$). We did find statistically significant differences between versions A and B, $Z = -2.811, p = 0.005$;

In this statement, Version B was ranked higher, followed by Version C, and Version A was ranked lower. Here, we can see that version B is significantly different and greater than A. However, no major differences were found between the remaining pairs. Looking at these results, we find that Version B is where the characters were perceived to be more aware of each other's intentions. Version C is a bit behind in score but is almost as good. For version A, it was unclear to the participants if the characters were aware or not of each other's intentions.

Q13 - The characters displayed different personalities. A topic yet to be discussed is that of the personalities presented. For this statement, version A had a mean of 3.81 ($SD = 2.132$); version B had a mean of 4.38 ($SD = 2.121$); and version C had a mean of 4.50 ($SD = 2.110$). We found no statistically significant differences between the versions ($\chi^2(2) = 2.742$, $p = 0.254$). Overall, the participants seemed to be unable to distinguish if the characters show or not different personalities. There is a positive tendency for versions B and C, while there is a negative tendency for version A.

Regarding the descriptions given by the participants on each character's personality, version A had the smallest number of personality characteristics given ($N = 15$ for both the axe and the sword characters). Versions B and C had the same number of personality characteristics given ($N = 21$ for the axe character and $N = 20$ for the sword character). With these characteristics, we performed a content analysis to better understand the differences in personalities. Recall that for versions A and B we expected the Axe character to be more defensive and the Sword character more aggressive, while for version C the Axe character would be more aggressive and the Sword character more defensive. In Table 5.12 we show the more frequent terms used to describe the personalities of each character in each version. Items with fewer than 2 occurrences were removed.

Table 5.12: Most frequent terms used to describe the personalities of each character in each version.

Version A				Version B				Version C			
Axe	f	**Sword**	f	**Axe**	f	**Sword**	f	**Axe**	f	**Sword**	f
Confident	8	*Aggressive*	6	*Confident*	12	*Aggressive*	3	*Confident*	12	*Scared*	11
Aggressive	3	*Confident*	4	*Aggressive*	3	*Scared*	3	*Aggressive*	9	*Confident*	4
Strong	2	*Scared*	3	Expressive	2	*Confident*	2	*Scared*	3	*Aggressive*	3
		Weak	3			level-headed	2				
						Neutral	2				

As we can see, regardless of the version, the Axe character is always perceived as *Confident* (note that participants reported that this weapon was stronger than the sword, which can have an impact on the gameplay and thus the expressed emotions). Putting this expression aside, for versions A and B, there is no clear distinction between the characters, as they are both perceived as *Aggressive*, however, in version A, the Sword character is perceived more times as *Aggressive* but is also viewed as *Scared* in versions A and B. The gameplay appears to have a bigger impact on the perception of the characters than the personality in these models (*Reactive* and *Predictive*).

However, in version C, there appear to be some differences in the perception of the characters. While both are perceived as *Aggressive* and *Scared*, the Axe character seems to be more often

Aggressive and the Sword character more frequently *Scared*. This aligns with how we designed the personalities. These results seem to suggest that in many cases participants will not assign personalities to the characters, but when they do, in version C, the personalities given align with those designed.

Questions on Participants' awareness of the expressions of the characters. We found that the participants were more aware of the expressions performed by the opposing character than those of the character they controlled, which is expected due to the third-person camera perspective used. Participants were aware of the expressions performed by the opposing character in 91% of the games played in version A ($SD = 0.296$); 88% in version B ($SD = 0.336$); and 97% in version C ($SD = 0.177$). Contrary to the reported awareness of the expressions of the controlled character of 75% of games played in version A ($SD = 0.296$); 69% in version B ($SD = 0.471$); and 69% in version C ($SD = 0.471$). We did not find statistically significant differences between the versions for the expression of the controlled character ($\chi^2(2) = 0.571$, $p = 0.751$) nor for the opposing character ($\chi^2(2) = 2.000$, $p = 0.368$).

5.4.4.A Subgroup Analysis

In an effort to better understand the results, we performed a subgroup analysis. We found that for a subgroup of *participants that enjoy and have played/watched others play action/fighting games multiple times* ($N = 16$) there were statistically significant improvements in multiple dimensions referring to the point of view of the characters *(Q7 - Q11)*. Again, for a detailed view of each statement of this subgroup analysis, see Appendix D.3.3. Figure 5.23 shows the averages and standard deviations of each statement for each version in this subgroup.

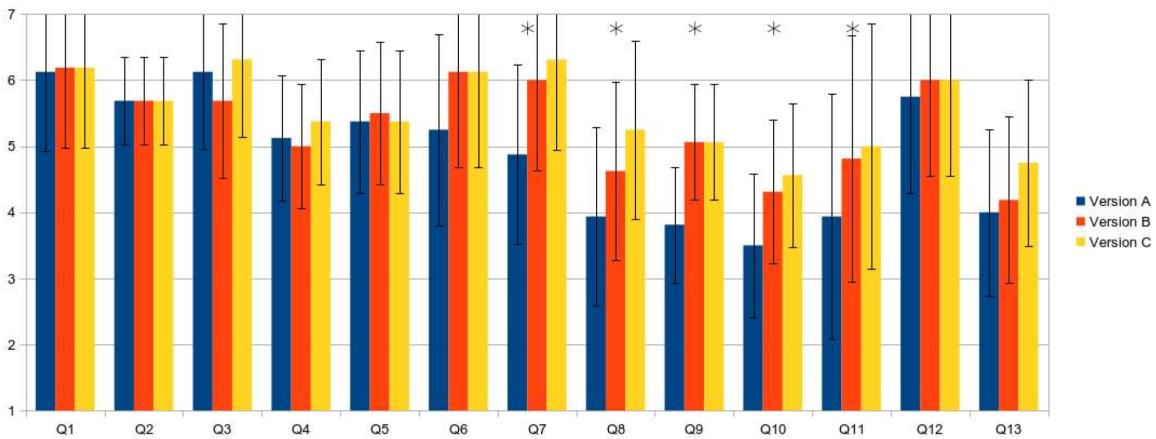


Figure 5.23: Averages and associated standard deviations of each statement for each version in the subgroup analysis. An asterisk (*) marks questions in which we found statistically significant differences in the Friedman test.

Q7 - The characters were aware of each other's actions. For this statement, version A had a mean of 4.88 ($SD = 1.544$); version B had a mean of 6.00 ($SD = 0.966$); and version C had a mean of 6.31 ($SD = 0.873$). Unlike the full sample analysis, here we did find statistically significant differences between versions ($\chi^2(2) = 9.591$, $p = 0.008$). Thus, we performed the post hoc analysis. No significant differences were found between versions C and B, $Z = -1.249, p = 0.212$. There were, however, significant differences between versions A and B, $Z = -2.501, p = 0.012$, and between versions A and C, $Z = -2.687, p = 0.007$.

These results suggest that both versions B and C are better than version A, and while not statistically different, version C is slightly better than version B. This means that the characters appear to be more aware of each other's actions in versions B and C than in version A, where C is rated higher than B, but not significantly.

Q8 - The characters could predict each others' actions. For this statement, version A had a mean of 3.94 ($SD = 1.569$); version B had a mean of 4.62 ($SD = 1.310$); and version C had a mean of 5.25 ($SD = 1.390$). Unlike the full sample analysis, we found statistically significant differences between the versions ($\chi^2(2) = 10.360$, $p = 0.006$). Thus, we performed the post hoc analysis. No significant differences were found between any pair: between versions A and B, $Z = -1.706, p = 0.088$; between versions C and B, $Z = -1.835, p = 0.067$; and between versions A and C, $Z = -2.311, p = 0.021$.

So, while a statistical difference was observed in the Friedman test, we found no major differences between pairs. We can still see that Version C was higher in mean value, followed by Version B, and Version C was the lowest in rank. If we wish to take meaning from these values, in general, the participants seemed to be unable to easily distinguish if the characters could predict each other's actions. Yet, for version A, there is a tendency for characters to be unable to perform a prediction, while in version C there is the opposite tendency, where characters seem more capable of predicting the opponent's actions.

Q9 - The characters were aware of each other's feelings. For this statement, version A had a mean of 3.81 ($SD = 1.642$); version B had a mean of 5.06 ($SD = 1.181$); and version C had a mean of 5.06 ($SD = 1.289$). Unlike the full sample analysis, we found statistically significant differences between the versions ($\chi^2(2) = 11.148, p = .004$). Thus, we performed the post hoc analysis. We found significant differences between versions A and B, $Z = -2.661, p = .008$. No significant differences were found between the remaining pairs: between versions C and B, $Z = -.040, p = 0.968$; and between versions A and C, $Z = -2.348, p = 0.019$.

In this statement, we see how even though two versions have the same average, B and C, because C has a wider standard deviation, there is no significant statistical difference between C and A. What can we take from these results? In general, the participants seemed to be unable to easily distinguish if the characters were aware of each other's feelings. Yet, for version A, there is a tendency for the characters to be perceived as unaware of others' feelings, while in versions B and C there is the opposite tendency, where characters seem more aware of each other's feelings.

Q10 - The characters could predict each others' feelings. For this statement, Version A had a mean of 3.50 ($SD = 1.549$); Version B had a mean of 4.31 ($SD = 1.078$); and Version C had a mean of 4.56 ($SD = 1.094$). Similarly to the complete sample analysis, we found statistically significant differences between versions ($\chi^2(2) = 6.318$, $p = 0.042$). Thus, we performed the post hoc analysis. No significant differences were found between versions A and B, $Z = -2.170, p = 0.030$, and between versions C and B, $Z = -1.265, p = 0.206$. We found significant differences between versions A and C, $Z = -2.438, p = 0.015$. Note that in the complete sample analysis, we did not find any significant differences between pairs.

Similarly to Q9, here versions B and C are very similar with values pointing towards the characters being able to predict each other's feelings, while version A suggests the opposite. Yet, overall, in all versions, participants appear to be unable to agree or disagree if the characters can predict the feelings of the other. Both this and Q9 seem to imply that having anticipatory responses has an impact, although small, on the participants' perception.

Q11 - The characters were aware of each other's intentions. For this statement, version A had a mean of 3.94 ($SD = 1.526$); version B had a mean of 4.81 ($SD = 0.911$); and version C had a mean of 5.00 ($SD = 1.317$). Similarly to the complete sample analysis, we found statistically significant differences between the versions ($\chi^2(2) = 8.341$, $p = 0.015$). Thus, we performed the post hoc analysis. No significant differences were found between versions C and B ($Z = -0.690, p = 0.490$) and between versions A and C ($Z = -2.201, p = 0.028$). We did find statistically significant differences between versions A and B, $Z = -2.488, p = 0.013$.

Although these results align with what we found in the complete sample analysis, here version C ranks highest, followed by version B, and then A. In these results, we find that version C is where the characters were perceived to be more aware of each other's intentions. Version B is a bit behind in score, but is almost as good. For version A, it was unclear to the participants if the characters were aware of each other's intentions or not.

5.4.4.B Other Comments

In addition to the results from these analyzes, we also gathered some information from the participants' comments during and after user testing and from the free-form comments at the end of the questionnaire.

Here, as in the previous evaluation, the participants report that the axe is a too powerful weapon compared to the sword and shield. While we made some adjustments to the damage on strikes, it appears that we needed to better fine-tune it for new players. Note that players that understand well how to block attacks found the axe less powerful than those who do not block attacks.

Additionally, some participants noted that because they focus so much on the combat, they do not focus as much on the characters' expressions. The same applies to the participant's awareness of the redesigned HUD, where participants became aware and often looked at it, or were completely unaware of it. Those aware of the HUD commented that it helped them better understand the character they were controlling.

5.4.5 Discussion

The main objective of this section was to describe our second evaluation on believability in Adfectus. Prior to the evaluation, we introduced a new emotional model, BALTAS, and some updates on the games' HUD to allow for character emotional expressions. We also bring in the improvements in emotion expressions from the previous study, including a new body expression of *Rage*, facial expressions, colored outlines, and background iconography.

The **BALTAS emotional** model introduces the *Emotivectors* [7] into *3motion* and allows for greater anticipatory behavior. This implementation takes the knowledge gathered from our previous implementation on the Virtual Coach application (see Section 4.3.1) and expands it with meta-predictors. Multiple emotivectors can output different or similar sensations. We define a Personality as a mapping between an emotion and a sensation, this solution is also similar to solutions we have seen previously (see Sections 4.3.1 and 5.2). To select which emotion is selected among the salient sensations, we also implemented several selection methods. The chosen method selects the emotion with the highest accumulated salience.

Having our generalist *Emotivector* concepts implemented, we looked into implementing *Emotivectors* specific to our use case. We created two, both related to health. The first is related to the health of one's own and allows the character to anticipate if they are close to defeat or still have a lot to give. The second is related to one's opponent's health and allows the character to anticipate if they are close to a victory or not. In both solutions, we also added anticipatory behaviors in which, if the characters believe certain events will happen (by comparing predictions with previous values and checking for significant differences), then they react in anticipation to them (e.g., if the character believes that the opponent will be able to hit it, when the opponent starts an attack, the character reacts negatively). These belief rules had an associated importance value that could be manipulated to further define the personality.

After defining our new emotional model, we went into the evaluation on believability. This evaluation mimics previous evaluations, namely those in Sections 3.5.1 and 5.2, where we compare different methods of computing emotions and ask participants to rate them. In this evaluation, we compared the *(A) Reactive Model*, the *(B) Predictive with Assertion Model*, and the *(C) BALTAS Model*. The first two models were two of the three in the previous evaluation (see Section 5.2) and help us understand whether the new additions of emotions expressions make a significant difference and how they compare with the BALTAS Model. We used a questionnaire containing some demographic questions followed by questions from the questionnaire on Believability, previously introduced in Section 3.5.1.A.

Our sample consisted of 32 participants between 20 and 33 years old, mostly men (22; 8 women; 2 preferred not to answer). 50% of the participants enjoy playing or watching action/fighting games. Most of the participants reported being aware of the expressions of the opposing character (varying between 88% and 97% of reported awareness in games played, depending on the version), while being less aware of their own character's expression (varying between 69% and 75% of reported awareness in games played, depending on the version).

Regarding the general believability of the interactions *(Q13)*, we concluded that neither the *(B) predictive with assertion model* nor the *(C) BALTAS model* appeared to improve the believability of character interactions compared to the more traditional purely *(A) reactive model*, $\chi^2(2) = 1.724$, $p = 0.422$. Where version *(C)* seemed to score the lowest among the three (5.50, $SD = 1.566$), followed by *(A)* (5.53, $SD = 1.319$), and version *(B)* the highest (5.78, $SD = 1.338$). However, most participants rated the interactions as believable regardless of the version, suggesting a positive experience even with only the reactive model. We believe that these high values are due to the emotion expression channels implemented, as there seems to be an overall improvement in believability, regardless of the emotion model used, compared with the previous evaluation (e.g. *(A) Reactive Model* increased from $M = 5.10$, $SD = 1.701$, to 5.53, $SD = 1.319$).

We did not find statistically significant results on most questions, yet most of the questions pointed toward a good perception of the character's actions and feelings, either from the point of view of the participant (Q1 - Q5) or the point of view of the characters (Q6 - Q11). For the questions where we did find significant differences (Q3, Q10, and Q11), here is what we found. From the point of view of the participant (Q3), there seems to be some confusion in understanding what the character is doing in version *(B)*, while the characters in the remaining versions have their feelings understood in a similar way. From the point of view of the character (Q10 and Q11), there appears to be little difference between versions *(B)* and *(C)*, while version *(A)* is worse. Meaning that characters in versions *(B)* and *(C)* are perceived as more aware of each other's intentions and as capable of predicting each other's feelings.

In an effort to better understand the results, we performed a subgroup analysis. We found that for a subgroup of *participants that enjoy and have played/watched others play action/fighting games multiple times* ($N = 16$) there were statistically significant improvements in multiple dimensions referring to the point of view of the characters *(Q7 - Q11)*. As seen previously in Q10 and Q11, in these statements, participants seem to rate versions *(B)* and *(C)* similarly, while rating version *(A)* lower. Note, however, that in most of these cases, although not significantly different, version *(C)* rates higher than version *(B)*. This suggests that further testing could result in statistically significant differences.

What can we gather from all of this? First, we found that **by improving emotion expression, there is an increase in overall believability**, even in more reactive solutions, as seen in version *(A) Reactive Model*; this shows the importance of performing a manipulation check, through the use of the TAI Method, which helps to ensure that the participant perceives the emotions correctly. We also found that from the perspective of the participant (Q1-Q5), there seems to be little difference between versions, suggesting that **participants moderately understand the characters regardless of version**. From the perspective of the characters (Q6-Q11), predictive models, versions *(B)* and *(C)*, seem to have a greater impact. The results suggest that **the characters in versions *(B)* and *(C)* are more aware of each other's actions, feelings, and intentions and are more capable of predicting each other's actions and feelings**. In these statements, there is a tendency for *(C)* to be ranked higher than *(B)*.

Looking back at the complete sample and focusing on the perception of the characters' personalities, results seemed to be inconclusive on whether or not the characters show different personalities. However, we performed a content analysis on the descriptions given of each character's personality. We found that the Axe character is always perceived as confident regardless of the version. Aside from this description, there is no clear distinction between characters in versions *(A)* and *(B)*, yet in version *(C)*, when a personality is perceived, it seems to align with the designed personality.

Closing this discussion, **we did not find statistically significant differences between versions in the rating of believability of interactions**, these findings suggest that **our hypothesis could not be validated**. It is important to note that most of the participants rated the interactions as believable regardless of the version. Furthermore, although not directly contributing to believability, the ***(C) BALTAS Model* seems to create characters more aware of others and more capable of expressing a unique personality**.

5.5 Summary

In this chapter, we discuss *Adfectus*, an arena game in which two characters, controlled by players, battle each other, our implementation of *3motion* in this use case, and several evaluations on believability and emotion perception.

5.5.1 Adfectus and 3Motion

Adfectus was developed as a testbed for *3motion*. It features two duelists, the characters controlled by the players. These characters can wield a sword and a shield, or an axe. They can perform attacks and blocks. Attacks can be interrupted, allowing feints to be performed. Additionally, attacks are subdivided as proposed in *3motion*: Interruptible Anticipation; Uninterruptible Anticipation; Uninterruptible Follow-through; Interruptible Follow-through. Thus, it allows for greater expression from characters during the duration of an attack. While a human player controls the duelists, their emotional expressions are independent and show how the duelist is experiencing combat. Duelists can express 4 emotions, *Confidence*, *Defensiveness*, *Rage*, and *Fear*. The expression of these emotions was progressively improved and, in the end, it featured body language, facial expressions, colored outlines, and background iconography.

3motion was implemented here to take advantage of action subdivisions. The steps were implemented as different models. *Perceive* step was implemented to store events of the world, specifically events on changes in characters, their health, position, actions (and their stages), and emotions, but also combat events that detail hits, misses, or blocks. *React* step had different implementations depending on the evaluation, but in the end, it was implemented in the BALTAS model, which incorporated multiple Emotivectors [7] that computed emotions for the character. *Decide* step, because the control is in the player's hands, was designed to mimic the input of the player. It was implemented in the GENA model and allowed the system to be aware of the actions that occur and their stages. Finally, the *Perform* step was implemented as an output of information for expression systems to consume.

5.5.2 First Evaluation on Believability

In this first version of the game, the emotion expressions were only performed through body movements. In this implementation, *3motion* was adorned with 3 emotional models (for the *React* step). These models focused on the calculation of a *health ratio*, the normalized difference between the health of both characters. The *Reactive Model* serves as the base model for comparison, as it is a common solution found in games. In it, the health ratio is computed every time the character receives a change in the world. The *Predictive Model* generates emotions only in the anticipation stage and in anticipation of how the *health ratio* changes. It computes predictions of the *health ratio* using weighted moving averages and their first and second derivatives. Essentially, instead of reacting to what is happening, the character reacts to what it thinks will happen. The third model, the *Predictive with Assertion Model*, follows the computations of the previous model and adds reactions to the follow-through stage of the actions, wherein, if the expectation of the *health ratio* mismatches with the actual ratio, a temporary boost is added to the health ratio to promote emotions to be computed.

Having our models, to understand which most improved the interaction's believability, we performed the first evaluation on believability. We resorted to user testing following a within-subjects design. As such, participants had the opportunity to experience all three models in various orders and answered a set of questions referring to the believability of the interaction and emotional behavioral patterns. The results for the entire sample show no significant differences between models, yet for participants more familiar with synthetic characters, both the *predictive* and *predictive with assertion* models result in more coherent emotion expression and better expression of characters' intentions compared to the *reactive* model, which indicates that, at least for this particular demographic, these models result in more believable interactions.

5.5.3 Applying the TAI Method in the Adfectus Characters

From the previous evaluation, we learned that participants had problems in identifying the expression of *Rage* and, due to the combat nature of the game, we wished to expand the emotion expression abilities of the characters. We reworked the expression of *Rage* to be more in line with the other emotions. Added *facial expressions*, through the use of emoji-like figures. Added *colored outlines* with colors inspired by research and media. And added *background iconography* in the shape of halos to reinforce the emotion expressed. To validate and understand if the improvements to the characters' animations and expressions were being perceived correctly, we employed the *TAI method* [22] to help us create a model of the perception of the characters.

After *construct* elicitation, we collected 13 meaningful constructs that describe how participants ($N = 29$) view the characters. The constructs, while varied, focused on descriptions of body movement (e.g. *Grounded/Shaky Posture*, *Head Down/Up*, *Offensive/Defensive Stance*).

After rating the animations ($N = 85$), we performed several data analyses, namely a PCA and a Cluster Analysis to create our model. In the Cluster Analysis, we concluded that 4 clusters were ideal for our study. These clusters have a one-to-one mapping for each emotion. In general, while the mapping is not perfect, namely, for the expression of *Defensiveness*, we can say that each emotion is perceived correctly and independently of the others. This supports that each emotion is being perceived correctly.

Interpreting the PCA components and the generated clusters, we created a more detailed view of how each emotion is perceived. **Confidence** has a relaxed posture with a cheerful facial expression and the weapon is held high and back. **Defensiveness** has a defensive stance with a grimacing facial expression and the weapon is held in front. **Fear** has a shaky posture and the weapon is held low. **Rage** has a prepared and ready posture with a grimacing facial expression and the weapon is held in front.

With these results, we have a clearer understanding of the perception of the user and can infer that the different emotion expressions are being correctly and independently perceived by the participants.

5.5.4 Second Evaluation on Believability

With the new expressions properly validated in the last evaluation, we aimed to improve our emotional model. To this end, we introduced a new emotional model, BALTAS Model. This model introduces the *Emotivectors* [7] into *3motion* and allows for greater anticipatory behavior. We implemented two *Emotivectors* specific to our use case. Both are related to health variations. The first is related to the health of one's own and allows the character to anticipate if they are close to defeat or still have a lot to give. The second is related to one's opponent's health and allows the character to anticipate if they are close to a victory or not. We also added anticipatory behaviors in which, if the characters believe certain events will happen, then they react in anticipation to them.

After defining our new emotional model, we conducted a new assessment of believability. In this evaluation, we compared the *Reactive Model*, the *Predictive with Assertion Model*, and the *BALTAS Model*. The first two models were two of the three in the previous evaluation. With a sample of 32 participants, we did not find significant differences between many of the questionnaire statements. Furthermore, we found **no significant differences between any version regarding the believability of character interactions**, however, most of the participants rated the interactions as believable regardless of the version. A comparison between these values and those from the first evaluation on believability suggests that **by improving emotion expression, there is an increase in overall believability**, regardless of version.

An analysis of the significantly different statements, either from the complete sample or from a subgroup with participants who *enjoy and have played/watched others play action/fighting games multiple times* ($N = 16$), shows that **participants understand the characters well regardless of the version** and the results suggest that **the characters in versions with the *Predictive with Assertion Model* and the *BALTAS Model* are more aware of each other's actions, feelings, and intentions and are more capable of predicting each other's actions and feelings**. Finally, although not directly contributing to believability, the ***(C) BALTAS Model* seems to create characters more aware of others and more capable of expressing a unique personality**.

This final evaluation concludes this chapter and the evaluations of this work. Recalling our initial hypothesis, it states that "by *subdividing actions according to the traditional principles of animation* and allowing for *emotional reactions and interruption of behavior* during their expression, the overall *real-time interaction will be perceived as more believable*". When we looked at our results, we did not seem to verify our initial hypothesis. Yet, through the multiple evaluations performed, if we were to summarize our findings in a single sentence, here is what it would be:

> By *subdividing actions according to the traditional principles of animation* and allowing *emotional reactions and interruption of behavior* during their expression, *characters will be perceived as more aware of others and more capable of expressing unique personalities*.

6

Conclusion

Contents

"Não ter já mais nada para dizer e continuar a escrever é um crime. Porque não tem o direito de continuar a escrever se não tem nada a dizer."

José Saramago

In this chapter we present a closure to this work along with a summary of our research and its findings. We start by reviewing a problem regarding believability we identified early at the start of this work and our hypothesis on how we would tackle it. Following, we touch on the most relevant related work that helped inform our decisions throughout this work. We then review the action subdivision and its implementation in 3motion*, along with the first evaluation using this agent model. We discuss our work on virtual coaches and its implications on our implementation of emotional models and the perception of emotion expression in synthetic characters. Subsequently, we recall the implementation of* Adfectus *along with the multiple evaluations on emotion expression and believability that help to understand the validity of our hypothesis and the impact of our research. To close, we present some future work and a takeaway message from this work.*

6.1 A Hobbit in the hole

We started by exploring the concept of immersion in films and video games. We discussed how clear intentions and well-timed interactions in films create anticipation and a sense of presence, leading to the suspension of disbelief. Video games achieve immersion through scripted scenes and limited player actions, particularly in first-person shooter games. However, open-world role-playing games struggle with immersion due to unnatural interactions with NPCs. In this setting, the lack of verbal and nonverbal communication between characters during action-oriented interactions results in mechanical and oversimplified sequences that break immersion. Faced with this problem, we wondered **how can we enable believable real-time multiparty interactions between synthetic characters?**

We hypothesized that by focusing on improving individual actions, instead of the whole interaction, and that **by *subdividing actions according to the traditional principles of animation* [9] and allowing for *emotional reactions and interruption of behavior* during their expression, the overall *real-time interaction will be perceived as more believable***. An action is divided into three stages: anticipation, action, and follow-through. Each stage takes some time to play out, allowing emotion expression and interpretation to be done under different contexts of the same action. By explicitly modeling a split of an action into the three stages, we expected the communication of the intentions of a character to be perceived more clearly and give a richer emotional context for all the characters involved in the scene, improving believability.

To test our hypothesis, we proposed the use and expansion of *3motion*, an agent model that makes use of action subdivision, along with the creation of a use case, *Adfectus*, an arena game in which two characters, controlled by players, battle each other. To validate the hypothesis, we propose a comparison between a version of *Adfectus* in which characters use action subdivision with *3motion* and a second version with atomic actions.

6.2 A room filled with Books

To better guide our development and research, we looked at different but intertwined subjects, namely, believability, emotion computation and expression, the principles of traditional animation, and works tackling similar problems we are focusing on.

To address our problem in the correct context we presented related work on believability and its different definitions by different authors. We proposed our own definition of believability as "the belief that an environment, its characters, and their interactions give the illusion of life". We also provide definitions for the related concepts of believable environments, characters, and interactions. Surrounding this topic, we delve into what believable agents are, how they need awareness and situatedness to be grounded in the world, and how the theory of mind and empathy allow characters to bond and be more believable. We then discussed works that tackle the assessment of believability, using subjective, objective, or gameplay-based approaches. We focused on the subjective assessment of believability made mainly through user-reports.

One of the crucial elements that make a character believable is emotion; thus, we then explored how emotions are defined and how they are expressed. We discussed their link with anticipation and how one influences the other, through action tendencies, and vice-versa, through context. We then move to how emotions are used in synthetic characters, looking into their internal representation, how emotions are created, and their external representation, how emotions are expressed. Regarding the internal representation, special attention was given to Emotivector, an emotion model that makes use of anticipation to generate emotions, which we make use in several solutions throughout the work. Regarding external representation, we discussed several emotion communication channels (primary or secondary), such as emotional body language, describing how we view emotion expression in body movements, facial expressions, describing how we view emotion expressions in facial movements, the use of colors to represent emotions, or other iconographic symbology for emotions. To close the study on emotions, a review of recognition and assessment of emotion expression is shown, touching on multiple methods for data collection. Here, as in believability, we focused on user-reports.

Connecting believability, emotion, and expression, we discuss the principles of traditional animation. These principles are a set of standardized practices followed by Disney's animators to create believable animations and can be applied to digital environments. We also briefly touch on an expanded set of principles targeted at game animations. Afterward, we discuss the subdivision of an action applied to an interactive experience to help create more believable interactions and characters.

Finally, we review a selection of recent works that explore multi-party interactions through communication and emotion expression between synthetic characters. We found that works tackle some of our multiple areas of interest, that is, multiparty interactions, emotion expression, real-time interactions, and believability, yet works tackling all these areas is rare or nonexistent. Those works tackling multi-party interactions tend to focus on decision-making but lack emotion expression, while those showing emotion expression tend to focus on human-agent interactions, but often lack in multi-party interactions.

6.3 Doing it because it is hard

With much anticipation... and action... and follow-through... we arrive at *3motion*, our solution to enable believable multi-party real-time interactions through the subdivision of actions and usage of non-verbal emotional communication.

We first introduced the concept of **action subdivision into three stages: anticipation stage, action stage, and follow-through stage**. **The anticipation stage is further subdivided into interruptible and uninterruptible sub-stages, and the follow-through stage is further subdivided into uninterruptible and interruptible sub-stages**. This solution allows a range of emotions to be communicated over the duration of a single action, which deepens the possible interactions and allows greater emotional involvement of the agents. After that, we discussed how the action subdivision integrates with the agents' internal behavior loop, which is split into four steps: Perceive, React, Decide, and Perform. Delving into each step, we discussed the perception of the environment, emotional responses, decision-making processes, and finally the performance of actions and expression of emotions using the action subdivision.

Then a text-based application was described as one of the use cases of *3motion*. Text and progress bar elements were used to display unfolding scenes with agents interacting. A user study was performed where 52 participants were asked to evaluate the believability of three videos showcasing a scene of application, differentiated by their use of action subdivision and of different timings. The results suggest that the video with action subdivision and correct timing, which essentially depicts the correct usage of *3motion*, was rated more believable in almost every statement compared to the other two videos.

6.4 A silence with consequences

Having introduced our agent model, we presented an architecture that allows for the creation of intelligent virtual agents acting as coaches in distance learning environments. While not directly related to *3motion* or *Adfectus*, this body of work presents several lessons on the implementation of synthetic characters capable of expressing emotions (e.g., making use of *Emotivectors*) and on emotion recognition, where we introduce the Triad Affect Interpretation (TAI) method as a method to understand participants' perception of emotions.

In this architecture, based on (1) the objective and subjective metrics provided by the users, as well as the learning environment, (2) the affective state of the users, inferred from the evolution of these metrics over time, how they create expectations, and how these expectations are actually met, and (3) the synthetic affective response of the virtual coach based on its synthetic personality, the system is able to select from a repertoire of dialogue trees gathered from a human pedagogical team, the one that better conveys the interpersonal regulation strategy appropriate for that situation. Having our system defined, we provided an overview of all components built in a first implementation of the approach and described a real-world case study featuring two virtual coaches, running on a mobile device during a 10-day evaluation period, whose results support the interest of students in the continued interaction with the empathic virtual coaches.

To better understand the impact on emotion expression/recognition of our prototype and the changes made to make the said expression more natural, we conducted two sets of experiments. The first one directly compared the first version to the second version (side-by-side) while expressing specific emotions, to both expert and non-expert participants. The results clearly supported that the second version was better rated at communicating each individual emotion. The follow-up experiment asked participants to recognize the individual emotions expressed by each version. The results showed that the emotion recognition of the second version had not improved in all cases, as was expected based on the first experiment. We noticed that, although the expression for some emotions was rated as being better in the first experiment, the second version had a lower recognition rate than the first version, as we could confirm in the second experiment. We also noticed that particular elements were reported to influence emotion recognition in different emotion expressions, pointing to separate needs for each family of emotions.

This discussion exemplifies why, when choosing the system that will better support the expression of emotions for intelligent virtual agents, in the sense that the portrayed emotions will be easier to recognize, asking the users to compare two different versions of the system, side by side, may not be the most adequate way to proceed. This study, therefore, mirrors the incremental formative evaluations that often underlie virtual agent development and offers a cautionary tale in evaluating aspects of a system during its development and its evolving context of use.

6.4.1 The TAI Method

In an effort to better understand the user's perception of our characters and their animations, we created the TAI method, which allows for the creation of a model of how people perceive different emotions in specific synthetic characters. The generated model provides insight into the different emotions, their similarities, and what features lead to their misinterpretation.

We started by selecting six animations from our characters, each representing a different emotion. Then, we collected meaningful constructs by asking participants to fill out a questionnaire containing triad comparison tasks, where three videos portraying animations of emotions were presented side-by-side, and asking participants to identify the two that are more alike and different from the third. Using the collected constructs, we designed a questionnaire in which each animation was to be rated against each construct using a bipolar scale. Using PCA and cluster analysis, we then created our model for the perception of emotions animations. PCA reduced the number of constructs to 3 correlated components: *(C1)* corresponds to how people viewed the face in general (if it was more contracted or expanded); *(C2)* relates to the movement of the mouth and its relation to the cheeks movement; and *(C3)* relates to the opening of the mouth and the visibility of the teeth. To further understand the similarities between the animations associated with the expression of emotions, we performed a cluster analysis that grouped the animations into 4 clusters: *Anger* and *Disgust* were grouped in Cluster α; *Fear* and *Surprise* in Cluster β; *Sadness* in Cluster γ; and *Happiness* in Cluster δ.

With the TAI method, it is possible to create a model of how people perceive different emotions that can provide insight into the different emotions that may be confused with each other due to specific features of the animation system. We can visualize which features are more prominent in each different cluster and thus how the different animations are more alike. *Anger* and *Disgust* are characterized mainly by a contraction of the face (Component 1); *Fear* and *Surprise* by the expansion of the face and the shape of the mouth being rounded (Components 1 and 2); *Happiness* by its smile, a curved shaped mouth (Component 2); and finally *Sadness* by the closed mouth with no teeth visible (Component 3). These features could be exaggerated or modulated to achieve the desired recognition level.

Our work on the virtual coaches informs the development of specific empathic synthetic characters; we can extrapolate this knowledge to other contexts as it shares many common features of synthetic characters' development. The emotion generation algorithm presented here is used as the basis for the emotional model in *3motion*, as it provides anticipatory capabilities to agents using Emotivectors. Furthermore, the lessons learned from the comparative studies help guide future evaluations of the main work, not only as cautionary tales but also as comparative data when considering emotion recognition tasks. Finally, the TAI method is also used in the context of a full-bodied animated character, showcasing its utility with different synthetic characters and helping to validate their expressions.

6.5 Learning from the past

At this point, we have evaluated *3motion* in a very simple setting with low interaction by the participants. In order to assess our agent model solution in a more dynamic environment (while recognizing the lessons learned from working with virtual coaches), we created *Adfectus*, an arena game in which two characters, controlled by players, battle each other.

6.5.1 Adfectus and 3Motion

Adfectus was developed as a testbed for *3motion*. It features two duelists, the characters controlled by the players. These characters can wield a sword and a shield, or an axe. They can perform attacks and blocks. Attacks can be interrupted, allowing feints to be performed. Additionally, attacks are subdivided as proposed in *3motion*: Interruptible Anticipation; Uninterruptible Anticipation; Uninterruptible Follow-through; Interruptible Follow-through. Thus, it allows for greater expression from characters during the duration of an attack. While a human player controls the duelists, their emotional expressions are independent and show how the duelist is experiencing combat. Duelists can express 4 emotions, *Confidence*, *Defensiveness*, *Anger*, and *Fear*. The expression of these emotions was progressively improved and, in the end, it featured body language, facial expressions, colored outlines, and background iconography.

3motion was implemented to take advantage of action subdivisions. The steps were implemented as different models. *Perceive* step was implemented to store events of the world, specifically events on changes in characters, their health, position, actions (and their stages), and emotions, but also combat events that detail hits, misses, or blocks. *React* step had different implementations depending on the evaluation, but in the end, it was implemented in the BALTAS model, which incorporated multiple Emotivectors [7] that computed emotions for the character. *Decide* step, because the control is in the player's hands, was designed to mimic the input of the player. It was implemented in the GENA model and allowed the system to be aware of the actions that occur and their stages. Finally, the *Perform* step was implemented as an output of information for expression systems to consume.

6.5.2 First Evaluation on Believability

We implemented the first version of the game. At this stage, the emotion expressions were only performed through body movements. In this implementation, *3motion* was adorned with 3 emotional models (for the *React* step). These models focused on the calculation of a *health ratio*, the normalized difference between the health of both characters. The *Reactive Model* serves as the base model for comparison, as it is a common solution found in games. In it, the health ratio is computed every time the character receives a change in the world. The *Predictive Model* generates emotions only in the anticipation stage and in anticipation of how the *health ratio* changes. It computes predictions of the *health ratio* using weighted moving averages and their first and second derivatives. Essentially, instead of reacting to what is happening, the character reacts to what it thinks will happen. The third model, the *Predictive with Assertion Model*, follows the computations of the previous model and adds reactions to the follow-through stage of the actions, wherein, if the expectation of the *health ratio* mismatches with the actual ratio, a temporary boost is added to the health ratio to promote emotions to be computed.

Having our models, to understand which most improved the interaction's believability, we performed the first evaluation on believability. We resorted to user testing following a within-subjects design. As such, participants had the opportunity to experience all three models in various orders and answered a set of questions referring to the believability of the interaction and emotional behavioral patterns. The results for the entire sample show no significant differences between models, yet for participants more familiar with synthetic characters, both the *predictive* and *predictive with assertion* models result in more coherent emotion expression and better expression of characters' intentions compared to the *reactive* model, which indicates that, at least for this particular demographic, these models result in more believable interactions.

6.5.3 Applying the TAI Method in Adfectus Characters

From the previous evaluation, we learned that participants had problems in identifying the expression of *Rage* and, due to the combat nature of the game, we wished to expand the emotion expression abilities of the characters. We reworked the expression of *Rage* to be more in line with the other emotions and added *facial expressions*, using emoji-like figures, *colored outlines* with colors inspired by research and media, and *background iconography* in the shape of halos to reinforce the emotion expressed. To validate and understand whether improvements to the animations and expressions of the characters were being perceived correctly, we employed the *TAI method* [22] to help us create a model of the perception of the characters.

After *construct* elicitation, we collected 13 meaningful constructs that describe how participants ($N = 29$) view the characters. The constructs, while varied, focused on descriptions of body movement (e.g. *Grounded/Shaky Posture*, *Head Down/Up*, *Offensive/Defensive Stance*).

After rating the animations ($N = 85$), we performed several data analyses, namely a PCA and a Cluster Analysis to create our model. In the Cluster Analysis, we concluded that 4 clusters were ideal for our study. These clusters have a one-to-one mapping for each emotion. In general, while the mapping is not perfect, namely, for the expression of *Defensiveness*, we can say that each emotion is perceived correctly and independently of the others. This confirms that each emotion is being perceived correctly.

Interpreting the PCA components and the generated clusters, we created a more detailed view of how each emotion is perceived. **Confidence** has a relaxed posture with a cheerful facial expression and the weapon is held high and back. **Defensiveness** has a defensive stance with a grimacing facial expression and the weapon is held in front. **Fear** has a shaky posture and the weapon is held low. **Rage** has a prepared and ready posture with a grimacing facial expression and the weapon is held in front.

With these results, we have a clearer understanding of the perception of the user and can infer that the different emotion expressions are being correctly and independently perceived by the participants.

6.5.4 Second Evaluation on Believability

With the new expressions properly validated in the last evaluation, our objective now was to improve our emotional model. Recall our hypothesis of improving believability through action subdivision, behavior interruption, and emotion expressions. With this in mind, we introduced the BALTAS Model. This emotional model introduces the *Emotivectors* [7] into *3motion* and allows for greater anticipatory behavior. We implemented two *Emotivectors* specific to our use case. Both are related to health variations. The first is related to the health of one's own and allows the character to anticipate if they are close to defeat or still have a lot to give. The second is related to one's opponent's health and allows the character to anticipate if they are close to a victory or not. We also added anticipatory behaviors in which, if the characters believe certain events will happen, then they react in anticipation to them.

After defining our new emotional model, we conducted a new assessment of believability. In this evaluation, we compared the *Reactive Model*, the *Predictive with Assertion Model*, and the *BALTAS Model*. The first two models were two of the three in the previous evaluation. With a sample of 32 participants, we did not find significant differences between many of the questionnaire statements. Furthermore, we found **no significant differences between any version on the believability of character interactions**, however, most of the participants rated the interactions as believable regardless of the version. A comparison between these values and those from the first evaluation on believability suggests that **by improving emotion expression, there is an increase in the overall believability of interactions**, regardless of version.

An analysis of the significantly different statements, either from the complete sample or from a subgroup with participants who *enjoy and have played/watched others play action/fighting games multiple times* ($N = 16$), shows that *participants understand the characters well regardless of the version* and the results suggest that *the characters in versions with the Predictive with Assertion Model and the BALTAS Model are more aware of each other's actions, feelings, and intentions and are more capable of predicting each other's actions and feelings*. Finally, although not directly contributing to believability, the *BALTAS Model seems to create characters more aware of others and more capable of expressing a unique personality*.

This final evaluation concludes the evaluations of this work. Recalling again our hypothesis, it states that "by *subdividing actions according to the traditional principles of animation* and allowing for *emotional reactions and interruption of behavior* during their expression, the overall *real-time interaction will be perceived as more believable*". When we looked at our results, although many statements seem to suggest improvements in many fields related to believability when using the BALTAS model, we did not seem to verify our initial hypothesis. Yet, through the multiple evaluations performed, if we were to summarize our findings in a single sentence, here is what it would be:

> **By *subdividing actions according to the traditional principles of animation* and allowing *emotional reactions and interruption of behavior* during their expression, characters will be perceived as more aware of others and more capable of expressing unique personalities.**

6.6 Having nothing more to say (... for now)

With this work coming to a close, we look into the future. Looking at what we developed, there are many lines of research that could be expanded. Not only is the concept of believability a field of many threads, but so are the areas exploring the use of timing, emotion expression, emotion recognition, interruption of behavior, and just development of interactive experiences in general. Let us explore some of the threads of research we find more relevant.

When applying the TAI Method, our evaluation showcased videos of the emotion expression, however, that is not how players experience the characters' expressions when playing *Adfectus*. So, if we were to apply the TAI Method using an interactive experience where participants could control the characters and see their expressions, how would that impact the constructs collected? Furthermore, until now we have made heavy use of user self-reporting to collect our data. However, in similar studies, objective data collection, such as physiological data, is not uncommon to help better understand the impact of the experience on the participant. **Does objective data, such as eye tracking, collected from a participant align with the responses the participant reports in the questionnaire?**

Our last evaluation on believability featured 32 participants. While enough for statistical significance, the results were not significantly different between versions in many cases. Part of the results suggests that, with a larger sample, we might find more significant results, mainly in questions that focus on the interpretation of character behavior from the perspective of the characters. Therefore, an important piece of future work is that of **conducting the experiment on believability with a larger sample size**.

Remaining on the topic of evaluations on believability. Our experiment asked people to be active participants in the experiment, in which their performance impacted their perception of the events. This explores only part of the possible channels of perception. Employing **an experiment where participants were passive**, in which no game interaction is required by them, essentially making them spectators of gameplay (with selected videos, for example), and asking them to fill out the believability questionnaire, would offer new views on the current application and helps understand its **impact on believability**.

In the current implementation of *3motion* and the BALTAS Model, we implemented a solution that allowed for the configuration of *Emotivectors* and the creation of personalities. We explored a very limited set of personalities that were aligned with our emotion expressions and the setting of the game. A path to future work is a **greater expansion of personalities and anticipatory behavior**, through the manipulation of importance values or the use of new *Emotivectors*. We could even ask, for example, if removing certain anticipatory behaviors helps create more believable interactions.

In *Adfectus*, we presented two duelists, which we have heavily discussed. These react to each other's actions and emotions, but what if a third or fourth entity is present? How would the characters react to them? To answer this question, we propose future work on the exploration of a crowd that would support the duelists. The crowd would have claques for each duelist and support it during combat. In this setting, while the duelists would be restricted to the same gameplay of fighting each other, the presence of the crowd allows for emotional reactions between the claque and the duelist, expanding the previous use case. With the previous use case, emotional reactions would favor antagonist responses to the opponent (e.g., if the opponent is fearful, my character is confident), and with the addition of the claque, we have a mix of antagonistic behavior targeting the opponent while having sympathetic reactions targeting the claque. With this future work, we ask **"How is believability impacted by the presence of more characters and their interactions?"**

Having said everything (... for now), we close with a humbling excerpt from one of my favorite books.

> *"Then the prophecies of the old songs have turned out to be true, after a fashion!" said Bilbo.*
>
> *"Of course!" said Gandalf. "And why should not they prove true? Surely you don't disbelieve the prophecies, because you had a hand in bringing them about yourself? You don't really suppose, do you, that all your adventures and escapes were managed by mere luck, just for your sole benefit? You are a very fine person, Mr. Baggins, and I am very fond of you; but* ***you are only quite a little fellow in a wide world after all!****"*
>
> ***"Thank goodness!"*** *said Bilbo laughing, and handed him the tabacco-jar.*
>
> *The Hobbit, J.R.R. Tolkien*

Bibliography

[1] E. Brown and P. Cairns, “A grounded investigation of game immersion,” in *Extended abstracts of the 2004 conference on Human factors and computing systems - CHI ’04*. New York, New York, USA: ACM Press, 2004, p. 1297. [Online]. Available: http://portal.acm.org/citation.cfm?doid=985921.986048

[2] S. Böcking, “Suspension of Disbelief,” in *The International Encyclopedia of Communication*. Chichester, UK: John Wiley & Sons, Ltd, 6 2008. [Online]. Available: https://onlinelibrary.wiley.com/doi/full/10.1002/9781405186407.wbiecs121https://onlinelibrary.wiley.com/doi/abs/10.1002/9781405186407.wbiecs121https://onlinelibrary.wiley.com/doi/10.1002/9781405186407.wbiecs121

[3] F. Noroozi, C. A. Corneanu, D. Kaminska, T. Sapinski, S. Escalera, and G. Anbarjafari, “Survey on Emotional Body Gesture Recognition,” *IEEE Transactions on Affective Computing*, vol. 12, no. 2, pp. 505–523, 4 2021. [Online]. Available: https://ieeexplore.ieee.org/document/8493586/

[4] R. M. Maatman, J. Gratch, and S. Marsella, “Natural behavior of a listening agent,” in *International Workshop on Intelligent Virtual Agents*, Springer. Kos Greece: Springer, 2005, pp. 25–36.

[5] E. Bevacqua, M. Mancini, and C. Pelachaud, “A listening agent exhibiting variable behaviour,” in *Intelligent Virtual Agents*, Springer. Tokyo Japan: Springer, 2008, pp. 262–269.

[6] J. Togelius, G. N. Yannakakis, S. Karakovskiy, and N. Shaker, “Assessing believability,” in *Believable Bots: Can Computers Play Like People?*, P. Hingston, Ed. Berlin, Germany: Springer Berlin Heidelberg, 2012, vol. 9783642323, pp. 215–230.

[7] C. Martinho and A. Paiva, “Using Anticipation to Create Believable Behaviour,” *Proceedings of the national conference on Artificial Intelligence*, vol. 1, pp. 175–180, 2006. [Online]. Available: http://www.aaai.org/Papers/AAAI/2006/AAAI06-028.pdf

[8] V. Nayak and M. Turk, “Emotional Expression in Virtual Agents Through Body Language,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Berlin, Heidelberg: Springer, Berlin, Heidelberg, 2005, vol. 3804 LNCS, pp. 313–320. [Online]. Available: https://link.springer.com/chapter/10.1007/11595755_38http://link.springer.com/10.1007/11595755_38

[9] F. Thomas, O. Johnston, and W. Rawls, *Disney animation: The illusion of life*. Burbank, California, USA: Disney Editions; Rev Sub edition, 1981, vol. 4.

[10] J. Mooney and J. M. Allbeck, "Rethinking NPC intelligence - A new reputation system," *Proceedings - Motion in Games 2014, MIG 2014*, pp. 55–60, 2014.

[11] J. Bates, "The role of emotion in believable agents," *Communications of the ACM*, vol. 37, no. 7, pp. 122–125, 1994.

[12] J. Orkin, "Symbolic representation of game world state: Toward real-time planning in games," in *AAAI Workshop - Technical Report*, vol. WS-04-04, 2004.

[13] M. Ochs, N. Sabouret, and V. Corruble, "Simulation of the dynamics of nonplayer characters' emotions and social relations in games," *IEEE Transactions on Computational Intelligence and AI in Games*, vol. 1, no. 4, pp. 281–297, 12 2009.

[14] T. Bosse and F. P. de Lange, "On virtual agents that regulate each other's emotions," *Web Intelligence and Agent Systems: An International Journal*, vol. 9, no. 1, pp. 57–67, 2011. [Online]. Available: https://www.medra.org/servlet/aliasResolver?alias=iospress&doi=10.3233/WIA-2011-0205

[15] K. Ijaz, A. Bogdanovych, and S. Simoff, "Enhancing the Believability of Embodied Conversational Agents through Environment-, Self- and Interaction-Awareness," in *Proceedings of the Thirty-Fourth Australasian Computer Science Conference - Volume 113*, ser. ACSC '11. AUS: Australian Computer Society, Inc., 2011, p. 107–116.

[16] R. Rodrigues and C. Martinho, "Towards Believable Interactions Between Synthetic Characters," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, J. Beskow, C. Peters, G. Castellano, C. O'Sullivan, I. Leite, and S. Kopp, Eds. Cham: Springer International Publishing, 2017, vol. 10498 LNAI, pp. 385–388. [Online]. Available: http://link.springer.com/10.1007/978-3-319-67401-8_48

[17] ——, "Measuring Believable Interactions, a Case Study: 3Motion," *Interaction with Agents and Robots: Different Embodiments, Common Challenges (Workshop at IVA 2017)*, 2017.

[18] R. Rodrigues, "Enriching Discrete Actions with Impactful Emotions," in *2019 8th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos, ACIIW 2019*. Cambridge, United Kingdom: IEEE, 2019, p. 39.

[19] A. P. Cláudio, M. B. Carmo, J. B. d. Silva, C. B. Alves, R. Costa, A. Marcos, E. Carvalho, V. Rocio, L. Morgado, L. Morgado, A. A. Varghese, S. Seixas, D. M. V. Barros, R. Rodrigues, and C. Martinho, "Empathic mediators for distance learning courses," *Proceedings ICGI - 2018. International Conference on Graphics and Interaction*, pp. 177–178, 2018. [Online]. Available: https://repositorioaberto.uab.pt/handle/10400.2/7970

[20] R. Rodrigues, R. Silva, R. Pereira, and C. Martinho, “Interactive Empathic Virtual Coaches Based on the Social Regulatory Cycle,” in *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*. Cambridge, United Kingdom: IEEE, 9 2019, pp. 69–75. [Online]. Available: https://ieeexplore.ieee.org/document/8925522/

[21] ——, “A cautionary tale of side-by-side evaluations while developing emotional expression for intelligent virtual agents,” in *Proceedings of the 22nd ACM International Conference on Intelligent Virtual Agents*. New York, NY, USA: ACM, 9 2022, pp. 1–8. [Online]. Available: https://dl.acm.org/doi/10.1145/3514197.3549672

[22] T. Ribeiro, R. Rodrigues, and C. Martinho, “Modeling the Interpretation of Animations to Help Improve Emotional Expression,” in *Proceedings of the 2023 International Conference on Autonomous Agents and Multiagent Systems*, ser. AAMAS ’23. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems, 2023, pp. 2301–2303.

[23] J. Silva, R. Rodrigues, and C. Martinho, “Believability, Anticipation, and... Timing Improving believability through timing manipulation,” in *Proceedings of the 18th International Conference on the Foundations of Digital Games*. New York, NY, USA: ACM, 4 2023, pp. 1–11. [Online]. Available: https://dl.acm.org/doi/10.1145/3582437.3582485

[24] C. Pacheco, “PAGAN for Character Believability Assessment,” 2020. [Online]. Available: https://davidmelhart.com/projects/mazing/index.html

[25] M. Gawron and U. B. C. G. Innovations, “Implementation of the hair lighting model in Unreal Engine 4 based on the state-of-the-art,” *Computer Game Innovations*, p. 29, 2018. [Online]. Available: https://www.researchgate.net/profile/Adam-Wojciechowski-3/publication/335524620_Computer_Game_Innovations_2018/links/5d6a9bbb92851c85388334b8/Computer-Game-Innovations-2018.pdf#page=29

[26] T. Alldieck, G. Pons-Moll, C. Theobalt, and M. Magnor, “Tex2Shape: Detailed Full Human Body Geometry from a Single Image,” in *{IEEE} International Conference on Computer Vision ({ICCV})*. IEEE, 2019.

[27] K. Zhou, B. L. Bhatnagar, and G. Pons-Moll, “Unsupervised Shape and Pose Disentanglement for 3D Meshes,” in *European Conference on Computer Vision (ECCV)*, 2020.

[28] A. Ranjan, T. Bolkart, S. Sanyal, and M. J. Black, “Generating 3D Faces using Convolutional Mesh Autoencoders,” in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018.

[29] Q. Reynaud, J.-Y. Donnart, and V. Corruble, “Evaluating the Impact of Anticipation on the Efficiency and Believability of Virtual Agents,” in *Intelligent Virtual Agents*, T. Bickmore, S. Marsella, and C. Sidner, Eds. Cham: Springer International Publishing, 2014, pp. 360–363. [Online]. Available: http://link.springer.com/10.1007/978-3-319-09767-1_47

[30] J. Silva, “Improving believability through timing manipulation,” Master’s thesis, Instituto Superior Técnico, University of Lisbon, 2022. [Online]. Available: https://fenix.tecnico.ulisboa.pt/cursos/meic-t/dissertacao/565303595503153

[31] J. Cooper, *Game Anim: Video Game Animation Explained*, 2nd ed. Princeton, New Jersey: CRC Press, 2021.

[32] H. Prendinger and M. Ishizuka, *Life-Like Characters: Tools, Affective Functions, and Applications*, ser. Cognitive Technologies, and and H. P. Mitsuru Ishizuka, Eds. Springer, 2004.

[33] M. Mori, K. F. MacDorman, and N. Kageki, “The uncanny valley,” *IEEE Robotics and Automation Magazine*, vol. 19, no. 2, pp. 98–100, 2012.

[34] C. Martinho, “Emotivector - Mecanismo Afectivo e Anticipatório para Personagens Sintéticas,” Ph.D. dissertation, Instituto Superior Técnico, 2007.

[35] J. Bates, “The nature of characters in interactive worlds and the Oz project,” Carnegie Mellon University, Tech. Rep. October 1992, 1993. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.50.6500

[36] A. Ortony, “On making believable emotional agents believable,” in *Emotions in Humans and Artifacts*. Cambridge, Massachusetts, USA: MIT, 2003, pp. 189–211. [Online]. Available: https://books.google.es/books?hl=es&lr=&id=jTgMIhy6YZMC&oi=fnd&pg=PA189&dq=On+making+believable+emotional+agents+believable&ots=I9fIhgy2AP&sig=Xd4a0oPqk679Kjk0s1gyhCeldfs

[37] P. Gomes, A. Paiva, C. Martinho, and A. Jhala, “Metrics for Character Believability in Interactive Narrative,” in *Interactive Storytelling*. London, UK: Springer International Publishing, 2013, vol. 8230, pp. 223–228. [Online]. Available: https://link.springer.com/10.1007/978-3-319-02756-2_27

[38] A. Bogdanovych, T. Trescak, and S. Simoff, “What makes virtual agents believable?” *Connection Science*, vol. 28, no. 1, pp. 83–108, 2016. [Online]. Available: https://doi.org/10.1080/09540091.2015.1130021

[39] S. Chandrasekharan and B. Esfandiari, “Software Agents and Situatedness: Being Where,” in *Proceedings of the Eleventh Mid-west Conference on Artificial Intelligence and Cognitive Science, 2000, AAAI Press, Menlo Park, CA*. Press, 2000. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.137.6488

[40] J. Lindblom and T. Ziemke, “Social Situatedness of Natural and Artificial Intelligence: Vygotsky and Beyond,” *Adaptive Behavior*, vol. 11, no. 2, pp. 79–96, 2003. [Online]. Available: https://doi.org/10.1177/10597123030112002

[41] K. J. Rohlfing, “Situatedness: The interplay between context (s) and situation,” *Journal of Cognition and Culture*, vol. 2, no. 2, pp. 132–157, 2003. [Online]. Available: http://www.ingentaconnect.com/content/brill/jocc/2003/00000003/00000002/art00002

[42] M. Costello, “Situatedness,” in *Encyclopedia of Critical Psychology*, T. Teo, Ed. Springer New York, 2014, pp. 1757–1762.

[43] A. Paiva, J. Dias, D. Sobral, R. Aylett, S. Woods, L. Hall, and C. Zoll, “Learning by feeling: Evoking empathy with synthetic characters,” *Applied Artificial Intelligence*, vol. 19, no. 3-4, pp. 235–266, 2005.

[44] M. A. Hogg and R. S. Tindale, *Blackwell Handbook of Social Psychology: Group Processes*, M. A. Hogg and R. S. Tindale, Eds. Oxford, UK: Blackwell Publishers Ltd, 1 2001. [Online]. Available: https://onlinelibrary.wiley.com/doi/book/10.1002/9780470998458http://doi.wiley.com/10.1002/9780470998458

[45] G. Pereira, J. Dimas, R. Prada, P. A. Santos, and A. Paiva, “A Generic Emotional Contagion Computational Model,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2011, vol. 6974 LNCS, no. PART 1, pp. 256–266. [Online]. Available: http://link.springer.com/10.1007/978-3-642-24600-5_29

[46] K. Niven, P. Totterdell, and D. Holman, “A Classification of Controlled Interpersonal Affect Regulation Strategies,” *Emotion*, vol. 9, no. 4, pp. 498–509, 2009.

[47] P. Gaussier, S. Moga, M. Quoy, and J. P. Banquet, “From perception-action loops to imitation processes: a bottom-up approach of learning by imitation,” *Applied Artificial Intelligence*, vol. 12, no. 7-8, pp. 701–727, 10 1998. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/088395198117596

[48] A. I. Goldman, *Theory of Mind*, E. Margolis, R. Samuels, and S. P. Stich, Eds. Oxford University Press, 5 2012, vol. 1. [Online]. Available: http://oxfordhandbooks.com/view/10.1093/oxfordhb/9780195309799.001.0001/oxfordhb-9780195309799-e-17

[49] D. Premack and G. Woodruff, “Does the chimpanzee have a theory of mind?” *The Behabioral and Brain Sciences*, 1978.

[50] M. Si, S. C. Marsella, and D. V. Pynadath, “Modeling Appraisal in Theory of Mind Reasoning,” in *Intelligent Virtual Agents*, H. Prendinger, J. Lester, and M. Ishizuka, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 334–347.

[51] M. L. Hoffman and Z. Raboteg-Saric, “Empathy and Moral Development: Implications for Caring and Justice,” *Contemporary Sociology*, vol. 30, no. 5, p. 487, 2001. [Online]. Available: http://www.jstor.org/stable/3089337?origin=crossref

[52] S. Hamdy and D. King, “Affect and believability in game characters - A review of the use of affective computing in games,” in *18th International Conference on Intelligent Games and Simulation, GAME-ON 2017*. EUROSIS, 2017, pp. 90–97.

[53] H. Warpefelt, M. Johansson, and H. Verhagen, “Analyzing the believability of game character behavior using the Game Agent Matrix,” in *Proceedings of the Seventh DiGRA International Conference: DeFragging Game Studies*. Digital Games Research Association, 2013.

[54] G. N. Yannakakis, J. Hallam, and H. H. Lund, “Entertainment capture through heart rate activity in physical interactive playgrounds,” *User Modeling and User-Adapted Interaction 2007 18:1*, vol. 18, no. 1, pp. 207–243, 9 2007. [Online]. Available: https://link.springer.com/article/10.1007/s11257-007-9036-7

[55] P. Rodrigues, A. Paiva, and J. Dias, “An Affect-Aware Intelligent Tutoring System for EmoRegulators A Restricted-Perception Wizard-of-Oz Approach,” p. 102, 2018.

[56] C. Pedersen, J. Togelius, and G. N. Yannakakis, “Modeling player experience for content creation,” *IEEE Transactions on Computational Intelligence and AI in Games*, vol. 2, no. 1, pp. 54–67, 3 2010.

[57] J. Pardal and C. Martinho, “Holiday Knight: a Videogame with Skill-based Challenge Generation,” p. 92, 2019. [Online]. Available: https://fenix.tecnico.ulisboa.pt/cursos/meic-t/dissertacao/283828618790385

[58] M. Marques and C. Martinho, “Profiling Players Through In-Game Animations,” p. 80, 2019.

[59] A. Ortony, G. L. Clore, and A. Collins, *The Cognitive Structure of Emotions*. Cambridge, UK: Cambridge University Press, 1990.

[60] N. Costa, “Believable Interactions Between Synthetic Characters,” Master’s thesis, Instituto Superior Técnico, University of Lisbon, 2015. [Online]. Available: https://fenix.tecnico.ulisboa.pt/cursos/meic-t/dissertacao/565303595500067

[61] A. Beck, B. Stevens, K. A. Bard, and L. Cañamero, “Emotional body language displayed by artificial agents,” *ACM Transactions on Interactive Intelligent Systems*, vol. 2, no. 1, pp. 1–29, 3 2012. [Online]. Available: https://dl.acm.org/doi/abs/10.1145/2133366.2133368https://dl.acm.org/doi/10.1145/2133366.2133368

[62] P. R. Kleinginna and A. M. Kleinginna, “A categorized list of emotion definitions, with suggestions for a consensual definition,” *Motivation and Emotion*, vol. 5, no. 4, pp. 345–379, 12 1981.

[63] K. R. Scherer, T. Banziger, and E. B. Roesch, *Blueprint for affective computing : a sourcebook*, 1st ed. New York, NY, USA: Oxford University Press, Inc., 2010.

[64] P. Ekman, *Emotions revealed: Recognizing faces and feelings to improve communication and emotional life*. New York, New York, USA: Macmillan, 2007.

[65] P. Ekman and W. V. Friesen, *Unmasking the face: A guide to recognizing emotions from facial clues*. SAN JOSE CA USA: Ishk, 2003.

[66] L. F. Barrett, *How emotions are made: The secret life of the brain.* Boston, USA: Pan MacMillan, 2017.

[67] P. Ekman, "An argument for basic emotions," *Cognition & emotion*, vol. 6, no. 3-4, pp. 169–200, 1992.

[68] A. Arya, S. DiPaola, and A. Parush, "Perceptually Valid Facial Expressions for Character-Based Applications," *International Journal of Computer Games Technology*, vol. 2009, pp. 1–13, 2009. [Online]. Available: http://www.hindawi.com/journals/ijcgt/2009/462315/

[69] M. Mäkäräinen and T. Takala, "An Approach for Creating and Blending Synthetic Facial Expressions of Emotion." in *Intelligent Virtual Agents*, Springer. Netherlands: Springer, 2009, pp. 243–249.

[70] R. Rodrigues, P. Ferreira, R. Prada, P. Paulino, and A. M. V. Simao, "Festarola: a Game for Improving Problem Solving Strategies," in *2019 11th International Conference on Virtual Worlds and Games for Serious Applications (VS-Games)*. IEEE, 9 2019, pp. 1–8. [Online]. Available: https://ieeexplore.ieee.org/document/8864524/

[71] C. Crivelli and A. J. Fridlund, "Inside-Out: From Basic Emotions Theory to the Behavioral Ecology View," *Journal of Nonverbal Behavior*, vol. 43, no. 2, p. 161–194, 2019.

[72] M. Miceli and C. Castelfranchi, *Expectancy and emotion*. Oxford University Press, 2015. [Online]. Available: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=psyc11&AN=2015-04579-000%5Cnhttp://dc8qa4cy3n.search.serialssolutions.com/?url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:journal&rfr_id=info:sid/Ovid:psyc11&rft.genre=article&rft_

[73] P. Pimentel, "Gameplay Aware Expression for Virtual Characters in Video Games," 2015.

[74] J. Dias and A. Paiva, "Feeling and Reasoning: A Computational Model for Emotional Characters," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, ser. Lecture Notes in Computer Science, C. Bento, A. Cardoso, and G. Dias, Eds. Springer Berlin Heidelberg, 2005, vol. 3808 LNCS, pp. 127–140. [Online]. Available: http://link.springer.com/10.1007/11595014_13

[75] S. He, Z. Liu, and W. Xiong, "An emotion model for virtual agent," in *Proceedings of the 2008 International Symposium on Computational Intelligence and Design, ISCID 2008*, vol. 1, 10 2008, pp. 511–514.

[76] E. Jacobs, J. Broekens, and C. Jonker, "Joy, distress, hope, and fear in reinforcement learning," in *13th International Conference on Autonomous Agents and*

Multiagent Systems, AAMAS 2014, ser. AAMAS '14, vol. 2. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems, 2014, pp. 1615–1616. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84911369211&partnerID=40&md5=cf583e64ab4a206b7f03d9ef9b80df11

[77] E. Doirado, "DogMate: Intent Recognition through Anticipation," 2009.

[78] I. Leite, "Long-term Interactions with Empathic Social Robots," Ph.D. dissertation, Instituto Superior Técnico, 2013.

[79] M. Lewis, J. M. Haviland-Jones, and L. F. Barrett, *Handbook of emotions*. New York, New York, USA: Guilford Press, 2010.

[80] J. Lee, Z. Wang, and S. Marsella, "Evaluating models of speaker head nods for virtual agents," in *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems*, International Foundation for Autonomous Agents and Multiagent Systems. Richland, SC: ACM, 2010, pp. 1257–1264.

[81] T. Dahle, "Culture and 3D animation : A study of how culture and body language affects the perception of animated 3D characters," Ph.D. dissertation, University of Skövde, 2019. [Online]. Available: http://urn.kb.se/resolve?urn=urn:nbn:se:his:diva-17242

[82] J. Patrício, "Conveying Believable Emotions Through Control Modulation," 2021. [Online]. Available: https://fenix.tecnico.ulisboa.pt/cursos/meic-a/dissertacao/1691203502344356

[83] P. Ekman and W. Friesen, "Facial action coding system: A technique for the measurement of facial movement." in *Differences among unpleasant feelings. Motivation and Emotion*. New York, New York, USA: Plenum Publishing Corp, 1978.

[84] S. McCloud, *Making Comics: Storytelling Secrets of Comics, Manga and Graphic Novels*. HarperCollins Publishers, 2006.

[85] S. Andrist, T. Pejsa, B. Mutlu, and M. Gleicher, "Designing effective gaze mechanisms for virtual agents," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM. Austin Texas USA: ACM, 2012, pp. 705–714.

[86] B. Lance and S. C. Marsella, "Emotionally expressive head and body movement during gaze shifts," in *International Workshop on Intelligent Virtual Agents*, Springer. Paris, France: Springer, 2007, pp. 72–85.

[87] C. Busso, Z. Deng, M. Grimm, U. Neumann, and S. Narayanan, "Rigid head motion in expressive speech animation: Analysis and synthesis," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 3, pp. 1075–1086, 2007.

[88] H. K. M. Meeren, C. C. R. J. van Heijnsbergen, and B. de Gelder, "Rapid perceptual integration of facial expression and emotional body language," *Proceedings*

of the National Academy of Sciences, vol. 102, no. 45, pp. 16 518–16 523, 11 2005. [Online]. Available: https://www.pnas.org/content/102/45/16518https://www.pnas.org/content/102/45/16518.abstracthttp://www.pnas.org/cgi/doi/10.1073/pnas.0507650102

[89] N. Kaya and H. H. Epps, "Color-emotion associations: Past experience and personal preference," in *AIC 2004 Color and Paints, Interim Meeting of the In ternational Color Association, Proceedings*. Porto Alegre, Brazil: Association Internationale de la Couleur, 2004, pp. 31–34.

[90] N. Kaya and H. H. Epps, "Relationship between Color and Emotion: A Study of College Students," *College student journal*, vol. 38, p. 396, 2004.

[91] C. M. de Melo and J. Gratch, "The effect of color on expression of joy and sadness in virtual humans," in *2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*. Netherlands: IEEE, 2009, pp. 1–7. [Online]. Available: http://ieeexplore.ieee.org/document/5349585/

[92] R. Plutchik and H. Kellerman, *Theories of Emotion*. Elsevier, 1980. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/C2013011313X

[93] J. Forlizzi, J. Lee, and S. Hudson, "The kinedit system," in *Proceedings of the conference on Human factors in computing systems - CHI '03*. New York, New York, USA: ACM Press, 2003, p. 377. [Online]. Available: http://portal.acm.org/citation.cfm?doid=642611.642677

[94] S. Malik, J. Aitken, and J. K. Waalen, "Communicating emotion with animated text," *Visual Communication*, vol. 8, no. 4, pp. 469–479, 2009. [Online]. Available: http://journals.sagepub.com/doi/10.1177/1470357209343375

[95] S. McCloud, *Understanding Comics: The Invisible Art*. New York, New York, USA: Harper-Perennial, 1993.

[96] N. S. Schutte, J. M. Malouff, L. E. Hall, D. J. Haggerty, J. T. Cooper, C. J. Golden, and L. Dornheim, "Development and validation of a measure of emotional intelligence," *Personality and Individual Differences*, vol. 25, no. 2, pp. 167–177, 8 1998. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0191886998000014

[97] P. Salovey and J. D. Mayer, "Emotional Intelligence," *Imagination, Cognition and Personality*, vol. 9, no. 3, pp. 185–211, 1990. [Online]. Available: http://journals.sagepub.com/doi/10.2190/DUGG-P24E-52WK-6CDG

[98] A. Mill, J. Allik, A. Realo, and R. Valk, "Age-related differences in emotion recognition ability: A cross-sectional study." *Emotion*, vol. 9, no. 5, pp. 619–630, 2009. [Online]. Available: http://doi.apa.org/getdoi.cfm?doi=10.1037/a0016562

[99] J. N. Bassili, "Emotion recognition: The role of facial movement and the relative importance of upper and lower areas of the face." *Journal of Personality and Social Psychology*, vol. 37,

no. 11, pp. 2049–2058, 1979. [Online]. Available: http://content.apa.org/journals/psp/37/11/2049

[100] E. Mower, M. J. Mataric, and S. Narayanan, "Human Perception of Audio-Visual Synthetic Character Emotion Expression in the Presence of Ambiguous and Conflicting Information," *IEEE Transactions on Multimedia*, vol. 11, no. 5, pp. 843–855, 2009. [Online]. Available: http://ieeexplore.ieee.org/document/4907039/

[101] D. Lottridge, M. Chignell, and A. Jovicic, "Affective Interaction: Understanding, Evaluating, and Designing for Human Emotion," *Reviews of Human Factors and Ergonomics*, vol. 7, no. 1, pp. 197–217, 2011.

[102] I. Lefter, C. M. Jonker, S. K. Tuente, W. Veling, and S. Bogaerts, "NAA: A multimodal database of negative affect and aggression," in *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*. San Antonio, Texas, USA: IEEE, 2017, pp. 21–27.

[103] B. Cheng, Z. Wang, Z. Zhang, Z. Li, D. Liu, J. Yang, S. Huang, and T. S. Huang, "Robust emotion recognition from low quality and low bit rate video: A deep learning approach," in *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*. San Antonio, Texas, USA: IEEE, 2017, pp. 65–70.

[104] G. Chanel, S. Avry, G. Molinari, M. Bétrancourt, and T. Pun, "Multiple users' emotion recognition: Improving performance by joint modeling of affective reactions," in *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*. San Antonio, Texas, USA: IEEE, 2017, pp. 92–97.

[105] L. Tian, M. Muszynski, C. Lai, J. D. Moore, T. Kostoulas, P. Lombardo, T. Pun, and G. Chanel, "Recognizing induced emotions of movie audiences: Are induced and perceived emotions the same?" in *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*. San Antonio, Texas, USA: IEEE, 2017, pp. 28–35.

[106] S. Ghosh, N. Ganguly, B. Mitra, and P. De, "Evaluating effectiveness of smartphone typing as an indicator of user emotion," in *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*. San Antonio, Texas, USA: IEEE, 2017, pp. 146–151.

[107] S. Afzal, B. Sengupta, M. Syed, N. Chawla, G. A. Ambrose, and M. Chetlur, "The ABC of MOOCs: Affect and its inter-play with behavior and cognition," in *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*. San Antonio, Texas, USA: IEEE, 2017, pp. 279–284.

[108] S. Kumano, R. Ishii, and K. Otsuka, "Comparing empathy perceived by interlocutors in multiparty conversation and external observers," in *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*. San Antonio, Texas, USA: IEEE, 2017, pp. 50–57.

[109] ——, "Computational model of idiosyncratic perception of others' emotions," in *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*. San Antonio, Texas, USA: IEEE, 2017, pp. 42–49.

[110] C. Sikora and W. Burleson, "The dance of emotion: Demonstrating ubiquitous understanding of human motion and emotion in support of human computer interaction," in *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*. San Antonio, Texas, USA: IEEE, 2017, pp. 548–555.

[111] Y. Ding, L. Shi, and Z. Deng, "Perceptual enhancement of emotional mocap head motion: An experimental study," in *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*. San Antonio, Texas, USA: IEEE, 2017, pp. 242–247.

[112] X. Lu, R. B. Adams, J. Li, M. G. Newman, and J. Z. Wang, "An investigation into three visual characteristics of complex scenes that evoke human emotion," in *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*. San Antonio, Texas, USA: IEEE, 2017, pp. 440–447.

[113] A. Hernández-García, "Perceived emotion from images through deep neural networks," in *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*. San Antonio, Texas, USA: IEEE, 2017, pp. 566–570.

[114] N. V. Mudrick, M. Taub, R. Azevedo, J. Rowe, and J. Lester, "Toward affect-sensitive virtual human tutors: The influence of facial expressions on learning and emotion," in *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*. San Antonio, Texas, USA: IEEE, 2017, pp. 184–189.

[115] J. Deanna Bailey, "Avatar and participant gender differences in the perception of uncanniness of virtual humans," in *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*. San Antonio, Texas, USA: IEEE, 2017, pp. 571–575.

[116] K. R. Scherer, "What are emotions? And how can they be measured?" *Social Science Information*, vol. 44, no. 4, pp. 695–729, 12 2005. [Online]. Available: https://journals.sagepub.com/doi/10.1177/0539018405058216

[117] D. Watson, L. A. Clark, and A. Tellegen, "Development and validation of brief measures of positive and negative affect: The PANAS scales." *Journal of Personality and Social Psychology*, vol. 54, no. 6, pp. 1063–1070, 1988. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/3397865/http://doi.apa.org/getdoi.cfm?doi=10.1037/0022-3514.54.6.1063

[118] E. Grossman, M. Donnelly, R. Price, D. Pickens, V. Morgan, G. Neighbor, and R. Blake, "Brain Areas Involved in Perception of Biological Motion," *Journal of Cognitive Neuroscience*, vol. 12, no. 5, pp. 711–720, 9 2000. [Online]. Available: http://mitprc.silverchair.com/jocn/article-pdf/12/5/711/1758799/089892900562417.pdfhttps://direct.mit.edu/jocn/article/12/5/711/3471/Brain-Areas-Involved-in-Perception-of-Biological

[119] R. McDonnell, “Appealing Virtual Humans,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Springer, Berlin, Heidelberg, 11 2012, vol. 7660 LNCS, pp. 102–111. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-642-34710-8_10http://link.springer.com/10.1007/978-3-642-34710-8_10

[120] J. Lasseter, “Principles of traditional animation applied to 3D computer animation,” *ACM SIGGRAPH Computer Graphics*, vol. 21, no. 4, pp. 35–44, 1987.

[121] C. Sun, B. Wu, R. Wang, X. Hu, X. Yang, and C. Cong, “Intrinsic motivated multi-agent communication,” in *Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS*, vol. 3. Online, 2021, pp. 1656–1658. [Online]. Available: www.ifaamas.org

[122] W. Al Enezi and C. Verbrugge, “Skeleton-based multi-agent opponent search,” in *IEEE Conference on Computatonal Intelligence and Games, CIG*, vol. 2021-Augus, 2021.

[123] A. Hosokawa and J. Morita, “Cognitive models leading to behavior-reading in a card game,” in *HAI 2021 - Proceedings of the 9th International User Modeling, Adaptation and Personalization Human-Agent Interaction*. Association for Computing Machinery, Inc, 11 2021, pp. 337–341.

[124] A. R. Maina-Kilaas, G. D. Montanez, C. Hom, K. Ginta, and C. Lay, “The Hero’s Dilemma: Survival Advantages of Intention Perception in Virtual Agent Games,” in *IEEE Conference on Computatonal Intelligence and Games, CIG*, vol. 2021-Augus, 2021.

[125] S. G. Ware and C. Siler, “The sabre narrative planner: Multi-agent coordination with intentions and beliefs,” in *Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS*, vol. 3, 2021, pp. 1686–1688. [Online]. Available: www.ifaamas.org

[126] S. Makri and P. Charalambous, “Towards a multi-agent non-player character road network: A Reinforcement Learning approach,” in *IEEE Conference on Computatonal Intelligence and Games, CIG*, vol. 2021-Augus, 2021. [Online]. Available: https://youtu.be/-qwFddCkyDo.

[127] T. Shintani, C. T. Ishi, and H. Ishiguro, “Analysis of role-based gaze behaviors and gaze aversions, and implementation of robot’s gaze control for multi-party dialogue,” in *HAI 2021 - Proceedings of the 9th International User Modeling, Adaptation and Personalization Human-Agent Interaction*. Association for Computing Machinery, Inc, 11 2021, pp. 332–336.

[128] D. M. Olson and Y. Xu, “Building Trust over Time in Human-Agent Relationships,” in *HAI 2021 - Proceedings of the 9th International User Modeling, Adaptation and Personalization Human-Agent Interaction*. Association for Computing Machinery, Inc, 11 2021, pp. 193–201. [Online]. Available: https://doi.org/10.1145/3472307.3484178

[129] T. Schneeberger, M. Scholtes, B. Hilpert, M. Langer, and P. Gebhard, "Can Social Agents elicit Shame as Humans do?" *2019 8th International Conference on Affective Computing and Intelligent Interaction, ACII 2019*, 2019.

[130] L. A. Dennis and N. Oren, "Explaining BDI agent behaviour through dialogue," in *Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS*, vol. 1, 2021, pp. 429–437. [Online]. Available: www.ifaamas.org

[131] I. Torre, S. Tuncer, D. McDuff, and M. Czerwinski, "Exploring the Effects of Virtual Agents' Smiles on Human-Agent Interaction: A Mixed-Methods Study," in *2021 9th International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 9 2021, pp. 1–8. [Online]. Available: https://ieeexplore.ieee.org/document/9597445/

[132] L. Yang, C. Achard, and C. Pelachaud, "Interruptions in Human-Agent Interaction," in *Proceedings of the 21st ACM International Conference on Intelligent Virtual Agents, IVA 2021*, vol. 1, no. 1. Association for Computing Machinery, 2021, pp. 206–208.

[133] T. Takahashi, K. Tanaka, K. Kobayashi, and N. Oka, "Melodic Emotional Expression Increases Ease of Talking to Spoken Dialog Agents," in *Proceedings of the 9th International Conference on Human-Agent Interaction*. New York, NY, USA: ACM, 11 2021, pp. 84–92. [Online]. Available: https://doi.org/10.1145/3472307.3484180https://dl.acm.org/doi/10.1145/3472307.3484180

[134] E. B. Cloude, F. Wortha, D. A. Dever, and R. Azevedo, "Negative emotional dynamics shape cognition and performance with MetaTutor: Toward building affect-aware systems," in *2021 9th International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 9 2021, pp. 1–8. [Online]. Available: https://ieeexplore.ieee.org/document/9597462/

[135] R. Asraf, C. Rozenshtein, and D. Sarne, "On the effect of user faults on her perception of agents faults in collaborative settings," in *HAI 2021 - Proceedings of the 9th International User Modeling, Adaptation and Personalization Human-Agent Interaction*. Association for Computing Machinery, Inc, 11 2021, pp. 372–376.

[136] Q. Roy, M. Ghafurian, W. Li, and J. Hoey, "Users, Tasks, and Conversational Agents: A Personality Study," in *HAI 2021 - Proceedings of the 9th International User Modeling, Adaptation and Personalization Human-Agent Interaction*. Association for Computing Machinery, Inc, 11 2021, pp. 174–182.

[137] D. S. Carvalho, J. Campos, M. Guimarães, A. Antunes, J. Dias, and P. A. Santos, "CHARET: Character-centered approach to emotion tracking in stories," in *Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS*, vol. 3, 2021, pp. 1457–1459. [Online]. Available: www.ifaamas.org

[138] A. Antunes, J. Campos, J. Dias, P. A. Santos, and R. Prada, "EEG Model: Emotional Episode Generation for Social Sharing of Emotions," in *Proceedings of the 21th ACM International*

Conference on Intelligent Virtual Agents, vol. 1, no. 1. New York, NY, USA: ACM, 9 2021, pp. 1–8. [Online]. Available: https://dl.acm.org/doi/10.1145/3472306.3478342

[139] D. Rato, M. Couto, and R. Prada, "Fitting the Room: Social Motivations for Context-Aware Agents," in *Proceedings of the 9th International Conference on Human-Agent Interaction*. New York, NY, USA: ACM, 11 2021, pp. 39–46. [Online]. Available: https://doi.org/10.1145/3472307.3484185https://dl.acm.org/doi/10.1145/3472307.3484185

[140] P. Raimbaud, A. Jovane, K. Zibrek, C. Pacchierotti, M. Christie, L. Hoyet, J. Pettré, and A. H. Olivier, "Reactive Virtual Agents: A Viewpoint-Driven Approach for Bodily Nonverbal Communication," in *Proceedings of the 21st ACM International Conference on Intelligent Virtual Agents, IVA 2021*, 2021, pp. 164–166. [Online]. Available: https://doi.org/10.1145/3472306.3478351

[141] T. Beinema, D. Davison, D. Reidsma, O. Banos, M. Bruijnes, B. Donval, A. F. Valero, D. Heylen, D. Hofs, G. Huizing, R. B. Kantharaju, R. Klaassen, J. Kolkmeier, K. Konsolakis, A. Pease, C. Pelachaud, D. Simonetti, M. Snaith, V. Traver, J. Van Loon, J. Visser, M. Weusthof, F. Yunus, H. Hermens, and H. Op Den Akker, "Agents United: An Open Platform for Multi-Agent Conversational Systems," in *Proceedings of the 21st ACM International Conference on Intelligent Virtual Agents, IVA 2021*. New York, NY, USA: ACM, 9 2021, pp. 17–24. [Online]. Available: https://dl.acm.org/doi/10.1145/3472306.3478352

[142] R. Rodrigues, "Expectancy and Emotions in Synthetic Characters," Master's thesis, Instituto Superior Técnico, University of Lisbon, 2016. [Online]. Available: https://fenix.tecnico.ulisboa.pt/cursos/meic-t/dissertacao/1972678479053108

[143] B. G. Breitmeyer and L. Ganz, "Implications of sustained and transient channels for theories of visual pattern masking, saccadic suppression, and information processing." *Psychological Review*, vol. 83, no. 1, pp. 1–36, 1 1976. [Online]. Available: /record/1979-20169-001http://doi.apa.org/getdoi.cfm?doi=10.1037/0033-295X.83.1.1

[144] M. Hibbeln, J. L. Jenkins, C. Schneider, J. S. Valacich, and M. Weinmann, "How Is Your User Feeling? Inferring Emotion Through Human-Computer interaction Devices," *MIS Quarterly*, vol. 41, no. 1, pp. 1–21, 1 2017. [Online]. Available: https://misq.org/how-is-your-user-feeling-inferring-emotion-through-human-computer-interaction-devices.html

[145] T. Ribeiro, "Perception of Emotion Expression in Synthetic Characters," Master's thesis, Instituto Superior Técnico, University of Lisbon, 2021. [Online]. Available: https://fenix.tecnico.ulisboa.pt/cursos/meic-a/dissertacao/1691203502344371

[146] R. Pereira, "Expressing Emotions through Animated Speech Balloons," Master's thesis, Instituto Superior Técnico, University of Lisbon, 2018. [Online]. Available: https://fenix.tecnico.ulisboa.pt/cursos/meic-t/dissertacao/1972678479054259

[147] R. Silva, “Supporting Affective Expression in Multi-party Interactions,” Master’s thesis, Instituto Superior Técnico, University of Lisbon, 2018. [Online]. Available: https://fenix.tecnico.ulisboa.pt/cursos/meic-t/dissertacao/1409728525632118

[148] R. Garrison, “Implications of Online and Blended Learning for the Conceptual Development and Practice of Distance Education,” 2009.

[149] T. Anderson and J. Dron, “Three generations of distance education pedagogy,” *International Review of Research in Open and Distance Learning*, vol. 12, pp. 80–97, 2011.

[150] A. Andrade, F. Nogueira, C. V. de Carvalho, P. Russo, and D. Gouveia, “Games to support problem-based learning,” in *2015 10th Iberian Conference on Information Systems and Technologies (CISTI)*. New Jersey, NJ, USA: IEEE, 2015, pp. 1–4.

[151] R. Leitão, J. M. F. Rodrigues, and A. F. Marcos, “Game-based learning: augmented reality in the teaching of geometric solids,” *International Journal of Art, Culture and Design Technologies (IJACDT)*, vol. 4, pp. 65–77, 2014.

[152] P. Rodrígues and J. Bidarra, “Transmedia storytelling and the creation of a converging space of educational practices,” *International Journal of Emerging Technologies in Learning*, vol. 9, p. 42, 2014.

[153] B. Cooper, P. Brna, and A. Martins, “Effective affective in intelligent systems–building on evidence of empathy in teaching and learning,” *Affective Interactions Towards a New Generation of Computer Interfaces*, vol. 1814/2000, p. 21–34, 2000. [Online]. Available: http://www.springerlink.com/index/j8v0l230t3503367.pdf

[154] F. Heider and M. Simmel, “An experimental study of apparent behaviour,” *The American Journal of Psychology*, vol. 57, no. 2, pp. 243–259, 1944. [Online]. Available: http://www.jstor.org/stable/1416950?origin=crossref%5Cnhttp://www.jstor.org/stable/10.2307/1416950

[155] N. D. Feshbach, “Studies of empathic behavior in children,” *Progress in experimental personality research*, vol. 8, p. 1–47, 1978.

[156] K. Niven, D. Holman, and P. Totterdell, “How to win friendship and trust by influencing people’s feelings: An investigation of interpersonal affect regulation and the quality of relationships,” *Human Relations*, vol. 65, no. 6, pp. 777–805, 2012.

[157] C. Reeck, D. R. Ames, and K. N. Ochsner, “The Social Regulation of Emotion: An Integrative, Cross-Disciplinary Model,” pp. 47–63, 2016.

[158] A. Pires, P. Alves-Oliveira, P. Arriaga, and C. Martinho, “Cubus: Autonomous Embodied Characters to Stimulate Creative Idea Generation in Groups of Children,” in *Proceedings of the 17th International Conference on Intelligent Virtual Agents - IVA 2017*. Cham, Switzerland: Springer International Publishing, 2017, pp. 360–373. [Online]. Available: http://link.springer.com/10.1007/978-3-319-67401-8_46

[159] J. Brooke, "SUS - A quick and dirty usability scale," in *Usability Evaluation in Industry*. Taylor & Francis Group, 2&4 Park Square, Milton Park, Abingdon, OX14 4RN: London: Taylor and Francis, 1996.

[160] ——, "SUS: A Retrospective," *Journal of Usability Studies*, vol. 8, no. 2, pp. 29–40, 2013.

[161] V. Sacharin, K. Schlegel, and K. R. Scherer, "Geneva Emotion Wheel Rating Study," 2012. [Online]. Available: https://archive-ouverte.unige.ch/unige:97849

[162] K. Goffin, "Repertory grid technique," *Essential Skills for Management Research*, vol. 1, pp. 199–225, 2002.

[163] R. Bell, D. Bannister, and F. Fransella, *A Manual for Repertory Grid Technique*. New York, New York, USA: John Wiley & Sons, Ltd, 2004.

[164] F. B. Tan and M. G. Hunter, "The Repertory Grid Technique: A Method for the Study of Cognition in Information Systems," *MIS Quarterly*, vol. 26, no. 1, p. 39, 3 2002. [Online]. Available: https://www.jstor.org/stable/10.2307/4132340?origin=crossref

[165] A. M. Curtis, T. M. Wells, P. B. Lowry, and T. Higbee, "An overview and tutorial of the repertory grid technique in information systems research," *Communications of the Association for Information Systems*, vol. 23, no. 1, p. 3, 2008.

[166] P. Samuels, "Advice on Exploratory Factor Analysis," p. 8, 2016.

[167] M. Wegrzyn, M. Vogt, B. Kireclioglu, J. Schneider, and J. Kissler, "Mapping the emotional face. How individual face parts contribute to successful emotion recognition," *PLOS ONE*, vol. 12, no. 5, p. e0177239, 5 2017. [Online]. Available: https://doi.org/10.1371/journal.pone.0177239https://dx.plos.org/10.1371/journal.pone.0177239

[168] A. R. Dores, F. Barbosa, C. Queirós, I. P. Carvalho, and M. D. Griffiths, "Recognizing emotions through facial expressions: a largescale experimental study," *International journal of environmental research and public health*, vol. 17, no. 20, p. 7420, 2020.

[169] O. Langner, R. Dotsch, G. Bijlstra, D. H. J. Wigboldus, S. T. Hawk, and A. D. Van Knippenberg, "Presentation and validation of the Radboud Faces Database," *Cognition and emotion*, vol. 24, no. 8, pp. 1377–1388, 2010.

[170] J. Patrício, "Conveying Believable Emotions Through Control Modulation," 2021. [Online]. Available: https://fenix.tecnico.ulisboa.pt/cursos/meic-a/dissertacao/1691203502344356

[171] A. Field, *Discovering Statistics using SPSS Statistics Third Edition*. SAGE Publications Inc., 2009.

[172] V. Nowlis and H. H. Nowlis, "THE DESCRIPTION AND ANALYSIS OF MOOD," *Annals of the New York Academy of Sciences*, vol. 65, no. 4, pp. 345–355, 1956.

[173] H. Schlosberg, “Three dimensions of emotion,” *Psychological Review*, 1954.

[174] S. L. Franconeri, A. Hollingworth, and D. J. Simons, “Do New Objects Capture Attention?” *Psychological Science*, vol. 16, no. 4, pp. 275–281, 4 2005. [Online]. Available: https://journals.sagepub.com/doi/abs/10.1111/j.0956-7976.2005.01528.xhttp://journals.sagepub.com/doi/10.1111/j.0956-7976.2005.01528.x

[175] F. I. Parke, “Perception-based animation rendering,” *The Journal of Visualization and Computer Animation*, vol. 2, no. 2, pp. 44–51, 4 1991. [Online]. Available: https://onlinelibrary.wiley.com/doi/full/10.1002/vis.4340020204https://onlinelibrary.wiley.com/doi/abs/10.1002/vis.4340020204https://onlinelibrary.wiley.com/doi/10.1002/vis.4340020204

[176] T. Calinski and J. Harabasz, “A dendrite method for cluster analysis,” *Communications in Statistics - Theory and Methods*, vol. 3, no. 1, pp. 1–27, 1974. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/03610927408827101

Video Games and Media References

[177] *Halo Infinite* [Video game] (2021). 343 Industries, Xbox Game Studios.

[178] *Guild Wars 2* [Video game] (2012). ArenaNet, NCSoft.

[179] *Warcraft III: Reign of Chaos* [Video game] (2002). Blizzard Entertainment.

[180] *StarCraft II* [Video game] (2010). Blizzard Entertainment.

[181] *Overwatch* [Video game] (2016). Blizzard Entertainment.

[182] *The Witcher 3: Wild Hunt* [Video game] (2015). CD PROJEKT RED.

[183] *Tomb Raider* [Video game] (2013). Crystal Dynamics, Square Enix.

[184] *Battlefield V* [Video game] (2018). DICE, Electronic Arts.

[185] *Horizon Zero Dawn* [Video game] (2020). Guerrilla, PlayStation PC LLC.

[186] *Doom* [Video game] (1993). id Software.

[187] *Fable II* [Video game] (2008). Lionhead Studios, Microsoft Game Studios.

[188] *The Lord of the Rings* (2003). New Line Cinema.

[189] *Darkest Dungeon* [Video game] (2016). Red Hook Studios.

[190] *Call of Duty: WWII* [Video game] (2017). Sledgehammer Games, Raven Software, Activision.

[191] *Assassin's Creed Odyssey* [Video game] (2018). Ubisoft.

[192] *Aladdin* (1992). Walt Disney Pictures, Walt Disney Studios Motion Pictures.

[193] *Inside Out* (2015). Walt Disney Pictures, Pixar Animation Studios, Walt Disney Studios Motion Pictures.

A

Additional Related Work

Contents

A.1 Notes on Metrics for Character Believability

In Gomes *et al.*[37], the audience's perception is asserted using Likert scales, one scale per dimension. The templates for the phrases to be rated, except for emotional expressiveness, by the subjects are:

- awareness: <***X***> perceives the world around him/her.
- behavior understandability: It is easy to understand what <***X***> is thinking about.
- personality: <***X***> has a personality.
- visual impact: <***X***>'s behavior draws my attention.
- predictability: <***X***>'s behavior is predictable.
- behavior coherence: <***X***>'s behavior is coherent.
- change with experience: <***X***>'s behavior changes according to experience.
- social: <***X***> interacts socially with other characters.

A.2 Summary of Affective Models Categories

In [63], the author describes five categories that incorporate different affective models: *Appraisal Theory Approaches*; *Anatomical Approaches*; *Rational Approaches*; *Communicative Approaches*; and, *Dimensional Theory Approaches*. Here follows a small description of each approach.

Anatomical Approaches try to emulate the neural structure that is behind an emotional response. They tend to specialize in a single emotion, since they see emotions as separate entities with their own systems, giving great importance to the systems that create the emotion. Although a very detailed approach, they focus on a more raw and basic emotional response and tend to specialize in only one emotion.

Rational Approaches "ponder what adaptive function does emotion serve"[73], attempting to incorporate an abstract version of an emotion from its implementation in humans into a model of intelligence. This approach is typically associated with artificial intelligence research, where models using this approach are usually used to further develop machine intelligence theories.

Communicative Approaches focus on the social component of emotion, that "serves an empathic objective to aid in communication and to transmit non-verbal cues"[73]. This approach is more usually used in social studies, crowd dynamics, and multi-agent systems, focusing on the outward emotional display, often simplifying internal workings for creating an emotion.

Dimensional Theory Approaches focuses on the view that emotions shouldn't be viewed as discrete and unrelated – a concept introduced by Nowlis *et al.*[172] showing that there were between six and twelve independent affective states (ie. sadness, anger, anxiety, etc.) – but as an end product of a system of undisclosed variables[173], where affective states are connected and their origins is an n-dimensional vector. Furthermore, "a big advantage of the dimensional approach is one can attempt to code the seemingly complex nature of human emotions as a combination of simpler internal factors"[73].

Appraisal Theory Approaches postulates that "all emotions come mostly from our own interpretations of events"[73], where our appraisal of the situation is the emotional response. The theory is best used in connecting awareness with emotion, focusing on the individual and their psychological response, where their own judgment of a situation is to blame as the source of their emotional response.

Ortony *et al.*[59] describe an emotional classification that states that emotion is structured into the categories of *Fortune-of-others*, *Prospect-based*, *Well-Being*, *Attribution* and *Attraction*, or more largely grouped into *consequences of events*, *actions of agents*, or *aspects of objects*. This is known as the **OCC Model** and many virtual emotional models that opt for the appraisal theory often base themselves on this model. The OCC Model attempts to incorporate all emotions, but with no relationship between them other than categorical, but not all models based on the OCC model incorporate all emotions. One good example is the model proposed by Ochs *et al.*[13] which focuses on the believability of the NPCs. They attempt to improve the experience by focusing on the personality, social relations, and roles of the NPCs inside a game. The emotions modeled, using the OCC, were joy/distress, hope/fear, and relief/disappointment. There is also an emotional decay component implemented to revert the emotional state to a neutral state after some time period. This model focuses on the NPCs and tries to increase believability through simulation of social relations.

Another variation on the OCC model comes from He *et al.*[75], it is a fuzzy emotional model for virtual agents. The emotions of events modeled were hope/fear, satisfaction/fear-confirmation, and relief/disappointment, and were based on three variables, desirability, "if an event is beneficial to an agent, it is desirable otherwise it is undesirable", importance, "we equate goal's importance with motivation intensity", and likelihood, "we equate likelihood with the possibility of what other virtual agents will do". An important feature is "how the model accounts for relations between agents as predictor of behavior for them"[73].

Using a partial OCC model implementation, the work of Jacobs *et al.*[76] focuses on the importance of the link between reward/punishment and an emotional response. They propose the use of an emotion label system that converts each agent's state transition, an initial joy/distress mapping that converts into a hope/fear mapping over time, related to the agent's previous knowledge, hopefully allowing the gathering of useful information for planning and decision making of the agent. This model's objective is focused on the modeling of the agent behavior but is important to note how the modeling anticipatory behavior improved the tested agent's performance.

FatiMA[74] offers an interesting appraisal theory application, where the OCC model is implemented by storing appraisals (valence based) in a numeral intensity value (-10;10). This model acting along with goal mechanisms and perceived events, models the complete range of emotions inside the OCC, being able to give individual personalities and coping mechanisms to deal with specific goals. This model however not take into account anticipation of behavior or emotions.

The appraisal theory approaches show potential when used in a more static NPC or Environment emotional association, giving a simpler and robust emotional model to the in-game interactions.

A.3 Perception of Behavior

We live in a dynamic visual world: "Objects often move about in our environment and we, as observers, are often moving as we look at those events"[118], but the "[r]eflected light carries too much information for the human visual system to process at once"[174]. Instead, some locations or objects are selectively prioritized at the expense of others. Often, priority is task or goal dependent (e.g. drivers might preferentially attend to red objects because of the importance of brake lights, stop lights, and stop signs), but keep in mind that "the actual cues that capture visual attention are luminance-based transients [e.g. motion], not new objects"[174]. Furthermore, "[t]he visual system seems designed to give highest priority to moving objects, though at low resolution"[175].

At the turn of the century, vision research demonstrated that visual perception is divided into two major systems which operate in parallel and semi-independently. These two channels are referred to as the transient and the sustained neuron systems [143, 175]. The *sustained system* has poor temporal response but has a good spatial resolution. It is specialized in pattern and detail detection and processing of structural or figural information. The *transient system* has poor spatial resolution but has a good temporal response. It is specialized for motion detection, being part of an "early warning system"[143]. While the *sustained system* can help recognize shapes, it requires time to do so, this is counterposed with the transient system that only detects motion, but does so quickly. In other words, "[f]or any neuron, the better its capacity to resolve details in space, the poorer its capacity to resolve details in time"[143].

In practice, in humans "the critical duration for brightness discrimination [. . .] is on the order of 100ms, whereas for form identification or acuity the critical duration can range from 200-350ms"[143]. Parke also states that "[a]n abrupt change in the scene caused by motion will mask detail perception for about 250ms"[175]. These times give us a good understanding of how long it takes for someone to perceive the motion (100ms) and to actually understand what it is (200-350ms).

B

Appendix on the Expression of Emotions in Virtual Coaches

Contents

B.1 Bubble System in Detail

This section expands on the Bubble System presented previously. As previously stated, the Bubble System controls all 2D elements used to communicate emotion and is subdivided into two components: balloons and backgrounds. For the backgrounds, the relevant aspects are texture and color, while for the balloons, the sprite, text, effects, colors, and animations will change accordingly.

Balloon Shape – We created one distinct speech bubble for each emotion, plus a neutral state, based on the ones traditionally used in Comics (e.g. spiked balloons for anger) as well as using design guidelines to convey the appropriate valence for the associated emotion. To that end, we followed a user-centric iterative process until reaching a desired response to a speech balloon (e.g. curves for positive valences, and lines/obtuse angles for negative valences).

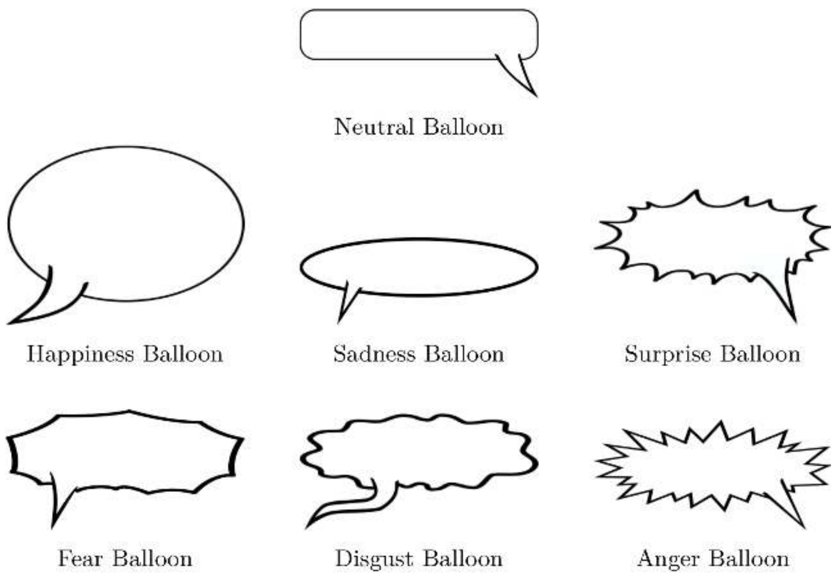


Figure B.1: Balloons shown with associated emotion. These balloons were colored and animated when appearing and disappearing.

We defined the neutral and anger balloon the standard comic way, as a rectangular and spiked shape respectively[1]. For happiness, we went with a more circular form, since curvy shapes tend to be perceived as positive, while sadness ended with an oval shape. Since surprise can either be positive or negative, we chose a mixture of balloons that had such valence. Therefore, it became similar to the anger balloon, but with fewer and more curved spikes. Disgust ended up having a wave format in order to convey the feeling of nausea. Finally fear, due to its negative valence, was created as multiple lines, creating obtuse angles.

Balloon Color – We decided to follow the same approach as Pires et al. [158], and draw inspiration from the Disney™ movie "Inside Out"[2]. The colors chosen follow the same representations as those

[1] Representation of scream (anger), narration (neutral), and dialog speech balloons as seen in Spider-Man Annual #5 - A Day at the Daily Bugle: `https://imgur.com/a/1J0l2tw`

[2] Inside Out, Walt Disney Pictures, Pixar Animation Studios (2015)

present in the movie, and for those emotions that were absent, a mix of other emotions/colors was used, to help communicate the appropriate valence.

Table B.1: Matching between colors and corresponding emotion as used in the application. The hexadecimal code used for the color is also shown.

Emotion	Color	Hex
Neutral		#8CDBA1FF
Happiness		#F0E64DFF
Surprise		#FFC358FF
Sadness		#1D33CEFF
Fear		#AE52ECFF
Disgust		#C5D137FF
Anger		#FF0000FF

Anger was represented with red, happiness with yellow, disgust with yellow-green, sadness with blue, and fear with purple. Surprise was defined as orange (aggressiveness of red and the positive valence of yellow), and for the neutral state, we decided upon a mix between green and blue, since it is usually seen as relaxed. For readability, the text color changes between black and white according to the balloon color.

Balloon Animation – We decided to create simple balloon animations as scaling/motion over time, following the pose to pose principle (see Section 2.8). In total, we ended up with 14 distinct animations, a pair (one to show the balloon, and another one to hide it) for each of the six emotions, plus another pair for the neutral state. To distinguish between different intensities, the speed of the animation was changed accordingly (slower speeds for lower intensities, faster speeds for higher ones).

Happiness was achieved with bounces, and sadness with a lower speed, due to how it is used in traditional animation; for surprise, we added a pendulum motion; for fear, we created a palpitation over time, since fear is usually accompanied by a strong heartbeat; disgust was achieved with a wave motion, with the idea of seasickness; and, finally, anger was defined with a chaotic path.

Text Animation – Regarding the text component of the speech balloons, we went with a parametric approach using animation curves (curves that define the text animation functions), since this allows us to have generalized effects and apply different curves to them, thus producing different animations (e.g. appear over time, jump, fade in/out, and shake).

The neutral animation was made with characters appearing over time; happiness ended up defined as a wave with a bell curve, which creates a jumping motion; surprise was defined as a swing, just like the pendulum motion of its balloon; sadness was created using a fade effect, and fear as jitters; disgust was created as a warp over a curve, which is just a displacement; and finally anger was defined with a shake.

Background Icons – Finally, we created icons for the backgrounds, and colored and animated the corresponding planes. For this, we either chose generic patterns, or specific icons, depending on whether we wanted the plane to be abstract, or to convey one of the specific application metrics discussed previously.

For abstract backgrounds, we wanted an image with any unrecognizable pattern, as long as we could define its color. For icons, we based them on the application metrics. The backgrounds were also given animations, mostly to change between them, and their respective color. However, while the colors used reflected the emotion felt, the iconography was not pursued extensively.

C

Appendix on 3motion

Contents

C.1 3motion Illustrative Scenario

This scenario is depicted in Figure C.1, where Hanna and Bob are about to cross a bridge. The boxes in white represent the action stage and emotion an agent is feeling at a given time, the arrows represent events being passed to the other agent.

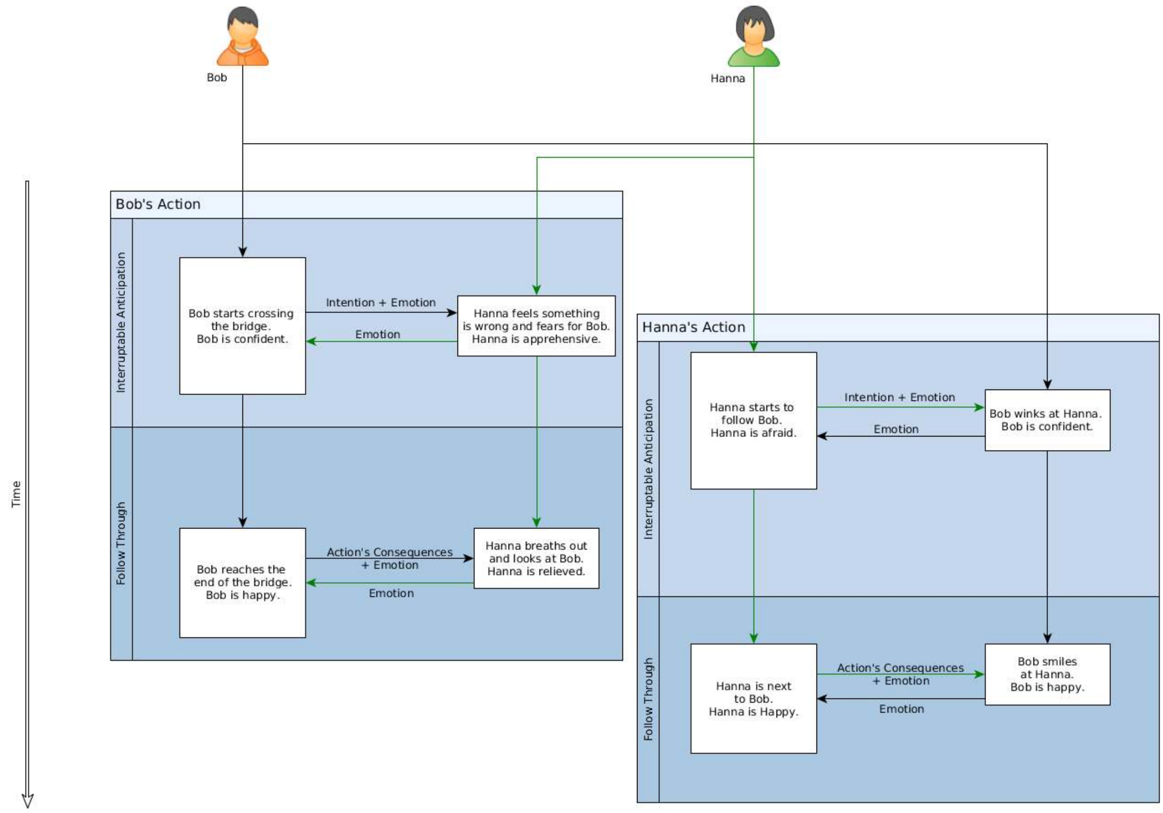


Figure C.1: Illustrative scenario where agents cross a bridge.

Bob confidently starts crossing the bridge, broadcasting his intention and emotion, some time later Hanna perceives the event and shows apprehension, but decides to follow Bob, showing fear in crossing the bridge. To this event Bob replies with confidence, maybe to try and calm Hanna down. Up to this point, the actions haven't left the interruptible anticipation stage and the agents could have canceled the action, but decided not to. It's important to note that in most systems, where the anticipation stage is not considered, these important non-verbal messages are completely ignored.

The first to reach the end of the bridge is Bob, showing happiness. Hanna is relieved that everything went fine and breaths out, and a little time later, she also reaches the end of the bridge next to Bob showing happiness, to which Bob replies with the same emotion.

Although with just two actions, this scene can be complex and help create more believable scenarios by using *3motion*.

D

Appendix on Designing Beliavable Interactions in *Adfectus*

Contents

D.1 On the First Evaluation on Believability

Here you will find the questionnaire used in this first evaluation on believability, seen in Section 5.2. The questionnaire is shown on the following pages.

Creating believable interactions in Adfectus

This questionnaire is designed to help find, out of three models, the one that creates the most believable interactions between characters in a game called Adfectus.

During this questionnaire, which should take about 20 minutes to complete, you will start by playing a tutorial version of the game followed by three different versions of the game, each with a different method of behavior. After playing a version of the game you will move to a new section of the questionnaire and answer a set of questions.

All the data collected through this questionnaire will only be used within the context of the project of which it is a part of. If you have any questions or concerns, please do so via the following email:
ANONYMIZED

Before agreeing to take part in this study, please read the following statements attentively:

1 - I have read and understood all information provided on this questionnaire. I had opportunity to ask questions, if necessary, and had them answered.

2 - I understand that participation in this study is voluntary and I may opt-out at any point, without needing to justify my decision. In which case, I will not be penalized and any data collected about me will be erased.

3 - All data will be stored privately and will be destroyed after five years in accordance with GDPR.

4 - I authorize the processing of the data collected within the context of this project for analysis, investigation and dissemination of results in scientific publications or conferences related to this project, by the author.

5 - I understand the data collected in this study will be used as explained above.

6 - In accordance with the details presented, I authorize my participation in this study and accept its conditions.

* Indica uma pergunta obrigatória

1. *

Marcar tudo o que for aplicável.

☐ I agree

Demographic questions

2. How old are you? *

3. What is your gender? *

Marcar apenas uma oval.

- Female
- Male
- Prefer not to answer
- Outra:

4. How often do you play video games? *

Marcar apenas uma oval.

- I make some time in my schedule to play video games.
- I play video games occasionally when the opportunity presents itself.
- I very rarely play video games.

5. Do you like action/fighting games, such as Chivalry and ForHonor?

Marcar apenas uma oval.

- I have played/watched others play them enough to understand I do not appreciate them.
- I enjoy them and have played/watched others play them multiple times.
- I am not familiar with these games and/or have not formed an opinion on them.

6. How often do you play action/fighting games, such as Chivalry or For honor? *

Marcar apenas uma oval.

- I make some time in my schedule to play action/fighting games.
- I play action/fighting games occasionally when the opportunity presents itself.
- I very rarely play action/fighting games.
- I never play action/fighting games.

7. Have you ever interacted with synthetic characters such as NPCs (Non-player characters)? *

Marcar apenas uma oval.

- I work with Synthetic Characters
- I interact regularly with Synthetic Characters
- I do not often interact with Synthetic Characters
- I have never interacted with Synthetic Characters
- I do not know what Synthetic Characters are

8. How do you value the expression of emotions by characters when playing games? *

Marcar apenas uma oval.

- I prefer when characters express emotions in games and that is crucial to my enjoyment
- I prefer when characters express emotions in games but that is not crucial to my enjoyment
- I'm indifferent to characters expressing emotions in games
- I prefer when characters don't express emotions in games but that is not crucial to my enjoyment
- I prefer when characters don't express emotions in games and that is crucial to my enjoyment

Tutorial

1. Open Adfectus and play the tutorial version of the game.
2. Once you feel comfortable with the controls and gameplay move to the next page of this questionnaire.

9. On a scale of 1 to 9 how much did you enjoy the game? *

Marcar apenas uma oval.

1	2	3	4	5	6	7	8	9
◯	◯	◯	◯	◯	◯	◯	◯	◯

10. After each play session you will be asked to rate the following set of statements on a scale of 1 (Strongly disagree) to 9 (Strongly agree) therefore it is very important that during the next play sessions, you **pay close attention to the characters' animations** (both yours and the opponent's), especially how and when they change. *

- I understood what the characters were doing.

- I understood what the characters were feeling.

- I understood the characters' intentions.

- The characters expressed emotions in a coherent manner.

- The characters expressed emotions in reaction to what happens (e.g. the character expresses fear when strongly hit by the opponent).

- The characters expressed emotions in anticipation of what could happen (e.g. the character becomes defensive when the opponent initiates an attack).

- The characters expressed emotions in reaction to what happened considering what was anticipated (e.g. the character expresses surprise when missing an attack the character was confident would connect).

- The characters' interactions were believable (i.e. coherent given the context in which they happen).

Marcar apenas uma oval.

( ) I understand

First version

1. Now, in game, go back to the main menu by pressing start or escape and selecting the main menu option.
2. Select the **first version of the game that has been attributed to you** and play until one player wins.

11. What version of the game did you play? *

Marcar apenas uma oval.

- Version A
- Version B
- Version C

12. I understood what the characters were doing. *

Marcar apenas uma oval.

	1	2	3	4	5	6	7	8	9	
Stro										Strongly agree

13. I understood what the characters were feeling. *

Marcar apenas uma oval.

	1	2	3	4	5	6	7	8	9	
Stro										Strongly agree

14. I understood the characters' intentions. *

Marcar apenas uma oval.

	1	2	3	4	5	6	7	8	9	
Stro										Strongly agree

15. The characters expressed emotions in a coherent manner. *

Marcar apenas uma oval.

	1	2	3	4	5	6	7	8	9	
Stro	◯	◯	◯	◯	◯	◯	◯	◯	◯	Strongly agree

16. The characters expressed emotions in reaction to what happens. *

Marcar apenas uma oval.

	1	2	3	4	5	6	7	8	9	
Stro	◯	◯	◯	◯	◯	◯	◯	◯	◯	Strongly agree

17. The characters expressed emotions in anticipation of what could happen. *

Marcar apenas uma oval.

	1	2	3	4	5	6	7	8	9	
Stro	◯	◯	◯	◯	◯	◯	◯	◯	◯	Strongly agree

18. The characters expressed emotions in reaction to what happened considering what was anticipated. *

Marcar apenas uma oval.

	1	2	3	4	5	6	7	8	9	
Stro	◯	◯	◯	◯	◯	◯	◯	◯	◯	Strongly agree

19. The characters' interactions were believable (i.e. coherent given the context in which they happen). *

Marcar apenas uma oval.

	1	2	3	4	5	6	7	8	9	
Stro	◯	◯	◯	◯	◯	◯	◯	◯	◯	Strongly agree

Second version

1. Now, in game, go back to the main menu by pressing start or escape and selecting the main menu option.
2. Select the **second version of the game that has been attributed to you** and play until one player wins.

20. What version of the game did you play? *

Marcar apenas uma oval.

- ◯ Version A
- ◯ Version B
- ◯ Version C

21. I understood what the characters were doing. *

Marcar apenas uma oval.

	1	2	3	4	5	6	7	8	9	
Stro	◯	◯	◯	◯	◯	◯	◯	◯	◯	Strongly agree

22. I understood what the characters were feeling. *

Marcar apenas uma oval.

	1	2	3	4	5	6	7	8	9	
Stro										Strongly agree

23. I understood the characters' intentions. *

Marcar apenas uma oval.

	1	2	3	4	5	6	7	8	9	
Stro										Strongly agree

24. The characters expressed emotions in a coherent manner. *

Marcar apenas uma oval.

	1	2	3	4	5	6	7	8	9	
Stro										Strongly agree

25. The characters expressed emotions in reaction to what happens. *

Marcar apenas uma oval.

	1	2	3	4	5	6	7	8	9	
Stro										Strongly agree

26. The characters expressed emotions in anticipation of what could happen. *

Marcar apenas uma oval.

	1	2	3	4	5	6	7	8	9	
Stro	◯	◯	◯	◯	◯	◯	◯	◯	◯	Strongly agree

27. The characters expressed emotions in reaction to what happened considering what was anticipated. *

Marcar apenas uma oval.

	1	2	3	4	5	6	7	8	9	
Stro	◯	◯	◯	◯	◯	◯	◯	◯	◯	Strongly agree

28. The characters' interactions were believable (i.e. coherent given the context in which they happen). *

Marcar apenas uma oval.

	1	2	3	4	5	6	7	8	9	
Stro	◯	◯	◯	◯	◯	◯	◯	◯	◯	Strongly agree

Third version

1. Now, in game, go back to the main menu by pressing start or escape and selecting the main menu option.
2. Select the **third version of the game that has been attributed to you** and play until one player wins.

29. What version of the game did you play? *

Marcar apenas uma oval.

- Version A
- Version B
- Version C

30. I understood what the characters were doing. *

Marcar apenas uma oval.

	1	2	3	4	5	6	7	8	9	
Stro										Strongly agree

31. I understood what the characters were feeling. *

Marcar apenas uma oval.

	1	2	3	4	5	6	7	8	9	
Stro										Strongly agree

32. I understood the characters’ intentions. *

Marcar apenas uma oval.

	1	2	3	4	5	6	7	8	9	
Stro										Strongly agree

33. The characters expressed emotions in a coherent manner. *

Marcar apenas uma oval.

	1	2	3	4	5	6	7	8	9	
Stro	◯	◯	◯	◯	◯	◯	◯	◯	◯	Strongly agree

34. The characters expressed emotions in reaction to what happens. *

Marcar apenas uma oval.

	1	2	3	4	5	6	7	8	9	
Stro	◯	◯	◯	◯	◯	◯	◯	◯	◯	Strongly agree

35. The characters expressed emotions in anticipation of what could happen. *

Marcar apenas uma oval.

	1	2	3	4	5	6	7	8	9	
Stro	◯	◯	◯	◯	◯	◯	◯	◯	◯	Strongly agree

36. The characters expressed emotions in reaction to what happened considering what was anticipated. *

Marcar apenas uma oval.

	1	2	3	4	5	6	7	8	9	
Stro	◯	◯	◯	◯	◯	◯	◯	◯	◯	Strongly agree

37. The characters' interactions were believable (i.e. coherent given the context in which they happen). *

Marcar apenas uma oval.

	1	2	3	4	5	6	7	8	9	
Stro	◯	◯	◯	◯	◯	◯	◯	◯	◯	Strongly agree

38. You have reached the end of the questionnaire. If you have any comments or suggestions regarding the study please write them below.

D.2 On the Application of the TAI Method in Adfectus

In this section, we present the different questionnaires used in this evaluation and then go into detail on certain choices made during the application of the TAI method on the Adfectus characters. Namely, we explore the selection of the best number of components for the PCA and the selection of the best number of clusters for the k-means Cluster Analysis.

D.2.1 Questionnaire for the Elicitation of Constructs

In the following pages, we present the questionnaire used in the first part of the TAI method, created to elicit constructs.

Completion Progress

Expressions on Synthetic Characters

Thank you for your participation in this study. This study is part of a Ph.D. dissertation at Instituto Superior Técnico that aims to **study the animation of expressions in synthetic characters**. Your participation and collaboration are, therefore, very much appreciated.

The experiment consists in watching very short videos (a few seconds each) depicting different expressions and comparing these animations. You are free to rewatch the videos as many times as you so desire. The experiment should take about **12 minutes**. All the answers will be anonymous and used solely for statistical purposes.

We also gently remind you that:
- participation is voluntary and you can withdraw at any time;
- you have the right to ask any question related to the experiment at any given time (e-mail: ricardo.proenca.rodrigues@tecnico.ulisboa.pt);
- you will not be identified at any stage of the study and individual results will not be shared;
- your participation does not involve physical or psychological risks;
- at some point you will be asked to complete one or more attention check tasks, failure in doing so will terminate the experiment in a fail state.

By proceeding to the questionnaire you are giving your consent.

Thank you for your time and consideration.

Continue

Completion Progress

Age:*

Gender:*

- ○ Female
- ○ Male
- ○ Prefer not to say
- ○ Other

Maternal Language:*

What is your experience with Synthetic Characters?

- ☐ Interacted with NPCs (non-player characters) in videogames
- ☐ Interacted with Virtual Assistants (e.g. Siri, Cortana, Google Assistant, in museums, etc.)
- ☐ Interacted with Social Robots (in scientific experiments, museums, etc.)
- ☐ Worked with or developed Synthetic Characters/Virtual Characters

Continue

The Experiment

In the following 6 sections you will be asked to compare 3 animations with different expressions by telling us the **two that are more alike**, **why** the two are more alike and how they **differ from the third**.

Here is an example:
Consider 3 different shapes, a **circle**, a **square** and a **triangle**. We could say for example that the **square** and the **triangle** are more **alike** because they are defined by **straight lines** and they are **different** from a **circle** because it has **curved lines**.

Please, do not reuse the same pair of characteristics more than once, even if it is the most obvious characteristic to distinguish between the animations.

The following section exemplifies how the questionnaire would be filled.

Continue

Example

Example of a possible answer.

The images below are representing three different shapes.

A B C

By comparing the 3 animations presented above, identify the two that are alike:

- [x] A
- [x] B
- [ ] C

Considering your previous answer, how are two of them alike and at the same time different from the third:

A and B are more alike because they are defined by straight lines and they are different from C because it has curved lines

Provide us with one characteristic that you found was alike in the two shapes:
(Use 1 to 3 words)

straight lines

Provide us with the opposite characteristic from the one mentioned above, describing the different shape:
(Use 1 to 3 words)

curved lines

Continue

Beginning of the experiment

The experiment itself starts in the next section.

Below you can find the example that was given before, in PDF format:

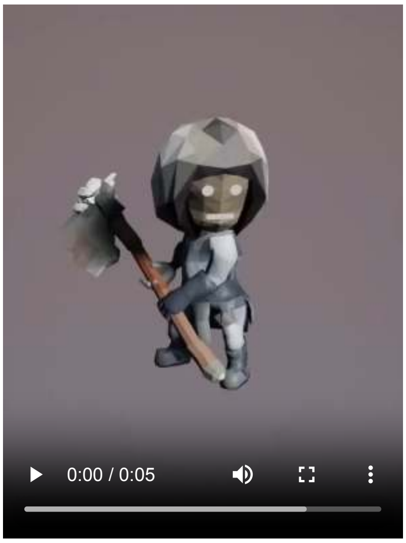
Example

Continue

Completion Progress

Comparison 1 of 4

The videos below are showing 3 animations of different emotion expressions.



A

B

C

In the following questions, please describe the observed behavior supporting your answers and not your interpretation of the behavior.
For instance, ***do not use*** *references to emotions (e.g., anger) but refer to the underlying behavior supporting your interpretation (e.g., grumbling).*

By comparing the 3 animations presented above, identify the two that are alike:*

☐ A ☐ B ☐ C

Considering your previous answer, how are two of them alike and at the same time different from the third:*

Provide us with one characteristic that you found was alike in the two animations:
(Use 1 to 3 words)*

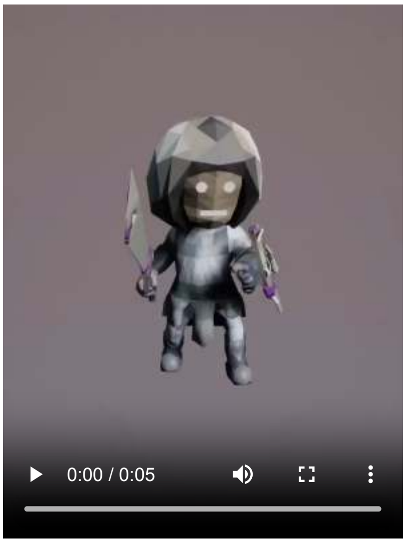
Provide us with the opposite characteristic from the one mentioned above, describing the different animation:
(Use 1 to 3 words)*

Continue

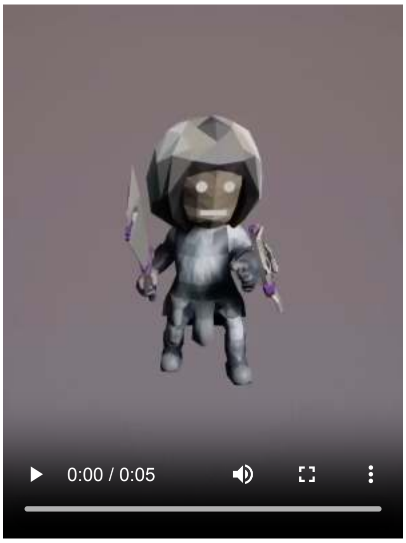
Completion Progress

Comparison 2 of 4

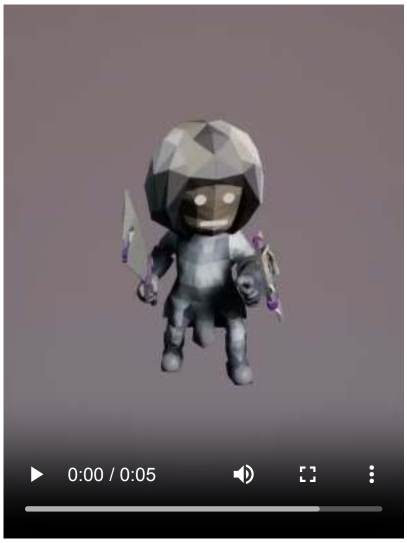
The videos below are showing 3 animations of different emotion expressions.



A



B



C

In the following questions, please describe the observed behavior supporting your answers and not your interpretation of the behavior.
*For instance, **do not use** references to emotions (e.g., anger) but refer to the underlying behavior supporting your interpretation (e.g., grumbling).*

By comparing the 3 animations presented above, identify the two that are alike:*

☐ A ☐ B ☐ C

Considering your previous answer, how are two of them alike and at the same time different from the third:*

*Please **do not repeat** characteristics from previous sections*

Provide us with one characteristic that you found was alike in the two animations:
(Use 1 to 3 words)*

Provide us with the opposite characteristic from the one mentioned above, describing the different animation:
(Use 1 to 3 words)*

Please select 'Agree' to show you are paying attention to this question:*

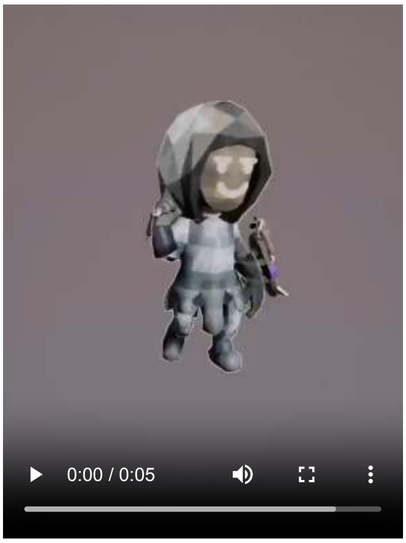
○ Strongly agree
○ Agree

- ○ Disagree
- ○ Strongly disagree

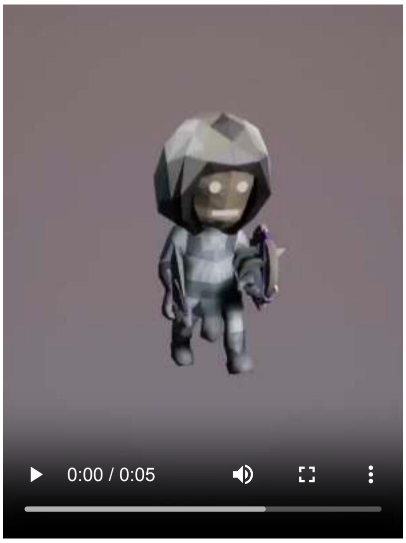
Continue

Comparison 3 of 4

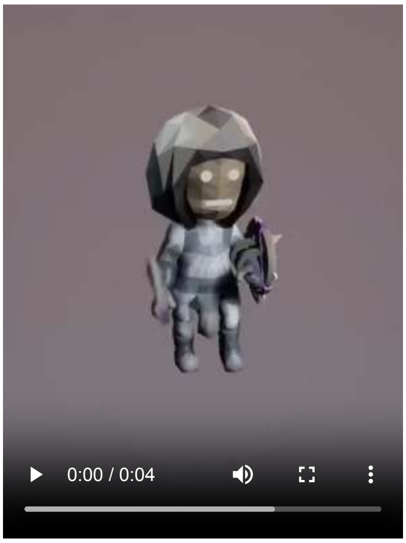
The videos below are showing 3 animations of different emotion expressions.



A



B



C

In the following questions, please describe the observed behavior supporting your answers and not your interpretation of the behavior.
*For instance, **do not use** references to emotions (e.g., anger) but refer to the underlying behavior supporting your interpretation (e.g., grumbling).*

By comparing the 3 animations presented above, identify the two that are alike:*

☐ A ☐ B ☐ C

Considering your previous answer, how are two of them alike and at the same time different from the third:*

*Please **do not repeat** characteristics from previous sections*

Provide us with one characteristic that you found was alike in the two animations:
(Use 1 to 3 words)*

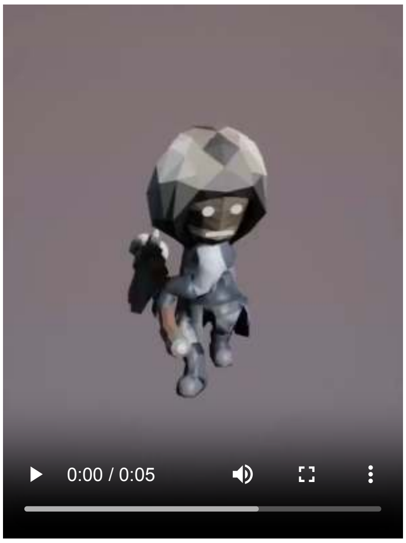
Provide us with the opposite characteristic from the one mentioned above, describing the different animation:
(Use 1 to 3 words)*

Continue

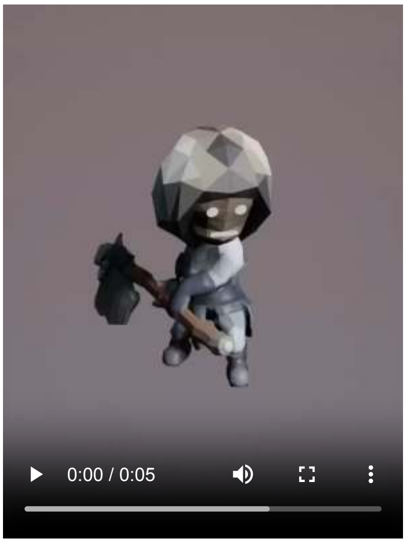
Completion Progress

Comparison 4 of 4

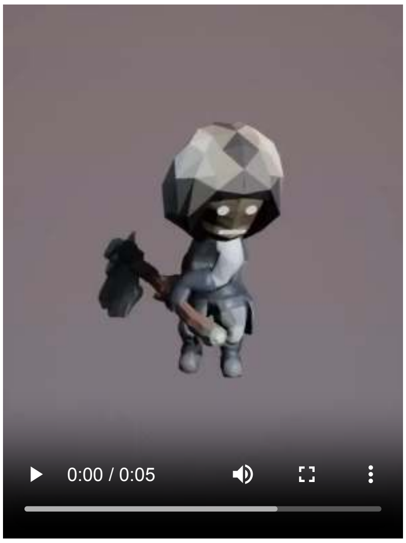
The videos below are showing 3 animations of different emotion expressions.



A



B



C

In the following questions, please describe the observed behavior supporting your answers and not your interpretation of the behavior.
*For instance, **do not use** references to emotions (e.g., anger) but refer to the underlying behavior supporting your interpretation (e.g., grumbling).*

By comparing the 3 animations presented above, identify the two that are alike:*

☐ A ☐ B ☐ C

Considering your previous answer, how are two of them alike and at the same time different from the third:*

*Please **do not repeat** characteristics from previous sections*

Provide us with one characteristic that you found was alike in the two animations:
(Use 1 to 3 words)*

Provide us with the opposite characteristic from the one mentioned above, describing the different animation:
(Use 1 to 3 words)*

Please select 'Disagree' to show you are paying attention to this question:*

○ Strongly agree

○ Agree

- ○ Disagree
- ○ Strongly disagree

Continue

Completion Progress

End of the experiment

Thank you for your participation in this study. **Press the submit button below to finish your participation.**

I would like to remind you that all data is used solely for statistical purposes.

For any further questions you can send an email to ricardo.proenca.rodrigues@tecnico.ulisboa.pt.

Submit

D.2.2 Questionnaire for the Rating of Elements

In the following pages, we present the questionnaire used in the second part of the TAI method, created to rate the elements, i.e. animations, against the constructs.

Expressions on Synthetic Characters

Thank you for your participation in this study. This study is part of a Ph.D. dissertation at Instituto Superior Técnico that aims to study the animation of expressions in synthetic characters. Your participation and collaboration are, therefore, very much appreciated.

You will be shown 16 short videos depicting different expressions and you will be asked to rate each animation against 12 different characteristics. The questionnaire should take about 20 minutes. All the answers will be anonymous and used solely for statistical purposes.

We also gently remind you that:

- participation is voluntary and you can withdraw at any time;
- you have the right to ask any question related to the experiment at any given time (e-mail: ricardo.proenca.rodrigues@tecnico.ulisboa.pt);
- you will not be identified at any stage of the study and individual results will not be shared;
- your participation does not involve physical or psychological risks.
- at some point, you will be asked to complete one or more attention-check tasks, failure in doing so will terminate the experiment in a fail state.

By proceeding to the questionnaire you are giving your consent.

Thank you for your time and consideration.

Age:

Gender:

Male

Female

Non-binary / third gender

Prefer not to say

Maternal Language

What is your experience with Synthetic Characters?

Interacted with NPCs (non-player characters) in videogames

Interacted with Virtual Assistants (e.g. Siri, Cortana, Google Assistant, in museums, etc.) ☐

Interacted with Social Robots (in scientific experiments, museums, etc.) ☐

Worked with or developed Synthetic Characters/Virtual Characters ☐

Rating of Animation 1 of 16

Rate the following animation:

Rate the animation using the following characteristics:

Grounded Posture	○	○	○	○	○	○	○	Shaky Posture
Natural Expression	○	○	○	○	○	○	○	Unnatural Expression
Weapon Held High	○	○	○	○	○	○	○	Weapon Held Low
Grimace Facial Expression	○	○	○	○	○	○	○	Cheerful Facial Expression
Ready to React	○	○	○	○	○	○	○	Relaxed
Prepared	○	○	○	○	○	○	○	Unprepared
Fast Walk Speed	○	○	○	○	○	○	○	Slow Walk Speed
Weapon Held in Front	○	○	○	○	○	○	○	Weapon Held Back
Leaning Back/Upright	○	○	○	○	○	○	○	Leaning Forward
Close Body Posture	○	○	○	○	○	○	○	Open Body Posture
Head Down/Low	○	○	○	○	○	○	○	Head Up/High
Offensive Stance	○	○	○	○	○	○	○	Defensive Stance

Rating of Animation 2 of 16

Rate the following animation:

Rate the animation using the following characteristics:

Grounded Posture	○	○	○	○	○	○	○	Shaky Posture
Natural Expression	○	○	○	○	○	○	○	Unnatural Expression
Weapon Held High	○	○	○	○	○	○	○	Weapon Held Low
Grimace Facial Expression	○	○	○	○	○	○	○	Cheerful Facial Expression
Ready to React	○	○	○	○	○	○	○	Relaxed
Prepared	○	○	○	○	○	○	○	Unprepared
Fast Walk Speed	○	○	○	○	○	○	○	Slow Walk Speed
Weapon Held in Front	○	○	○	○	○	○	○	Weapon Held Back
Leaning Back/Upright	○	○	○	○	○	○	○	Leaning Forward
Close Body Posture	○	○	○	○	○	○	○	Open Body Posture
Head Down/Low	○	○	○	○	○	○	○	Head Up/High
Offensive Stance	○	○	○	○	○	○	○	Defensive Stance

Rating of Animation 3 of 16

Rate the following animation:

Rate the animation using the following characteristics:

Grounded Posture	○	○	○	○	○	○	○	Shaky Posture
Natural Expression	○	○	○	○	○	○	○	Unnatural Expression
Weapon Held High	○	○	○	○	○	○	○	Weapon Held Low
Grimace Facial Expression	○	○	○	○	○	○	○	Cheerful Facial Expression
Ready to React	○	○	○	○	○	○	○	Relaxed
Prepared	○	○	○	○	○	○	○	Unprepared
Fast Walk Speed	○	○	○	○	○	○	○	Slow Walk Speed
Weapon Held in Front	○	○	○	○	○	○	○	Weapon Held Back
Leaning Back/Upright	○	○	○	○	○	○	○	Leaning Forward
Close Body Posture	○	○	○	○	○	○	○	Open Body Posture
Head Down/Low	○	○	○	○	○	○	○	Head Up/High
Offensive Stance	○	○	○	○	○	○	○	Defensive Stance

Rating of Animation 4 of 16

Rate the following animation:

Rate the animation using the following characteristics:

Grounded Posture	○	○	○	○	○	○	○	Shaky Posture
Natural Expression	○	○	○	○	○	○	○	Unnatural Expression
Weapon Held High	○	○	○	○	○	○	○	Weapon Held Low
Grimace Facial Expression	○	○	○	○	○	○	○	Cheerful Facial Expression
Ready to React	○	○	○	○	○	○	○	Relaxed
Prepared	○	○	○	○	○	○	○	Unprepared
Fast Walk Speed	○	○	○	○	○	○	○	Slow Walk Speed
Weapon Held in Front	○	○	○	○	○	○	○	Weapon Held Back
Leaning Back/Upright	○	○	○	○	○	○	○	Leaning Forward
Close Body Posture	○	○	○	○	○	○	○	Open Body Posture
Head Down/Low	○	○	○	○	○	○	○	Head Up/High
Offensive Stance	○	○	○	○	○	○	○	Defensive Stance

Please select 'Disagree' to show you are paying attention to this question:

- ○ Strongly disagree
- ○ Disagree

Somewhat disagree

Neither agree nor disagree

Somewhat agree

Agree

Strongly agree

Rating of Animation 5 of 16

Rate the following animation:

Rate the animation using the following characteristics:

Grounded Posture	○	○	○	○	○	○	○	Shaky Posture
Natural Expression	○	○	○	○	○	○	○	Unnatural Expression
Weapon Held High	○	○	○	○	○	○	○	Weapon Held Low
Grimace Facial Expression	○	○	○	○	○	○	○	Cheerful Facial Expression
Ready to React	○	○	○	○	○	○	○	Relaxed
Prepared	○	○	○	○	○	○	○	Unprepared
Fast Walk Speed	○	○	○	○	○	○	○	Slow Walk Speed
Weapon Held in Front	○	○	○	○	○	○	○	Weapon Held Back
Leaning Back/Upright	○	○	○	○	○	○	○	Leaning Forward
Close Body Posture	○	○	○	○	○	○	○	Open Body Posture
Head Down/Low	○	○	○	○	○	○	○	Head Up/High
Offensive Stance	○	○	○	○	○	○	○	Defensive Stance

Rating of Animation 6 of 16

Rate the following animation:

Rate the animation using the following characteristics:

Grounded Posture	◯	◯	◯	◯	◯	◯	◯	Shaky Posture
Natural Expression	◯	◯	◯	◯	◯	◯	◯	Unnatural Expression
Weapon Held High	◯	◯	◯	◯	◯	◯	◯	Weapon Held Low
Grimace Facial Expression	◯	◯	◯	◯	◯	◯	◯	Cheerful Facial Expression
Ready to React	◯	◯	◯	◯	◯	◯	◯	Relaxed
Prepared	◯	◯	◯	◯	◯	◯	◯	Unprepared
Fast Walk Speed	◯	◯	◯	◯	◯	◯	◯	Slow Walk Speed
Weapon Held in Front	◯	◯	◯	◯	◯	◯	◯	Weapon Held Back
Leaning Back/Upright	◯	◯	◯	◯	◯	◯	◯	Leaning Forward
Close Body Posture	◯	◯	◯	◯	◯	◯	◯	Open Body Posture
Head Down/Low	◯	◯	◯	◯	◯	◯	◯	Head Up/High
Offensive Stance	◯	◯	◯	◯	◯	◯	◯	Defensive Stance

Rating of Animation 7 of 16

Rate the following animation:

Rate the animation using the following characteristics:

Grounded Posture	○	○	○	○	○	○	○	Shaky Posture
Natural Expression	○	○	○	○	○	○	○	Unnatural Expression
Weapon Held High	○	○	○	○	○	○	○	Weapon Held Low
Grimace Facial Expression	○	○	○	○	○	○	○	Cheerful Facial Expression
Ready to React	○	○	○	○	○	○	○	Relaxed
Prepared	○	○	○	○	○	○	○	Unprepared
Fast Walk Speed	○	○	○	○	○	○	○	Slow Walk Speed
Weapon Held in Front	○	○	○	○	○	○	○	Weapon Held Back
Leaning Back/Upright	○	○	○	○	○	○	○	Leaning Forward
Close Body Posture	○	○	○	○	○	○	○	Open Body Posture
Head Down/Low	○	○	○	○	○	○	○	Head Up/High
Offensive Stance	○	○	○	○	○	○	○	Defensive Stance

Rating of Animation 8 of 16

Rate the following animation:

Rate the animation using the following characteristics:

Grounded Posture	○	○	○	○	○	○	○	Shaky Posture
Natural Expression	○	○	○	○	○	○	○	Unnatural Expression
Weapon Held High	○	○	○	○	○	○	○	Weapon Held Low
Grimace Facial Expression	○	○	○	○	○	○	○	Cheerful Facial Expression
Ready to React	○	○	○	○	○	○	○	Relaxed
Prepared	○	○	○	○	○	○	○	Unprepared
Fast Walk Speed	○	○	○	○	○	○	○	Slow Walk Speed
Weapon Held in Front	○	○	○	○	○	○	○	Weapon Held Back
Leaning Back/Upright	○	○	○	○	○	○	○	Leaning Forward
Close Body Posture	○	○	○	○	○	○	○	Open Body Posture
Head Down/Low	○	○	○	○	○	○	○	Head Up/High
Offensive Stance	○	○	○	○	○	○	○	Defensive Stance

Rating of Animation 9 of 16

Rate the following animation:

Rate the animation using the following characteristics:

Grounded Posture	○	○	○	○	○	○	○	Shaky Posture
Natural Expression	○	○	○	○	○	○	○	Unnatural Expression
Weapon Held High	○	○	○	○	○	○	○	Weapon Held Low
Grimace Facial Expression	○	○	○	○	○	○	○	Cheerful Facial Expression
Ready to React	○	○	○	○	○	○	○	Relaxed
Prepared	○	○	○	○	○	○	○	Unprepared
Fast Walk Speed	○	○	○	○	○	○	○	Slow Walk Speed
Weapon Held in Front	○	○	○	○	○	○	○	Weapon Held Back
Leaning Back/Upright	○	○	○	○	○	○	○	Leaning Forward
Close Body Posture	○	○	○	○	○	○	○	Open Body Posture
Head Down/Low	○	○	○	○	○	○	○	Head Up/High
Offensive Stance	○	○	○	○	○	○	○	Defensive Stance

Rating of Animation 10 of 16

Rate the following animation:

Rate the animation using the following characteristics:

Grounded Posture	○	○	○	○	○	○	○	Shaky Posture
Natural Expression	○	○	○	○	○	○	○	Unnatural Expression
Weapon Held High	○	○	○	○	○	○	○	Weapon Held Low
Grimace Facial Expression	○	○	○	○	○	○	○	Cheerful Facial Expression
Ready to React	○	○	○	○	○	○	○	Relaxed
Prepared	○	○	○	○	○	○	○	Unprepared
Fast Walk Speed	○	○	○	○	○	○	○	Slow Walk Speed
Weapon Held in Front	○	○	○	○	○	○	○	Weapon Held Back
Leaning Back/Upright	○	○	○	○	○	○	○	Leaning Forward
Close Body Posture	○	○	○	○	○	○	○	Open Body Posture
Head Down/Low	○	○	○	○	○	○	○	Head Up/High
Offensive Stance	○	○	○	○	○	○	○	Defensive Stance

Rating of Animation 11 of 16

Rate the following animation:

Rate the animation using the following characteristics:

Grounded Posture								Shaky Posture
Natural Expression								Unnatural Expression
Weapon Held High								Weapon Held Low
Grimace Facial Expression								Cheerful Facial Expression
Ready to React								Relaxed
Prepared								Unprepared
Fast Walk Speed								Slow Walk Speed
Weapon Held in Front								Weapon Held Back
Leaning Back/Upright								Leaning Forward
Close Body Posture								Open Body Posture
Head Down/Low								Head Up/High
Offensive Stance								Defensive Stance

Rating of Animation 12 of 16

Rate the following animation:

Rate the animation using the following characteristics:

Grounded Posture	○	○	○	○	○	○	○	Shaky Posture
Natural Expression	○	○	○	○	○	○	○	Unnatural Expression
Weapon Held High	○	○	○	○	○	○	○	Weapon Held Low
Grimace Facial Expression	○	○	○	○	○	○	○	Cheerful Facial Expression
Ready to React	○	○	○	○	○	○	○	Relaxed
Prepared	○	○	○	○	○	○	○	Unprepared
Fast Walk Speed	○	○	○	○	○	○	○	Slow Walk Speed
Weapon Held in Front	○	○	○	○	○	○	○	Weapon Held Back
Leaning Back/Upright	○	○	○	○	○	○	○	Leaning Forward
Close Body Posture	○	○	○	○	○	○	○	Open Body Posture
Head Down/Low	○	○	○	○	○	○	○	Head Up/High
Offensive Stance	○	○	○	○	○	○	○	Defensive Stance

Rating of Animation 13 of 16

Rate the following animation:

Rate the animation using the following characteristics:

Grounded Posture	○	○	○	○	○	○	○	Shaky Posture
Natural Expression	○	○	○	○	○	○	○	Unnatural Expression
Weapon Held High	○	○	○	○	○	○	○	Weapon Held Low
Grimace Facial Expression	○	○	○	○	○	○	○	Cheerful Facial Expression
Ready to React	○	○	○	○	○	○	○	Relaxed
Prepared	○	○	○	○	○	○	○	Unprepared
Fast Walk Speed	○	○	○	○	○	○	○	Slow Walk Speed
Weapon Held in Front	○	○	○	○	○	○	○	Weapon Held Back
Leaning Back/Upright	○	○	○	○	○	○	○	Leaning Forward
Close Body Posture	○	○	○	○	○	○	○	Open Body Posture
Head Down/Low	○	○	○	○	○	○	○	Head Up/High
Offensive Stance	○	○	○	○	○	○	○	Defensive Stance

Please select 'Agree' to show you are paying attention to this question:

○ Strongly disagree

○ Disagree

Somewhat disagree

Neither agree nor disagree

Somewhat agree

Agree

Strongly agree

Rating of Animation 14 of 16

Rate the following animation:

Rate the animation using the following characteristics:

Grounded Posture	○	○	○	○	○	○	○	Shaky Posture
Natural Expression	○	○	○	○	○	○	○	Unnatural Expression
Weapon Held High	○	○	○	○	○	○	○	Weapon Held Low
Grimace Facial Expression	○	○	○	○	○	○	○	Cheerful Facial Expression
Ready to React	○	○	○	○	○	○	○	Relaxed
Prepared	○	○	○	○	○	○	○	Unprepared
Fast Walk Speed	○	○	○	○	○	○	○	Slow Walk Speed
Weapon Held in Front	○	○	○	○	○	○	○	Weapon Held Back
Leaning Back/Upright	○	○	○	○	○	○	○	Leaning Forward
Close Body Posture	○	○	○	○	○	○	○	Open Body Posture
Head Down/Low	○	○	○	○	○	○	○	Head Up/High
Offensive Stance	○	○	○	○	○	○	○	Defensive Stance

Rating of Animation 15 of 16

Rate the following animation:

Rate the animation using the following characteristics:

Grounded Posture	◯	◯	◯	◯	◯	◯	◯	Shaky Posture
Natural Expression	◯	◯	◯	◯	◯	◯	◯	Unnatural Expression
Weapon Held High	◯	◯	◯	◯	◯	◯	◯	Weapon Held Low
Grimace Facial Expression	◯	◯	◯	◯	◯	◯	◯	Cheerful Facial Expression
Ready to React	◯	◯	◯	◯	◯	◯	◯	Relaxed
Prepared	◯	◯	◯	◯	◯	◯	◯	Unprepared
Fast Walk Speed	◯	◯	◯	◯	◯	◯	◯	Slow Walk Speed
Weapon Held in Front	◯	◯	◯	◯	◯	◯	◯	Weapon Held Back
Leaning Back/Upright	◯	◯	◯	◯	◯	◯	◯	Leaning Forward
Close Body Posture	◯	◯	◯	◯	◯	◯	◯	Open Body Posture
Head Down/Low	◯	◯	◯	◯	◯	◯	◯	Head Up/High
Offensive Stance	◯	◯	◯	◯	◯	◯	◯	Defensive Stance

Please select 'Somewhat agree' to show you are paying attention to this question:

- ◯ Strongly disagree
- ◯ Disagree

Somewhat disagree

Neither agree nor disagree

Somewhat agree

Agree

Strongly agree

Rating of Animation 16 of 16

Rate the following animation:

Rate the animation using the following characteristics:

Grounded Posture	○	○	○	○	○	○	○	Shaky Posture
Natural Expression	○	○	○	○	○	○	○	Unnatural Expression
Weapon Held High	○	○	○	○	○	○	○	Weapon Held Low
Grimace Facial Expression	○	○	○	○	○	○	○	Cheerful Facial Expression
Ready to React	○	○	○	○	○	○	○	Relaxed
Prepared	○	○	○	○	○	○	○	Unprepared
Fast Walk Speed	○	○	○	○	○	○	○	Slow Walk Speed
Weapon Held in Front	○	○	○	○	○	○	○	Weapon Held Back
Leaning Back/Upright	○	○	○	○	○	○	○	Leaning Forward
Close Body Posture	○	○	○	○	○	○	○	Open Body Posture
Head Down/Low	○	○	○	○	○	○	○	Head Up/High
Offensive Stance	○	○	○	○	○	○	○	Defensive Stance

Thank you for your participation in this study. Press the → button below to finish your participation.
I would like to remind you that all data is used solely for statistical purposes.
For any further questions, you can send an email to ricardo.proenca.rodrigues@tecnico.ulisboa.pt.

Feel free to leave any comment on the experiment:

D.2.3 On the selection of the best number of Components of the PCA

When computing a PCA, we need to choose a number of relevant components for the analysis and the following steps. Multiple indicators can be used, including using the inflection point in the scree plot, Kaiser's Criterion, Jolliffe's Criterion, and/or selecting a minimum cumulative variance the selected components have to overcome [171]. An initial analysis was run to obtain the variance and eigenvalues for each component of the data.

Table D.1: Components' variance and Eigenvalues on the performed PCA. The components marked in bold have an Eigenvalue greater than 1, while the components in italic have an Eigenvalue greater than 0.7.

Component	Eigenvalue	% of Variance	Cumulative %
1	**3.879**	**35.261**	**35.261**
2	**1.620**	**14.724**	**49.985**
3	**1.146**	**10.421**	**60.407**
4	*.958*	*8.708*	*69.115*
5	*.831*	*7.557*	*76.672*
6	.642	5.832	82.504
7	.574	5.221	87.724
8	.523	4.756	92.480
9	.375	3.405	95.885
10	.299	2.718	98.603
11	.154	1.397	100.000

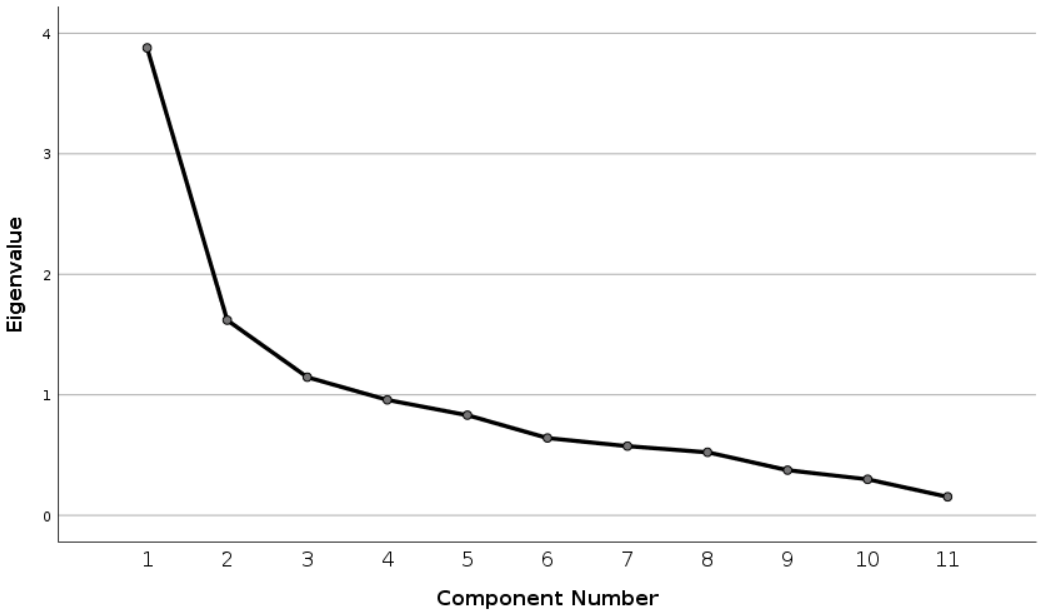


Figure D.1: Scree plot of the PCA. Note how the inflection point justifies retaining 3 components.

Looking at our results, we can gather the following results:

- Using the inflection point of the Scree plot, the recommended number of components is 3.
- Using Kaiser's Criterion (Eigenvalues > 1), the recommended number of components is 3.
- Using Jolliffe's Criterion (Eigenvalues > 0.7), the recommended number of components is 5.

– Looking at the cumulative variance, 3 components only represent 60% of the results, while 5 components reach 76% of the results.

With the resulting Rotated Component Matrix with three components (see Table 5.4), we find no cross-loading. With these metrics, we continued the analysis with 3 components.

D.2.4 On the selection of the best number of clusters for the k-means Cluster Analysis

To determine the appropriate number of clusters to use in the analysis of the model created from PCA, we perform multiple K-Mean analyses with varying numbers of clusters (from 2 to 10 clusters). Then to inquire about the number of relevant clusters, we use several methods, namely, the Silhouette and Elbow Methods[145], and the Calinski-Harabasz Index[176]. As an additional indicator, we also computed a two-step cluster solution provided by SPSS[1], which automatically selects a number of clusters.

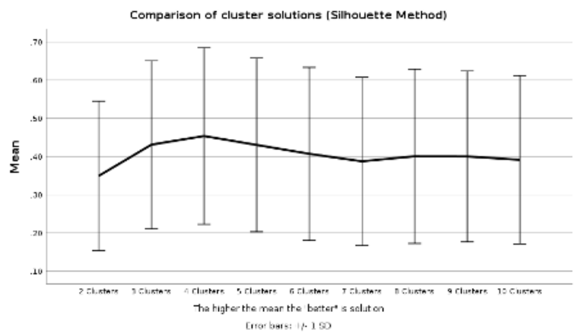


Figure D.2: Comparison of differently numbered cluster solutions through the Silhouette Method. The higher the mean, the "better" the solution. In this case, the best is the 4 Clusters solution.

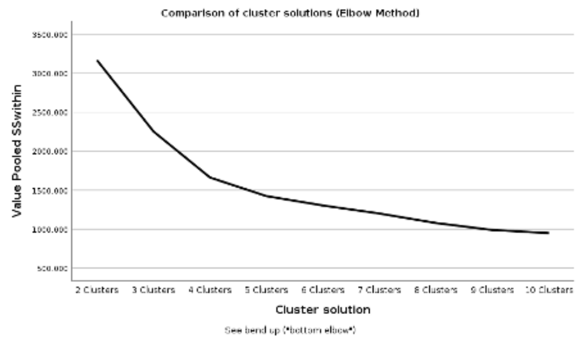


Figure D.3: Comparison of differently numbered cluster solutions through the Elbow Method. The number prior to an accentuated bend is the "best" solution. In this case, the best is the 4 Clusters solution.

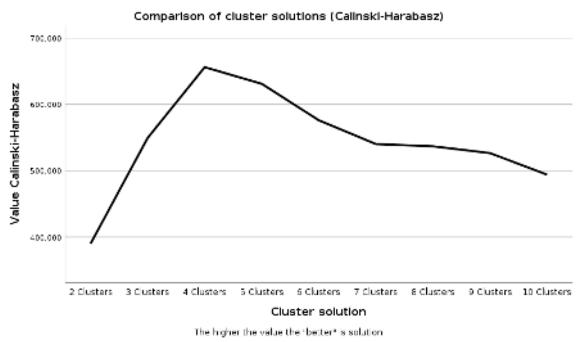


Figure D.4: Comparison of differently numbered cluster solutions through the Calinski-Harabasz Method. The higher the value, the "better" the solution. In this case, the best is the 4 Clusters solution.

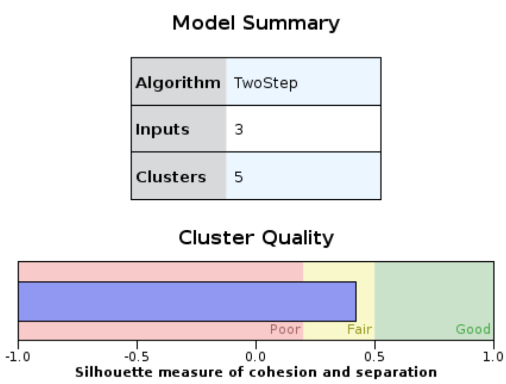


Figure D.5: Comparison of differently numbered cluster solutions through the Elbow Method. The number prior to an accentuated bend is the "best" solution. In this case, the best is the 4 Clusters solution.

After running the previously mentioned methods, here are our findings:

[1] IBM Corp. Released 2019. IBM SPSS Statistics for Linux, Version 26.0. Armonk, NY: IBM Corp

- The Silhouette method points towards the use of 4 Clusters (see Figure D.2).
- The Elbow method points towards the use of 4 Clusters (see Figure D.3).
- The Calinski-Harabasz method points towards the use of 4 Clusters.
- The Two-Step Clustering method selected 5 Clusters.

We decided to use a total of 4 Clusters for our Cluster Analysis, given that 3 out of 4 methods pointed to that value.

D.3 On the Second Evaluation on Believability

Here you will find the questionnaire used in this second evaluation along with a more detailed exploration of the data analysis performed in the second evaluation on Believability in Section 5.4.4.

D.3.1 Second Questionnaire on Interaction Believability

In the following pages, we present the questionnaire used in the second evaluation on believability in *Adfectus*.

Creating believable interactions in Adfectus

This questionnaire is part of Ricardo Rodrigues' Ph.D. Thesis at Instituto Superior Técnico, University of Lisbon, to help find, out of three models, the one that creates the most believable interactions between characters in a game called Adfectus. This thesis is supervised by professor Carlos Martinho.

During this questionnaire, which should take about 25 minutes to complete, you will start by playing a tutorial version of the game followed by three different versions of the game, each with a different method of behavior. After playing a version of the game you will move to a new section of the questionnaire and answer a set of questions.

All the data collected through this questionnaire will only be used within the context of the thesis of which it is a part of. If you have any questions or concerns, please do so via the following email:
ricardo.proenca.rodrigues@tecnico.ulisboa.pt

Please read the following statements attentively:

1 - I have read and understood all information provided on this questionnaire. I had opportunity to ask questions, if necessary, and had them answered.

2 - I understand that participation in this study is voluntary and I may opt-out at any point, without needing to justify my decision. In which case, I will not be penalized and any data collected about me will be erased.

3 - All data will be stored privately and will be destroyed after five years in accordance with GDPR.

4 - I authorize the processing of the data collected within the context of this project for analysis, investigation and dissemination of results in scientific publications or conferences related to this project, by the author.

5 - I understand the data collected in this study will be used as explained above.

6 - In accordance with the details presented, I authorize my participation in this study and accept its conditions.

By proceeding to the questionnaire you are agreeing with the statements above and giving your consent to your participation in this study.

Thank you for your time and consideration.

* Indica uma pergunta obrigatória

Demographic questions

1. How old are you? *

2. What is your gender? *

Marcar apenas uma oval.

- Female
- Male
- Prefer not to answer
- Outra:

3. How often do you play video games? *

Marcar apenas uma oval.

- I make some time in my schedule to play video games.
- I play video games occasionally when the opportunity presents itself.
- I very rarely play video games.

4. Do you like action/fighting games, such as Chivalry and ForHonor? *

Marcar apenas uma oval.

- I have played/watched others play them enough to understand I do not appreciate them.
- I enjoy them and have played/watched others play them multiple times.
- I am not familiar with these games and/or have not formed an opinion on them.

5. How often do you play action/fighting games, such as Chivalry or For honor? *

Marcar apenas uma oval.

- I make some time in my schedule to play action/fighting games.
- I play action/fighting games occasionally when the opportunity presents itself.
- I very rarely play action/fighting games.
- I never play action/fighting games.

6. Have you ever interacted with synthetic characters such as NPCs (Non-player characters)? *

Marcar apenas uma oval.

- I work with Synthetic Characters
- I interact regularly with Synthetic Characters
- I do not often interact with Synthetic Characters
- I have never interacted with Synthetic Characters
- I do not know what Synthetic Characters are

7. How do you value the expression of emotions by characters when playing games? *

Marcar apenas uma oval.

- I prefer when characters express emotions in games and that is crucial to my enjoyment
- I prefer when characters express emotions in games but that is not crucial to my enjoyment
- I'm indifferent to characters expressing emotions in games
- I prefer when characters don't express emotions in games but that is not crucial to my enjoyment
- I prefer when characters don't express emotions in games and that is crucial to my enjoyment

Tutorial

1. Open Adfectus and play the tutorial version of the game.
2. Once you feel comfortable with the controls and gameplay move to the next page of this questionnaire.

8. I enjoyed playing the tutorial. *

Select the column below that best represents how you feel about each statement.

Marcar apenas uma oval.

	1	2	3	4	5	6	7	
Stro	◯	◯	◯	◯	◯	◯	◯	Strongly agree

9. I could easily control the character. *

Select the column below that best represents how you feel about each statement.

Marcar apenas uma oval.

	1	2	3	4	5	6	7	
Stro	◯	◯	◯	◯	◯	◯	◯	Strongly agree

10. After each play session you will be asked to rate the following set of statements on a scale of 1 (Strongly disagree) to 7 (Strongly agree) therefore it is very important that during the next play sessions, you **pay close attention to the characters' animations and expressions** (both yours and the opponent's), especially how and when they change. *

- I understood what the characters were doing.

- I could predict the characters' actions.

- I understood what the characters were feeling.

- I could predict the characters' feelings.

- I understood the characters' intentions.

- The characters were aware of each other.

- The characters were aware of each other's actions.

- The characters could predict each others' actions.

- The characters were aware of each other's feelings.

- The characters could predict each others' feelings.

- The characters were aware of each other's intentions.

- The characters displayed different personalities.

Marcar apenas uma oval.

◯ I understand

First version

1. Now, in game, go back to the main menu by pressing start or escape and selecting the main menu option.
2. Select the **first version of the game that has been attributed to you** and play until one player wins.

11. What version of the game did you play? *

Marcar apenas uma oval.

- Version A
- Version B
- Version C

12. Regarding your perception of the **characters' expressions during the combat**:

Select the items below that best represents your experience.

Marcar tudo o que for aplicável.

- [ ] I was aware of the expressions of the character I was controlling.
- [ ] I was aware of the expressions of the character my opponent was controlling.

13. I understood what the characters were doing. *

Select the column below that best represents how you feel about each statement.

Marcar apenas uma oval.

	1	2	3	4	5	6	7	
Stro								Strongly agree

14. I could predict the characters' actions. *

Select the column below that best represents how you feel about each statement.

Marcar apenas uma oval.

	1	2	3	4	5	6	7	
Stro								Strongly agree

15. I understood what the characters were feeling. *

Select the column below that best represents how you feel about each statement.

Marcar apenas uma oval.

	1	2	3	4	5	6	7	
Stro	◯	◯	◯	◯	◯	◯	◯	Strongly agree

16. I could predict the characters' feelings. *

Select the column below that best represents how you feel about each statement.

Marcar apenas uma oval.

	1	2	3	4	5	6	7	
Stro	◯	◯	◯	◯	◯	◯	◯	Strongly agree

17. I understood the characters' intentions. *

Select the column below that best represents how you feel about each statement.

Marcar apenas uma oval.

	1	2	3	4	5	6	7	
Stro	◯	◯	◯	◯	◯	◯	◯	Strongly agree

18. The characters were aware of each other. *

Select the column below that best represents how you feel about each statement.

Marcar apenas uma oval.

	1	2	3	4	5	6	7	
Stro	◯	◯	◯	◯	◯	◯	◯	Strongly agree

19. The characters were aware of each other's actions. *

Select the column below that best represents how you feel about each statement.

Marcar apenas uma oval.

	1	2	3	4	5	6	7	
Stro	◯	◯	◯	◯	◯	◯	◯	Strongly agree

20. The characters could predict each others' actions. *

Select the column below that best represents how you feel about each statement.

Marcar apenas uma oval.

	1	2	3	4	5	6	7	
Stro	◯	◯	◯	◯	◯	◯	◯	Strongly agree

21. The characters were aware of each other's feelings. *

Select the column below that best represents how you feel about each statement.

Marcar apenas uma oval.

	1	2	3	4	5	6	7	
Stro	◯	◯	◯	◯	◯	◯	◯	Strongly agree

22. The characters could predict each others' feelings. *

Select the column below that best represents how you feel about each statement.

Marcar apenas uma oval.

	1	2	3	4	5	6	7	
Stro	◯	◯	◯	◯	◯	◯	◯	Strongly agree

23. The characters were aware of each other's intentions. *

Select the column below that best represents how you feel about each statement.

Marcar apenas uma oval.

	1	2	3	4	5	6	7	
Stro	◯	◯	◯	◯	◯	◯	◯	Strongly agree

24. The interaction between characters in this scene was believable. *

Select the column below that best represents how you feel about each statement.

Marcar apenas uma oval.

	1	2	3	4	5	6	7	
Stro	◯	◯	◯	◯	◯	◯	◯	Strongly agree

25. The characters displayed different personalities. *

Select the column below that best represents how you feel about each statement.

Marcar apenas uma oval.

	1	2	3	4	5	6	7	
Stro	◯	◯	◯	◯	◯	◯	◯	Strongly agree

26. Please provide one or two adjectives to describe the personality of the character holding an axe.

Skip this question if you did not found differences in the characters' personalities.

____

27. Please provide one or two adjectives to describe the personality of the character holding a sword and shield.

Skip this question if you did not found differences in the characters' personalities.

____

Second version

1. Now, in game, go back to the main menu by pressing start or escape and selecting the main menu option.
2. Select the **second version of the game that has been attributed to you** and play until one player wins.

28. What version of the game did you play? *

Marcar apenas uma oval.

- Version A
- Version B
- Version C

29. Regarding your perception of the **characters' expressions during the combat**:

Select the items below that best represents your experience.

Marcar tudo o que for aplicável.

- [ ] I was aware of the expressions of the character I was controlling.
- [ ] I was aware of the expressions of the character my opponent was controlling.

30. I understood what the characters were doing. *

Select the column below that best represents how you feel about each statement.

Marcar apenas uma oval.

	1	2	3	4	5	6	7	
Stro								Strongly agree

31. I could predict the characters' actions. *

Select the column below that best represents how you feel about each statement.

Marcar apenas uma oval.

	1	2	3	4	5	6	7	
Stro	◯	◯	◯	◯	◯	◯	◯	Strongly agree

32. I understood what the characters were feeling. *

Select the column below that best represents how you feel about each statement.

Marcar apenas uma oval.

	1	2	3	4	5	6	7	
Stro	◯	◯	◯	◯	◯	◯	◯	Strongly agree

33. I could predict the characters' feelings. *

Select the column below that best represents how you feel about each statement.

Marcar apenas uma oval.

	1	2	3	4	5	6	7	
Stro	◯	◯	◯	◯	◯	◯	◯	Strongly agree

34. I understood the characters' intentions. *

Select the column below that best represents how you feel about each statement.

Marcar apenas uma oval.

	1	2	3	4	5	6	7	
Stro	◯	◯	◯	◯	◯	◯	◯	Strongly agree

35. The characters were aware of each other. *

Select the column below that best represents how you feel about each statement.

Marcar apenas uma oval.

	1	2	3	4	5	6	7	
Stro								Strongly agree

36. The characters were aware of each other's actions. *

Select the column below that best represents how you feel about each statement.

Marcar apenas uma oval.

	1	2	3	4	5	6	7	
Stro								Strongly agree

37. The characters could predict each others' actions. *

Select the column below that best represents how you feel about each statement.

Marcar apenas uma oval.

	1	2	3	4	5	6	7	
Stro								Strongly agree

38. The characters were aware of each other's feelings. *

Select the column below that best represents how you feel about each statement.

Marcar apenas uma oval.

	1	2	3	4	5	6	7	
Stro								Strongly agree

39. The characters could predict each others' feelings. *

Select the column below that best represents how you feel about each statement.

Marcar apenas uma oval.

	1	2	3	4	5	6	7	
Stro								Strongly agree

40. The characters were aware of each other's intentions. *

Select the column below that best represents how you feel about each statement.

Marcar apenas uma oval.

	1	2	3	4	5	6	7	
Stro								Strongly agree

41. The interaction between characters in this scene was believable. *

Select the column below that best represents how you feel about each statement.

Marcar apenas uma oval.

	1	2	3	4	5	6	7	
Stro								Strongly agree

42. The characters displayed different personalities. *

Select the column below that best represents how you feel about each statement.

Marcar apenas uma oval.

	1	2	3	4	5	6	7	
Stro								Strongly agree

43. Please provide one or two adjectives to describe the personality of the character holding an axe.

Skip this question if you did not found differences in the characters' personalities.

44. Please provide one or two adjectives to describe the personality of the character holding a sword and shield.

Skip this question if you did not found differences in the characters' personalities.

Third version

1. Now, in game, go back to the main menu by pressing start or escape and selecting the main menu option.
2. Select the **third version of the game that has been attributed to you** and play until one player wins.

45. What version of the game did you play? *

Marcar apenas uma oval.

- Version A
- Version B
- Version C

46. Regarding your perception of the **characters' expressions during the combat**:

Select the items below that best represents your experience.

Marcar tudo o que for aplicável.

- [ ] I was aware of the expressions of the character I was controlling.
- [ ] I was aware of the expressions of the character my opponent was controlling.

47. I understood what the characters were doing. *

Select the column below that best represents how you feel about each statement.

Marcar apenas uma oval.

	1	2	3	4	5	6	7	
Stro	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Strongly agree

48. I could predict the characters' actions. *

Select the column below that best represents how you feel about each statement.

Marcar apenas uma oval.

	1	2	3	4	5	6	7	
Stro	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Strongly agree

49. I understood what the characters were feeling. *

Select the column below that best represents how you feel about each statement.

Marcar apenas uma oval.

	1	2	3	4	5	6	7	
Stro	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Strongly agree

50. I could predict the characters' feelings. *

Select the column below that best represents how you feel about each statement.

Marcar apenas uma oval.

	1	2	3	4	5	6	7	
Stro	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Strongly agree

51. I understood the characters' intentions. *

Select the column below that best represents how you feel about each statement.

Marcar apenas uma oval.

	1	2	3	4	5	6	7	
Stro	◯	◯	◯	◯	◯	◯	◯	Strongly agree

52. The characters were aware of each other. *

Select the column below that best represents how you feel about each statement.

Marcar apenas uma oval.

	1	2	3	4	5	6	7	
Stro	◯	◯	◯	◯	◯	◯	◯	Strongly agree

53. The characters were aware of each other's actions. *

Select the column below that best represents how you feel about each statement.

Marcar apenas uma oval.

	1	2	3	4	5	6	7	
Stro	◯	◯	◯	◯	◯	◯	◯	Strongly agree

54. The characters could predict each others' actions. *

Select the column below that best represents how you feel about each statement.

Marcar apenas uma oval.

	1	2	3	4	5	6	7	
Stro	◯	◯	◯	◯	◯	◯	◯	Strongly agree

55. The characters were aware of each other's feelings. *

Select the column below that best represents how you feel about each statement.

Marcar apenas uma oval.

	1	2	3	4	5	6	7	
Stro	◯	◯	◯	◯	◯	◯	◯	Strongly agree

56. The characters could predict each others' feelings. *

Select the column below that best represents how you feel about each statement.

Marcar apenas uma oval.

	1	2	3	4	5	6	7	
Stro	◯	◯	◯	◯	◯	◯	◯	Strongly agree

57. The characters were aware of each other's intentions. *

Select the column below that best represents how you feel about each statement.

Marcar apenas uma oval.

	1	2	3	4	5	6	7	
Stro	◯	◯	◯	◯	◯	◯	◯	Strongly agree

58. The interaction between characters in this scene was believable. *

Select the column below that best represents how you feel about each statement.

Marcar apenas uma oval.

	1	2	3	4	5	6	7	
Stro	◯	◯	◯	◯	◯	◯	◯	Strongly agree

59. The characters displayed different personalities. *

Select the column below that best represents how you feel about each statement.

Marcar apenas uma oval.

	1	2	3	4	5	6	7	
Stro								Strongly agree

60. Please provide one or two adjectives to describe the personality of the character holding an axe.

Skip this question if you did not found differences in the characters' personalities.

61. Please provide one or two adjectives to describe the personality of the character holding a sword and shield.

Skip this question if you did not found differences in the characters' personalities.

62. You have reached the end of the questionnaire. If you have any comments or suggestions regarding the study please write them below.

D.3.2 Data Analysis in Depth

Our group of participants ($N = 32$), collected by convenience sampling, was between 20 and 33 years old ($M = 24$, $SD = 3.542$). 8 were women, 22 were men, and 2 preferred not to identify their gender. Regarding the participants' habits with video games, 50.0% reported making time to play video games, while 37.0% played when the opportunity presented itself, and the rest rarely played. Regarding action/fighting games, 50.0% of the participants enjoy playing or watching this genre of games, 37.5% was not familiar with this genre, and the remaining 12.5% did not enjoy this genre of games. Furthermore, 65.6% of the participants rarely or never played these types of games, and only 2 participants report actively scheduling time to play these types of games. Regarding the experience with synthetic characters, 75% interact regularly with synthetic characters (4 of which work with these types of characters). Furthermore, regarding the expression of emotions by characters, most participants (90.6%) prefer characters that express emotions, but only 21.9% consider it crucial for their enjoyment.

A Shapiro-Wilk test determined that only a few of the data sets of the statements were normally distributed, thus we decided to use non-parametric tests. To understand whether there was a significant difference between versions, for each statement we performed a Friedman test; furthermore, where appropriate, a post hoc analysis with Wilcoxon signed-rank tests was performed with a Bonferroni correction applied, resulting in a significance level set at $p < 0.017$. We made available a document that records every test performed on the data[2]. Figure D.6 shows the averages and standard deviations of each statement for each version.

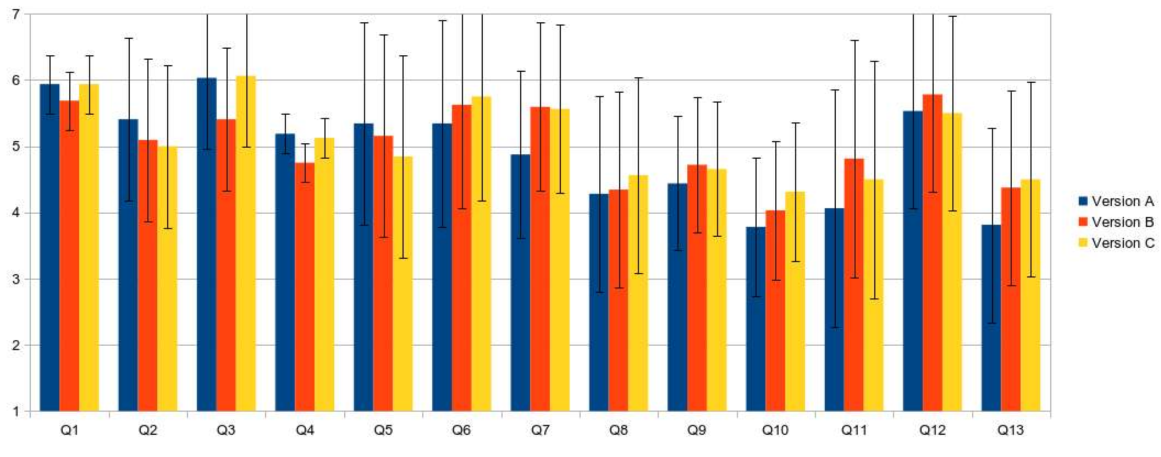


Figure D.6: Averages and associated standard deviations of each statement for each version.

Q1 - I understood what the characters were doing. For this statement, version A had a mean of 5.94 ($SD = 1.268$); version B had a mean of 5.69 ($SD = 1.230$); and version C had a mean of 5.94 ($SD = 1.076$). Figure D.7 shows the distribution of the Likert scale values by version for this question. We found no statistically significant differences between the versions ($\chi^2(2) = 2.844$, $p = 0.241$).

[2] A document detailing every test performed on the data can be found at `https://drive.google.com/file/d/1zUArcBC93n2eGZ6GsIZGo3O30csJnWeR/view?usp=sharing`

Overall, the participants seemed to understand what the characters were doing regardless of the version.

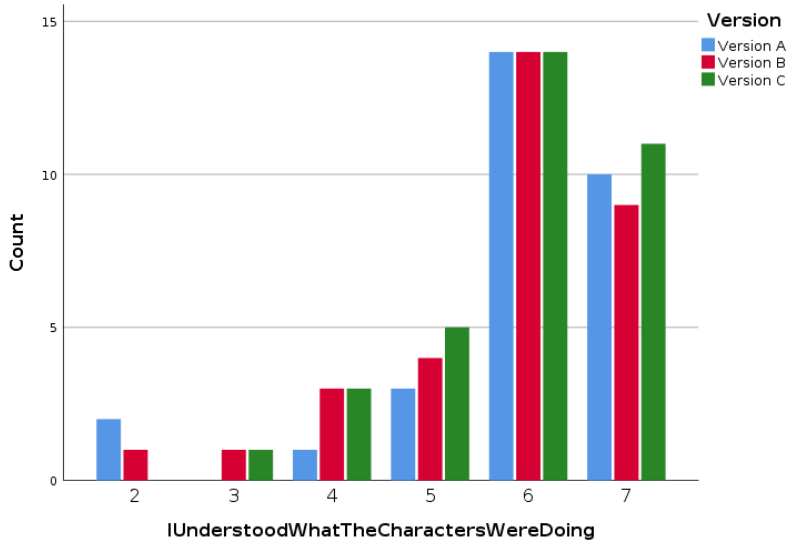


Figure D.7: Bar graph displaying the number of occurrences of each value of the Likert scale for the *Q1 - I understood what the characters were doing* - by played version.

Q2 - I could predict the characters' actions. For this statement, version A had a mean of 5.41 ($SD = 1.043$); version B had a mean of 5.09 ($SD = 1.532$); and version C had a mean of 5.00 ($SD = 1.566$). Figure D.8 shows the distribution of the Likert scale values by version for this question. We found no statistically significant differences between the versions ($\chi^2(2) = 0.889$, $p = 0.641$). Overall, the participants seemed moderately able to predict the characters' actions regardless of version.

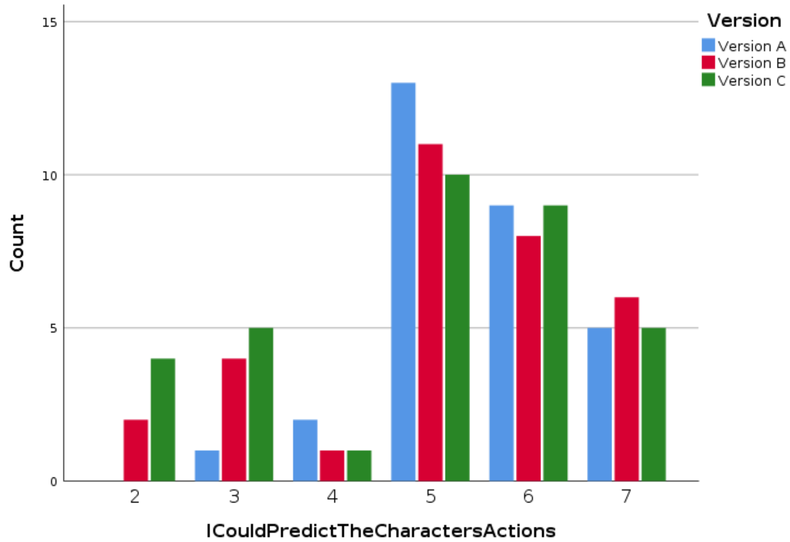


Figure D.8: Bar graph displaying the number of occurrences of each value of the Likert scale for the *Q2 - I could predict the characters' actions* - by played version.

Q3 - I understood what the characters were feeling. For this statement, version A had a mean of 6.03 ($SD = 1.470$); version B had a mean of 5.41 ($SD = 1.478$); and version C had a mean of 6.06 ($SD = 1.014$). Figure D.9 shows the distribution of the Likert scale values by version for this question. We found statistically significant differences between the versions ($\chi^2(2) = 6.179$, $p = 0.046$). Thus, we performed the post hoc analysis. No significant differences were found between any pair: between versions A and B, $Z = -2.156, p = 0.031$; between versions C and B, $Z = -2.244, p = 0.025$; and between versions A and C, $Z = -0.053, p = 0.958$.

So, while a statistical difference was observed in the Friedman test, we found no statistically significant differences between pairs. We can still see that Version C was higher in mean value, followed by Version A, and Version B was the lowest in rank. Versions A and C are also very similar in values, as seen by the low difference in the Wilcoxon test ($Z = -0.053, p = 0.958$). If we wish to take meaning from these values, then in all versions the characters' feelings are understood, and in versions A and C the characters' feelings are slightly better understood than in version B.

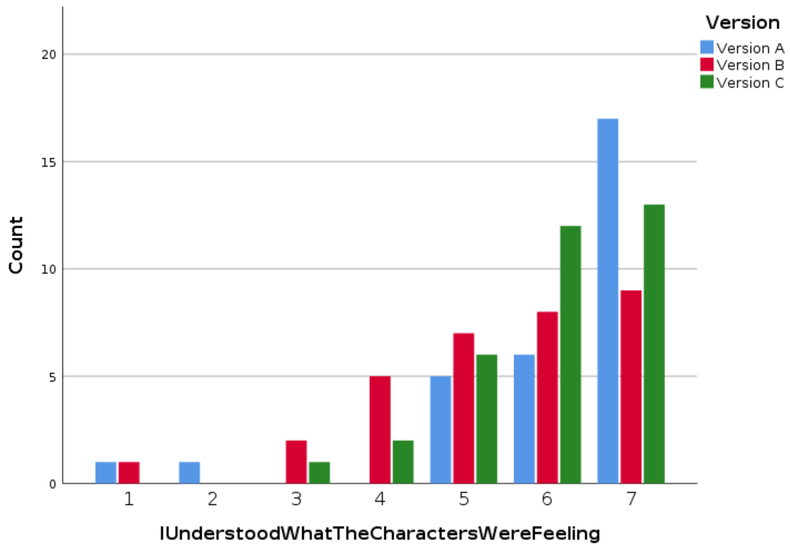


Figure D.9: Bar graph displaying the number of occurrences of each value of the Likert scale for the *Q3 - I understood what the characters were feeling* - by played version.

Q4 - I could predict the characters' feelings. For this statement, version A had a mean of 5.19 ($SD = 1.355$); version B had a mean of 4.75 ($SD = 1.796$); and version C had a mean of 5.13 ($SD = 1.476$). Figure D.10 shows the distribution of the Likert scale values by version for this question. We found no statistically significant differences between the versions ($\chi^2(2) = 2.135$, $p = 0.344$). Overall, the participants seemed to be able to predict the characters' feelings regardless of version.

Q5 - I understood the characters' intentions. For this statement, version A had a mean of 5.34 ($SD = 1.335$); version B had a mean of 5.16 ($SD = 1.588$); and version C had a mean of 4.84 ($SD = 1.417$). Figure D.11 shows the distribution of the Likert scale values by version for this question. We found no statistically significant differences between the versions ($\chi^2(2) = 2.000$, $p = 0.368$).

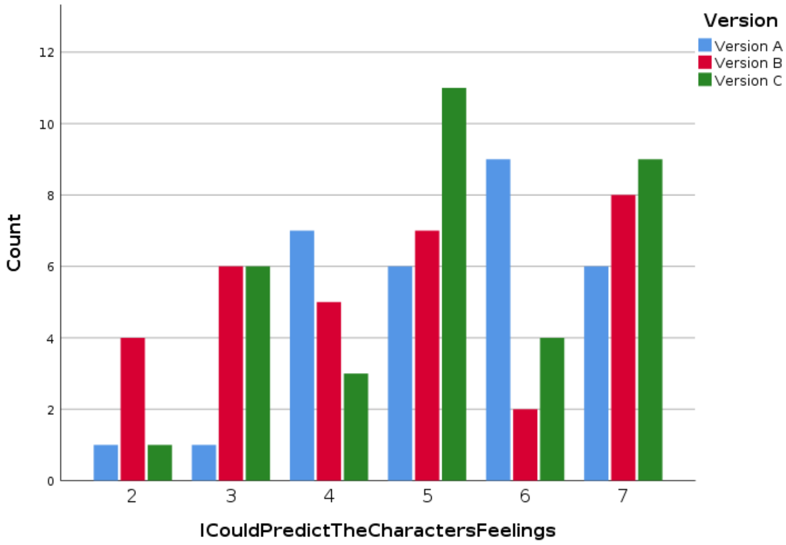


Figure D.10: Bar graph displaying the number of occurrences of each value of the Likert scale for the *Q4 - I could predict the characters' feelings* - by played version.

Overall, the participants seemed to moderately understand the characters' intentions regardless of version.

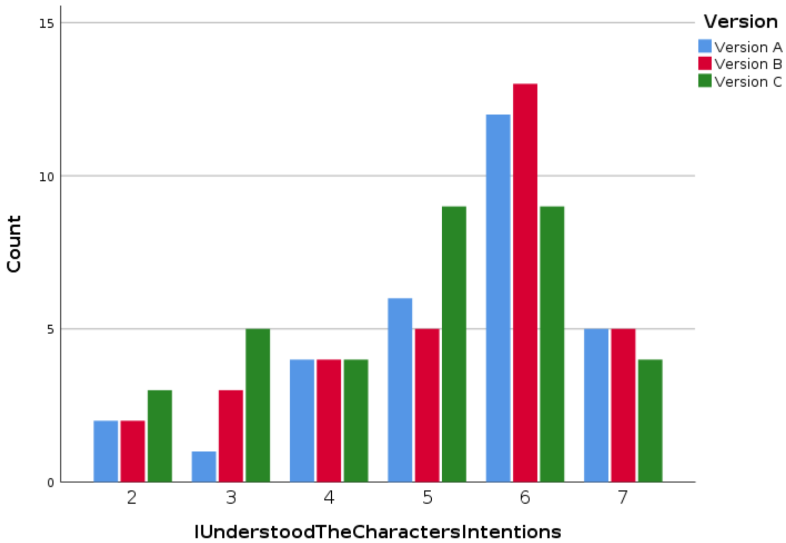


Figure D.11: Bar graph displaying the number of occurrences of each value of the Likert scale for the *Q5 - I understood the characters' intentions* - by played version.

Q6 - The characters were aware of each other. For this statement, version A had a mean of 5.34 ($SD = 1.473$); version B had a mean of 5.63 ($SD = 1.385$); and version C had a mean of 5.75 ($SD = 1.459$). Figure D.12 shows the distribution of the Likert scale values by version for this question. We found no statistically significant differences between the versions ($\chi^2(2) = 1.101$, $p = 0.577$). Overall, the participants seemed to find the characters to be aware of each other regardless of version.

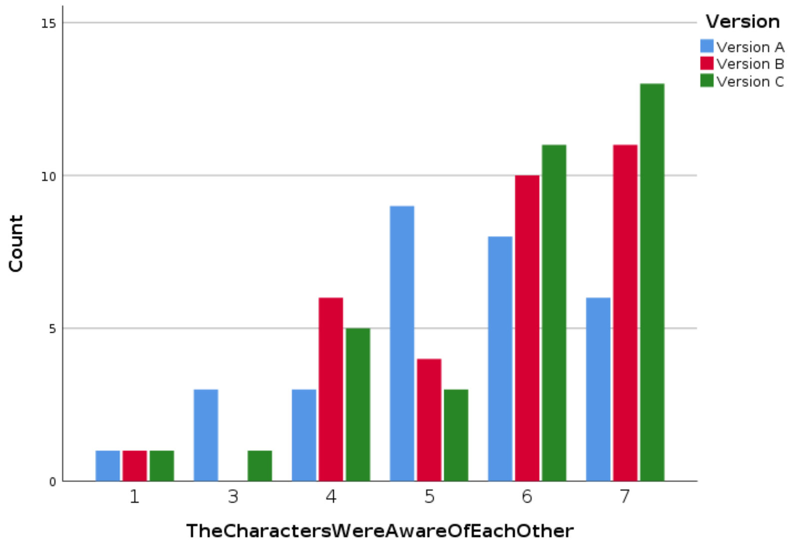


Figure D.12: Bar graph displaying the number of occurrences of each value of the Likert scale for the *Q6 - The characters were aware of each other* - by played version.

Q7 - The characters were aware of each other's actions. For this statement, version A had a mean of 4.87 ($SD = 1.737$); version B had a mean of 5.59 ($SD = 1.188$); and version C had a mean of 5.56 ($SD = 1.605$). Figure D.13 shows the distribution of the Likert scale values by version for this question. We found no statistically significant differences between the versions ($\chi^2(2) = 3.140$, $p = 0.208$). Overall, the participants seemed to find the characters to be aware of each other's actions regardless of version.

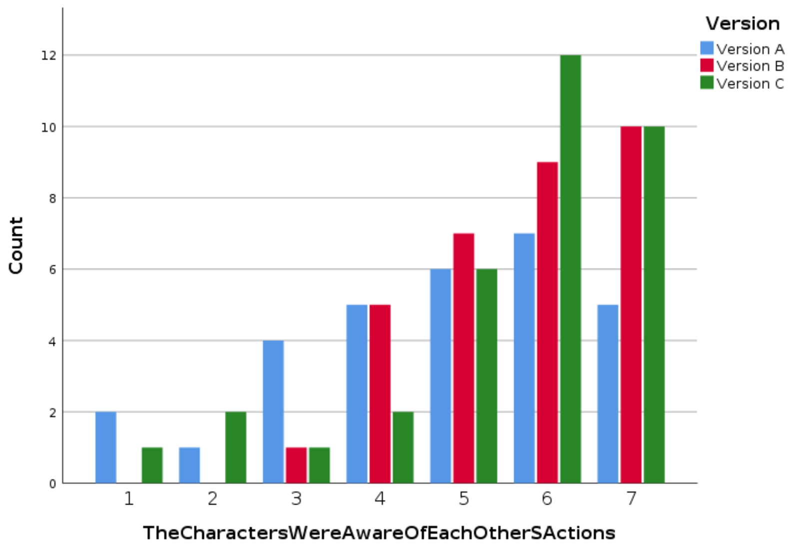


Figure D.13: Bar graph displaying the number of occurrences of each value of the Likert scale for the *Q7 - The characters were aware of each other's actions* - by played version.

Q8 - The characters could predict each others' actions. For this statement, version A had a mean of 4.28 ($SD = 1.631$); version B had a mean of 4.34 ($SD = 1.578$); and version C had a mean of 4.56 ($SD = 1.564$). Figure D.14 shows the distribution of the Likert scale values by version for this question. We found no statistically significant differences between the versions ($\chi^2(2) = 3.979$, $p = 0.137$). Overall, the participants seemed to not be able to distinguish if the characters could predict each other's actions, but there is a small tendency towards the characters being able to make the prediction regardless of version.

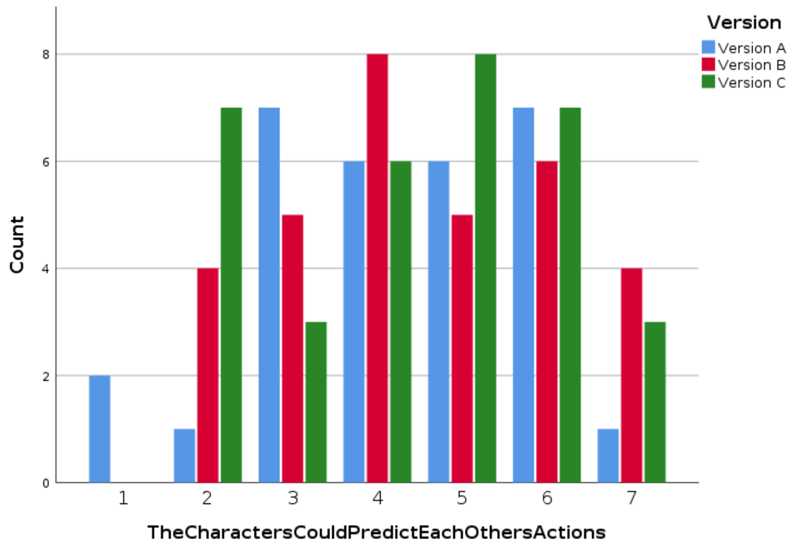


Figure D.14: Bar graph displaying the number of occurrences of each value of the Likert scale for the *Q8 - The characters could predict each others' actions* - by played version.

Q9 - The characters were aware of each other's feelings. For this statement, version A had a mean of 4.44 ($SD = 1.605$); version B had a mean of 4.72 ($SD = 1.591$); and version C had a mean of 4.66 ($SD = 1.473$). Figure D.15 shows the distribution of the Likert scale values by version for this question. We found no statistically significant differences between the versions ($\chi^2(2) = 3.980$, $p = 0.137$). Overall, the participants seemed to not be able to distinguish if the characters were aware of each other's feelings, but there is a small tendency towards the characters' active awareness regardless of version.

Q10 - The characters could predict each others' feelings. For this statement, Version A had a mean of 3.78 ($SD = 1.699$); Version B had a mean of 4.03 ($SD = 1.534$); and Version C had a mean of 4.31 ($SD = 1.378$). Figure D.16 shows the distribution of the Likert scale values by version for this question. We found statistically significant differences between the versions ($\chi^2(2) = 6.660$, $p = 0.036$). Thus, we performed the post hoc analysis. No significant differences were found between any pair: between versions A and B, $Z = -0.947, p = 0.344$; between versions C and B, $Z = -1.554, p = 0.120$; and between versions A and C, $Z = -1.434, p = 0.152$.

So, while a statistical difference was observed in the Friedman test, we found no major differences between pairs. We can still see that Version C was higher in mean value, followed by Version B, and Version A was the lowest in rank. If we wish to take meaning from these values, then in all versions participants seem to not be able to agree or disagree if characters are able to predict each other's feelings. Yet, for version A, there is a tendency for characters to be unable to perform a prediction,

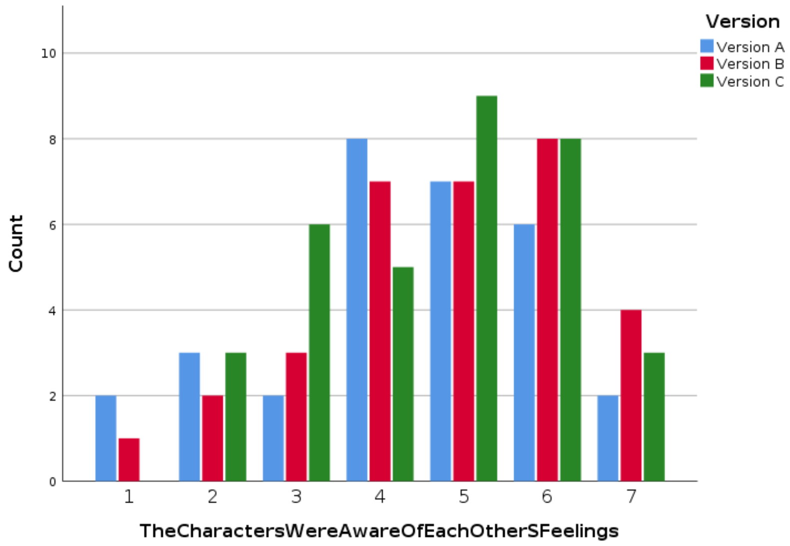


Figure D.15: Bar graph displaying the number of occurrences of each value of the Likert scale for the *Q9 - The characters were aware of each other's feelings* - by played version.

while in version C there is the opposite tendency, where characters seem more capable of predicting the opponent's feelings.

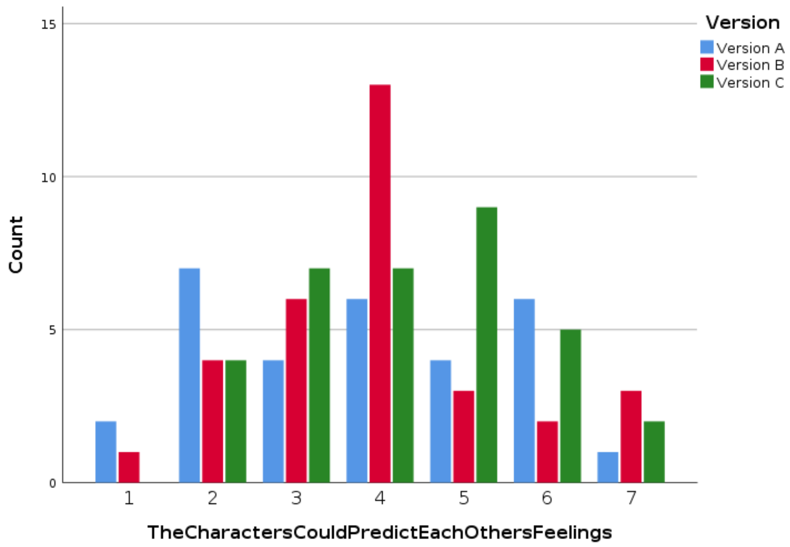


Figure D.16: Bar graph displaying the number of occurrences of each value of the Likert scale for the *Q10 - The characters could predict each others' feelings* - by played version.

Q11 - The characters were aware of each other's intentions. For this statement, version A had a mean of 4.06 ($SD = 1.625$); version B had a mean of 4.81 ($SD = 1.230$); and version C had a mean of 4.50 ($SD = 1.437$). Figure D.17 shows the distribution of the Likert scale values by version for this question. We found statistically significant differences between the versions ($\chi^2(2) = 7.814$, $p = 0.020$). Thus, we performed the post hoc analysis. No significant differences were found between versions C and B ($Z = -1.201, p = 0.230$) and between versions A and C ($Z = -1.496, p = 0.135$). We did find statistically significant differences between versions A and B, $Z = -2.811, p = 0.005$.

In this statement, Version B was ranked higher, followed by Version C, and Version A was ranked lower. Here, we can see that version B is significantly different and greater than A. However, no major differences were found between the remaining pairs. Looking at these results, we find that Version B is where the characters were perceived to be more aware of each other's intentions. Version C is a bit behind in score but is almost as good. For version A, it was unclear to the participants if the characters were aware or not of each other's intentions.

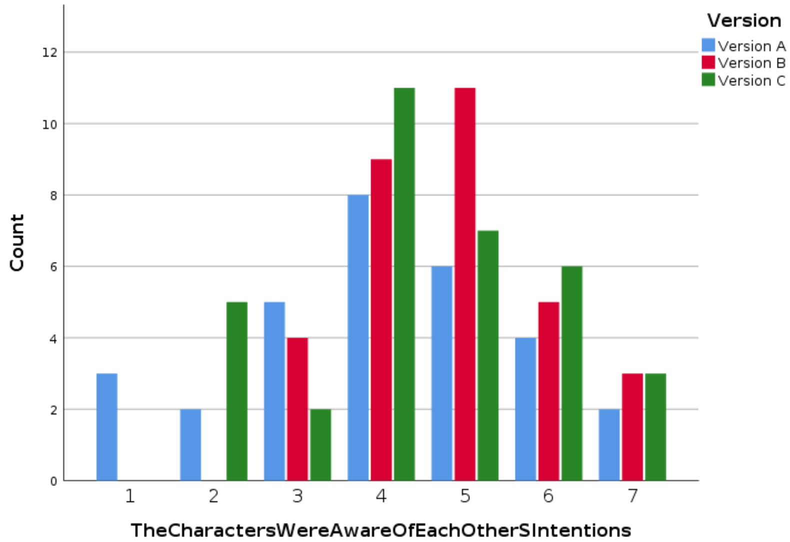


Figure D.17: Bar graph displaying the number of occurrences of each value of the Likert scale for the *Q11 - The characters were aware of each other's intentions* - by played version.

Q12 - The interaction between characters in this scene was believable. For this statement, version A had a mean of 5.53 ($SD = 1.319$); version B had a mean of 5.78 ($SD = 1.338$); and version C had a mean of 5.50 ($SD = 1.566$). Figure D.18 shows the distribution of the Likert scale values by version for this question. We found no statistically significant differences between the versions ($\chi^2(2) = 1.724$, $p = 0.422$). Overall, the participants seemed to find the interaction believable regardless of version.

Q13 - The characters displayed different personalities. For this statement, version A had a mean of 3.81 ($SD = 2.132$); version B had a mean of 4.38 ($SD = 2.121$); and version C had a mean of 4.50 ($SD = 2.110$). Figure D.19 shows the distribution of the Likert scale values by version for this question. We found no statistically significant differences between the versions ($\chi^2(2) = 2.742$, $p = 0.254$). Overall, the participants seemed to be unable to distinguish whether the characters showed or not different personalities. There is a positive tendency for versions B and C, while there is a negative tendency for version A, but due to the big standard deviation, it is impossible to infer any information from these tendencies.

Regarding the descriptions given by the participants on each character's personality, version A had the smallest number of personality characteristics given ($N = 15$ for both the axe and the sword characters). Versions B and C had the same number of personality characteristics given ($N = 21$ for

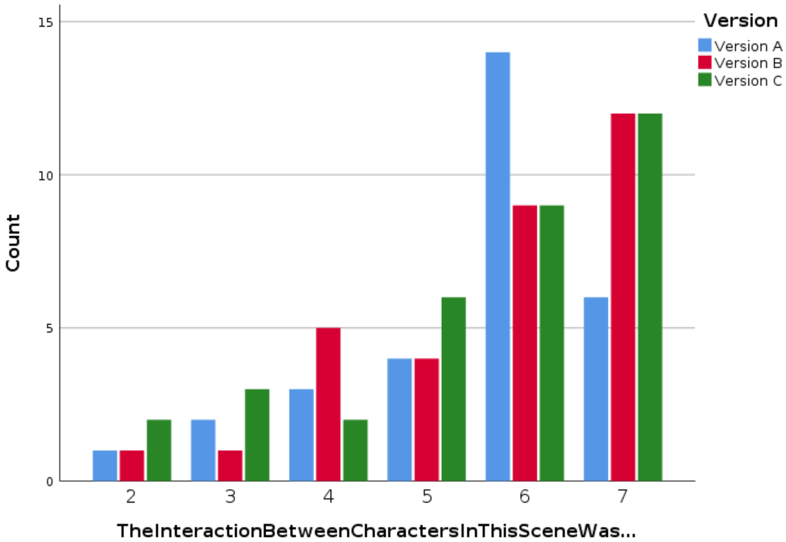


Figure D.18: Bar graph displaying the number of occurrences of each value of the Likert scale for the *Q12 - The interaction between characters in this scene was believable* - by played version.

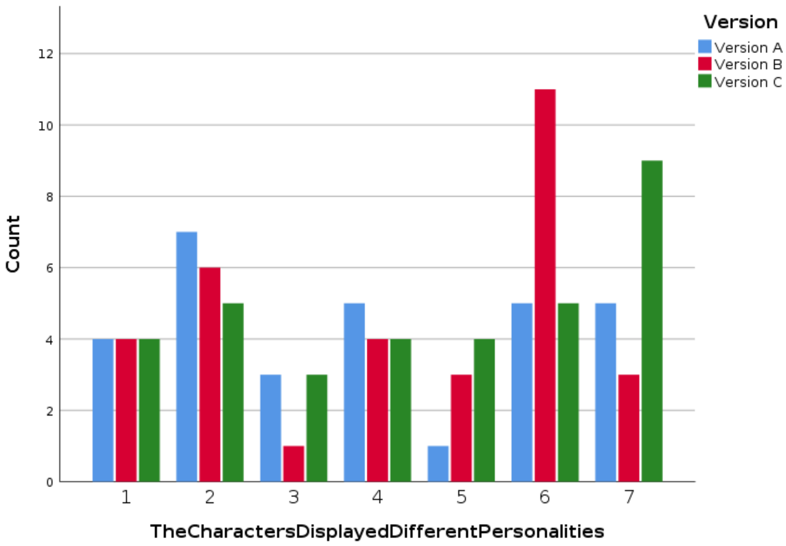


Figure D.19: Bar graph displaying the number of occurrences of each value of the Likert scale for the *Q13 - The characters displayed different personalities* - by played version.

the axe character and $N = 20$ for the sword character). With these characteristics, we performed a content analysis to better understand the differences in personalities. Recall that for versions A and B we expected the Axe character to be more defensive and the Sword character more aggressive, while for version C the Axe character would be more aggressive and the Sword character more defensive. In Table D.2 we show the more frequent terms used to describe the personalities of each character in each version. Items with fewer than 2 occurrences were removed.

Table D.2: Most frequent terms used to describe the personalities of each character in each version.

Version A				Version B				Version C			
Axe	f	**Sword**	f	**Axe**	f	**Sword**	f	**Axe**	f	**Sword**	f
Confident	8	*Aggressive*	6	*Confident*	12	*Aggressive*	3	*Confident*	12	*Scared*	11
Aggressive	3	*Confident*	4	*Aggressive*	3	*Scared*	3	*Aggressive*	9	*Confident*	4
Strong	2	*Scared*	3	Expressive	2	*Confident*	2	*Scared*	3	*Aggressive*	3
		Weak	3			level-headed	2				
						Neutral	2				

As we can see, regardless of version, the Axe character is always perceived as *Confident* (note that participants reported that this weapon was stronger than the sword, which can have an impact on the gameplay and thus the expressed emotions). Putting this expression aside, for versions A and B, there is no clear distinction between the characters, as they are both perceived as *Aggressive*, however, in version A, the Sword character is perceived more times as *Aggressive* but is also viewed as *Scared* in versions A and B. Gameplay seems to have a bigger impact on the perception of the characters than the personality in these models (*Reactive* and *Predictive*).

However, in version C, there appear to be some differences in the perception of the characters. While both are perceived as *Aggressive* and *Scared*, the Axe character seems to be more often *Scared* and the Sword character more frequently *Scared*. This aligns with how we designed the personalities. These results seem to suggest that in many cases participants will not assign personalities to the characters, but when they do, in version C, the personalities given align with those designed.

Questions on Participants' awareness of the expressions of the characters. We found that the participants were more aware of the expressions performed by the opposing character than those of the character they controlled, which is expected due to the third-person camera perspective used. Participants were aware of the expressions performed by the opposing character in 91% of the games played in version A ($SD = 0.296$); 88% in version B ($SD = 0.336$); and 97% in version C ($SD = 0.177$). Contrary to the reported awareness of the expressions performed by the controlled character of 75% of games played in version A ($SD = 0.296$); 69% in version B ($SD = 0.471$); and 69% in version C ($SD = 0.471$). We did not find statistically significant differences between the versions for the expression of the controlled character ($\chi^2(2) = 0.571$, $p = 0.751$) nor for the opposing character ($\chi^2(2) = 2.000$, $p = 0.368$).

D.3.3 Subgroup Analysis in Depth

In an effort to better understand the results, we performed a subgroup analysis. We found that for a subgroup of *participants that enjoy and have played/watched others play action/fighting games* *multiple times* ($N = 16$) there were statistically significant improvements in multiple dimensions referring to the point of view of the characters, that is, *Q7 - Q11*. Here follows an overview of each statement analyzed within this subgroup. Figure D.20 shows the averages and standard deviations of each statement for each version in this subgroup.

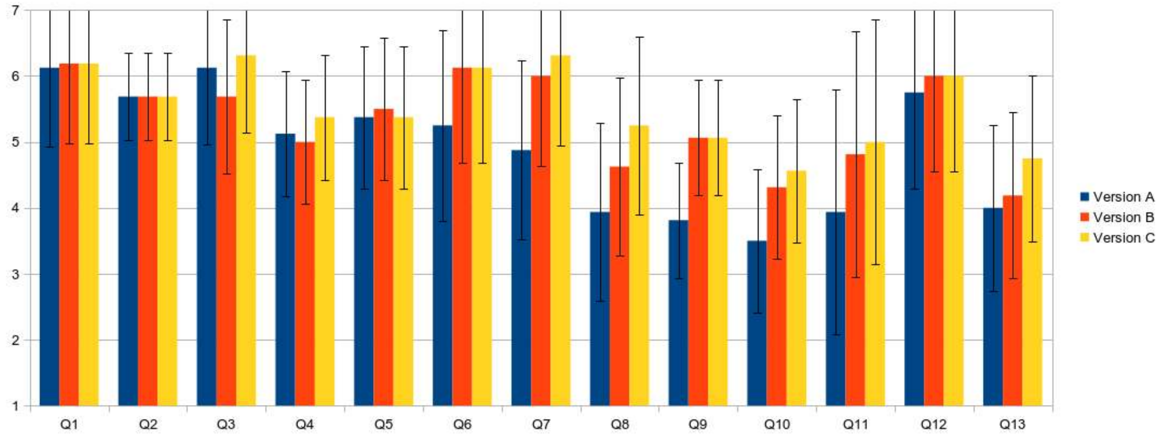


Figure D.20: Averages and associated standard deviations of each statement for each version in the subgroup analysis.

Q1 - I understood what the characters were doing. For this statement, version A had a mean of 6.13 ($SD = 1.204$); version B had a mean of 6.19 ($SD = 0.655$); and version C had a mean of 6.19 ($SD = 1.167$). Figure D.21 shows the distribution of the Likert scale values by version for this question. We found no statistically significant differences between the versions ($\chi^2(2) = 1.182$, $p = 0.554$). Overall, as in the complete sample analysis, the participants seemed to understand what the characters were doing regardless of the version. We have noted, however, an increase in averages.

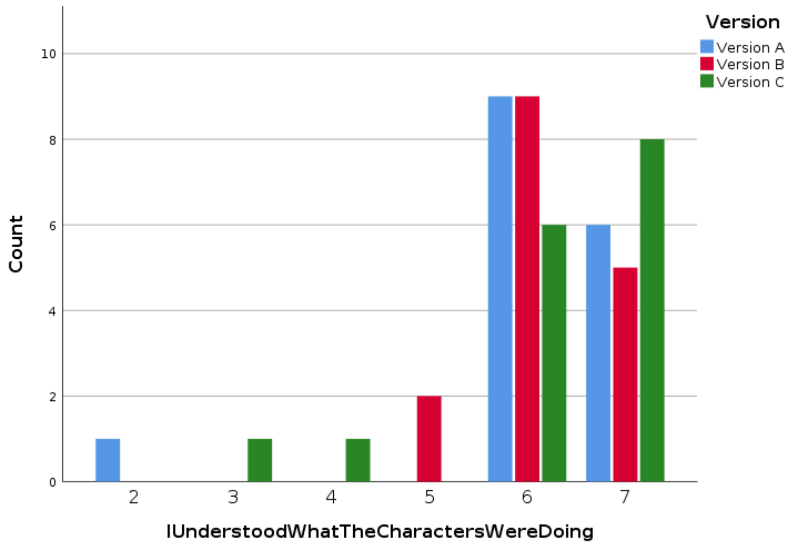


Figure D.21: Bar graph displaying the number of occurrences of each value of the Likert scale for the *Q1 - I understood what the characters were doing* - by played version in the subgroup analysis.

Q2 - I could predict the characters' actions. For this statement, version A had a mean of 5.69 ($SD = 0.946$); version B had a mean of 5.69 ($SD = 1.078$); and version C had a mean of 5.69 ($SD = 1.448$). Figure D.22 shows the distribution of the Likert scale values by version for this question. We found no statistically significant differences between the versions ($\chi^2(2) = 1.676$, $p = 0.433$). Overall, we did not find major changes compared to the complete sample analysis, the participants seemed to be able to predict the characters' actions regardless of version.

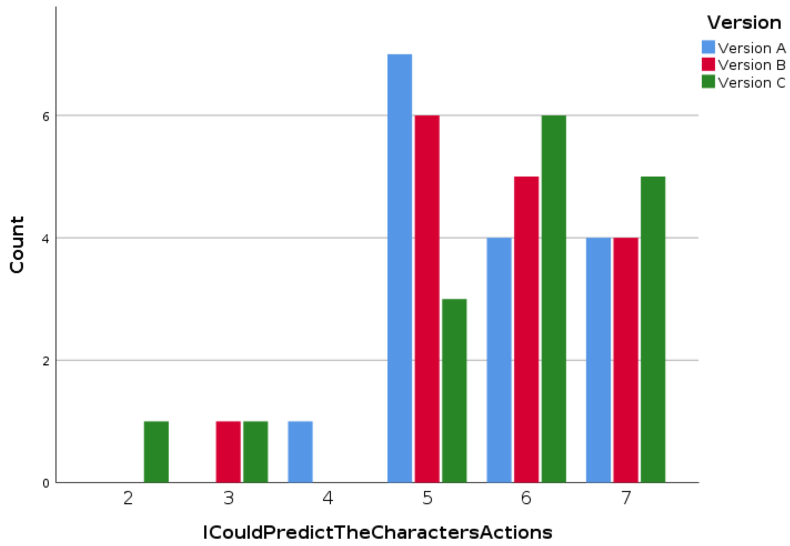


Figure D.22: Bar graph displaying the number of occurrences of each value of the Likert scale for the *Q2 - I could predict the characters' actions* - by played version in the subgroup analysis.

Q3 - I understood what the characters were feeling. For this statement, version A had a mean of 6.12 ($SD = 1.360$); version B had a mean of 5.69 ($SD = 1.352$); and version C had a mean of 6.31 ($SD = 0.873$). Figure D.23 shows the distribution of the Likert scale values by version for this question. Unlike the complete sample analysis, here we did not find statistically significant differences between versions ($\chi^2(2) = 1.512$, $p = 0.469$).

Q4 - I could predict the characters' feelings. For this statement, version A had a mean of 5.13 ($SD = 1.088$); version B had a mean of 5.00 ($SD = 1.862$); and version C had a mean of 5.38 ($SD = 1.455$). Figure D.24 shows the distribution of the Likert scale values by version for this question. We found no statistically significant differences between the versions ($\chi^2(2) = 2.039$, $p = 0.361$). Overall, we did not find major changes compared to the complete sample analysis, the participants seemed to be able to predict the characters' feelings regardless of version.

Q5 - I understood the characters' intentions. For this statement, version A had a mean of 5.38 ($SD = 1.258$); version B had a mean of 5.50 ($SD = 1.317$); and version C had a mean of 5.37 ($SD = 1.088$). Figure D.25 shows the distribution of the Likert scale values by version for this question. We found no statistically significant differences between the versions ($\chi^2(2) = 0.167$, $p = 0.920$).

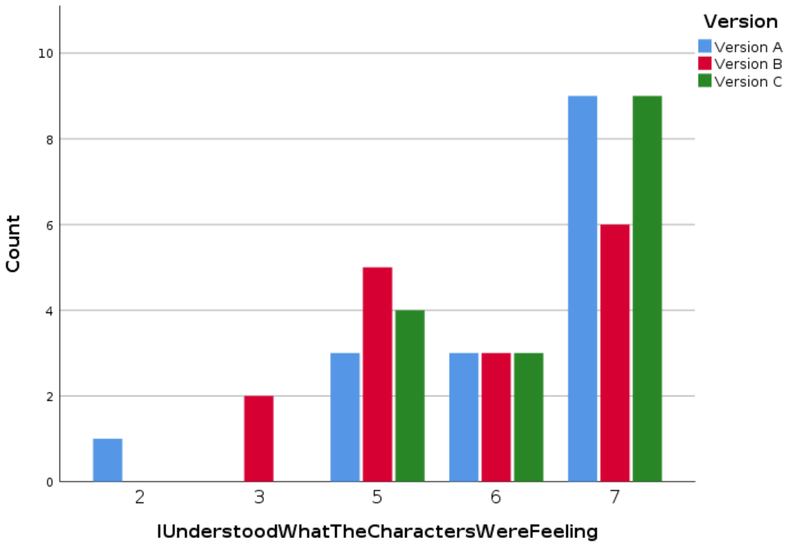


Figure D.23: Bar graph displaying the number of occurrences of each value of the Likert scale for the *Q3 - I understood what the characters were feeling* - by played version in the subgroup analysis.

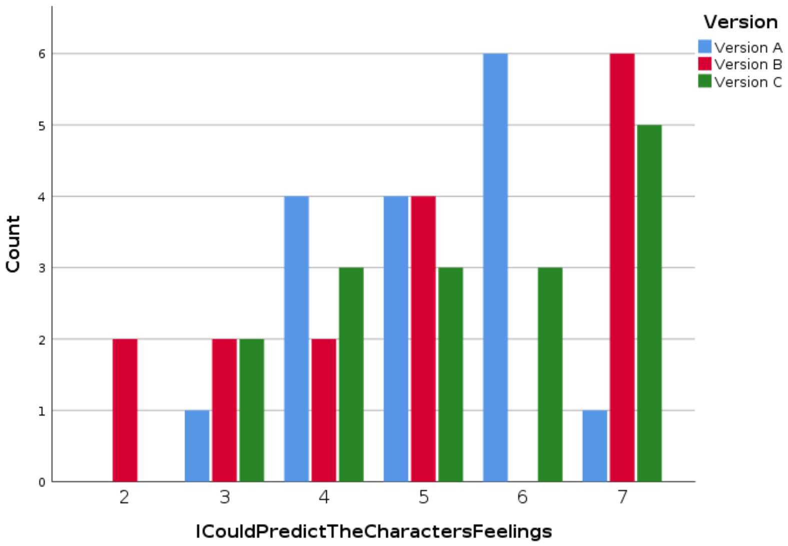


Figure D.24: Bar graph displaying the number of occurrences of each value of the Likert scale for the *Q4 - I could predict the characters' feelings* - by played version in the subgroup analysis.

Overall, we did not find major changes compared to the complete sample analysis, the participants seemed to understand the characters' intentions regardless of version.

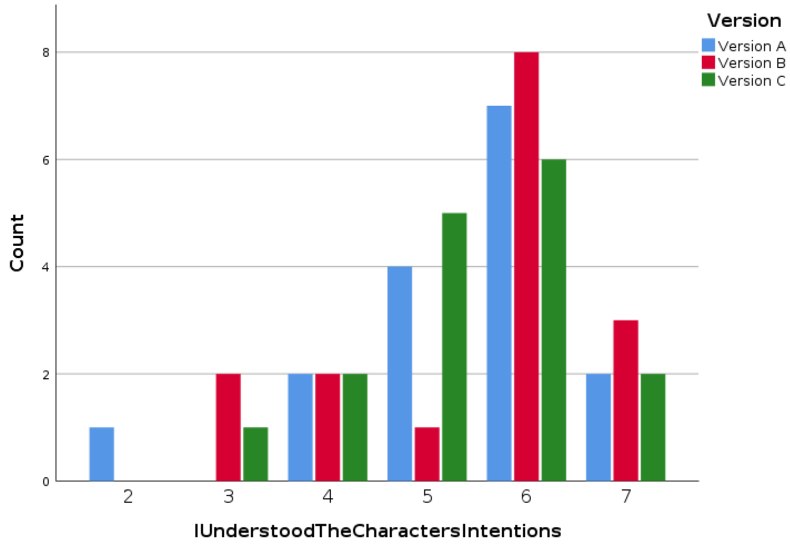


Figure D.25: Bar graph displaying the number of occurrences of each value of the Likert scale for the *Q5 - I understood the characters' intentions* - by played version in the subgroup analysis.

Q6 - The characters were aware of each other. For this statement, version A had a mean of 5.25 ($SD = 1.438$); version B had a mean of 6.13 ($SD = 0.885$); and version C had a mean of 6.13 ($SD = 1.025$). Figure D.26 shows the distribution of the Likert scale values by version for this question. We did not find statistically significant differences between versions ($\chi^2(2) = 5.907$, $p = 0.052$), but note the proximity to a significant difference in the value of p. Overall, the participants seemed to find the characters to be aware of each other regardless of version. Although the result is equivalent to the one seen in the complete sample analysis, versions B and C seem greater on average when compared to the complete sample (we see this trend repeated on multiple statements).

Q7 - The characters were aware of each other's actions. For this statement, version A had a mean of 4.88 ($SD = 1.544$); version B had a mean of 6.00 ($SD = 0.966$); and version C had a mean of 6.31 ($SD = 0.873$). Figure D.27 shows the distribution of the Likert scale values by version for this question. Unlike the full sample analysis, here we did find statistically significant differences between versions ($\chi^2(2) = 9.591$, $p = 0.008$). Thus, we performed the post hoc analysis. No significant differences were found between versions C and B, $Z = -1.249, p = 0.212$. There were, however, significant differences between versions A and B, $Z = -2.501, p = 0.012$, and between versions A and C, $Z = -2.687, p = 0.007$.

These results suggest that both versions B and C are better than version A, and while not statistically different, version C is slightly better than version B. This means that the characters appear to be more aware of each other's actions in versions B and C than in version A, where C is rated higher than B.

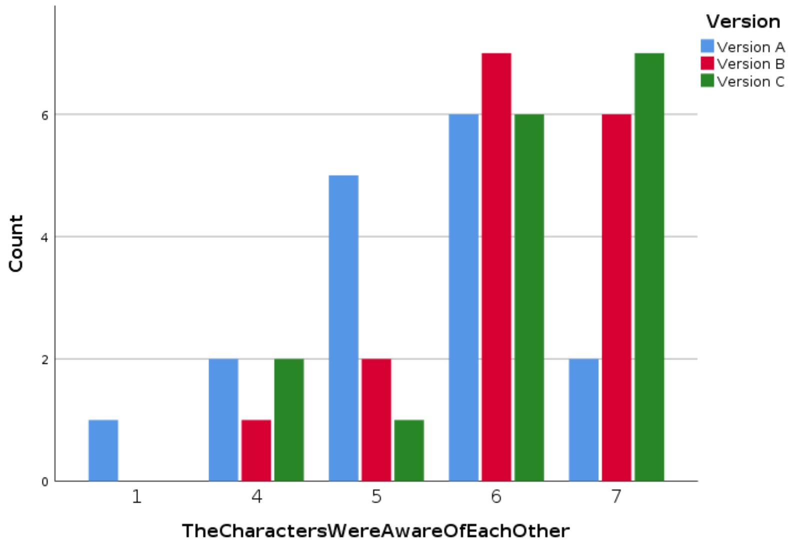


Figure D.26: Bar graph displaying the number of occurrences of each value of the Likert scale for the *Q6 - The characters were aware of each other -* by played version in the subgroup analysis.

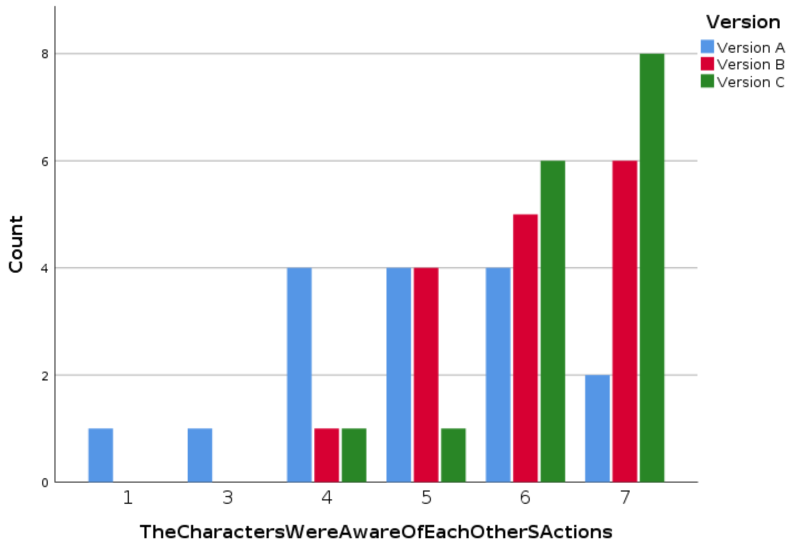


Figure D.27: Bar graph displaying the number of occurrences of each value of the Likert scale for the *Q7 - The characters were aware of each other's actions -* by played version in the subgroup analysis.

Q8 - The characters could predict each others' actions. For this statement, version A had a mean of 3.94 ($SD = 1.569$); version B had a mean of 4.62 ($SD = 1.310$); and version C had a mean of 5.25 ($SD = 1.390$). Figure D.28 shows the distribution of the Likert scale values by version for this question. Unlike the full sample analysis, we found statistically significant differences between the versions ($\chi^2(2) = 10.360$, $p = 0.006$). Thus, we performed the post hoc analysis. No significant differences were found between any pair: between versions A and B, $Z = -1.706, p = 0.088$; between versions C and B, $Z = -1.835, p = 0.067$; and between versions A and C, $Z = -2.311, p = 0.021$.

So, while a statistical difference was observed in the Friedman test, we found no major differences between pairs. We can still see that Version C was higher in mean value, followed by Version B, and Version C was the lowest in rank. If we wish to take meaning from these values, in general, the participants seemed to be unable to easily distinguish if the characters could predict each other's actions. Yet, for version A, there is a tendency for characters to be unable to perform a prediction, while in version C there is the opposite tendency, where characters seem more capable of predicting the opponent's actions.

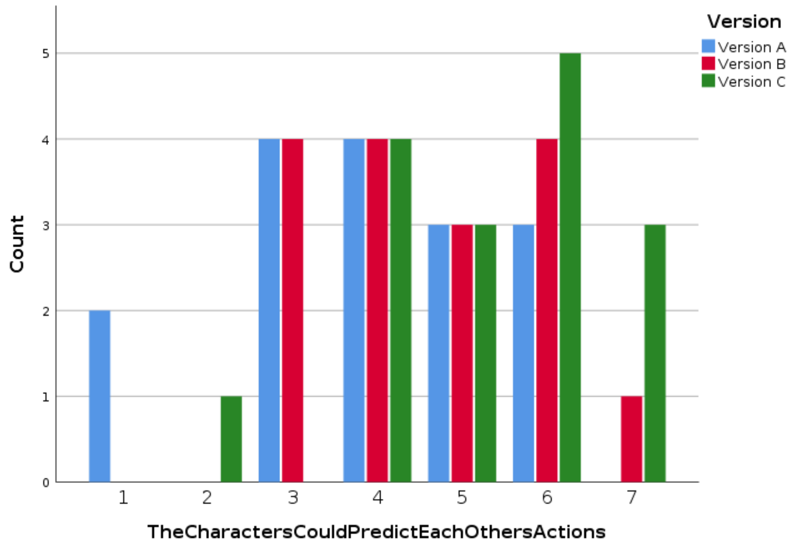


Figure D.28: Bar graph displaying the number of occurrences of each value of the Likert scale for the *Q8 - The characters could predict each others' actions* - by played version in the subgroup analysis.

Q9 - The characters were aware of each other's feelings. For this statement, version A had a mean of 3.81 ($SD = 1.642$); version B had a mean of 5.06 ($SD = 1.181$); and version C had a mean of 5.06 ($SD = 1.289$). Figure D.29 shows the distribution of the Likert scale values by version for this question. Unlike the full sample analysis, we found statistically significant differences between the versions ($\chi^2(2) = 11.148$, $p = .004$). Thus, we performed the post hoc analysis. We found significant differences between versions A and B, $Z = -2.661, p = .008$. No significant differences were found between the remaining pairs: between versions C and B, $Z = -.040, p = 0.968$; and between versions A and C, $Z = -2.348, p = 0.019$.

In this statement, we see how even though two versions have the same average, B and C, because C has a wider standard deviation, there is no significant statistical difference between C and A. What can we take from these results? In general, the participants seemed to be unable to easily distinguish if the characters were aware of each other's feelings. Yet, for version A, there is a tendency for the characters to be perceived as unaware of others' feelings, while in versions B and C there is the opposite tendency, where characters seem more aware of each other's feelings.

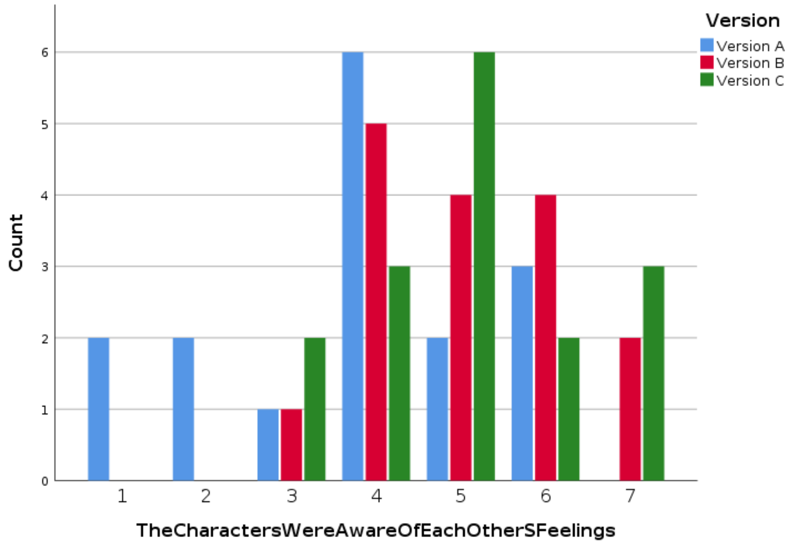


Figure D.29: Bar graph displaying the number of occurrences of each value of the Likert scale for the *Q9 - The characters were aware of each other's feelings* - by played version in the subgroup analysis.

Q10 - The characters could predict each others' feelings. For this statement, Version A had a mean of 3.50 ($SD = 1.549$); Version B had a mean of 4.31 ($SD = 1.078$); and Version C had a mean of 4.56 ($SD = 1.094$). Figure D.30 shows the distribution of the Likert scale values by version for this question. Similarly to the complete sample analysis, we found statistically significant differences between versions ($\chi^2(2) = 6.318$, $p = 0.042$). Thus, we performed the post hoc analysis. No significant differences were found between versions A and B, $Z = -2.170, p = 0.030$, and between versions C and B, $Z = -1.265, p = 0.206$. We found significant differences between versions A and C, $Z = -2.438, p = 0.015$. Note that in the complete sample analysis, we did not find any significant differences between pairs.

Similarly to Q9, here versions B and C are very similar with values pointing towards the characters being able to predict each other's feelings, while version A suggests the opposite. Yet, overall, in all versions, participants appear to be unable to agree or disagree if the characters can predict the feelings of the other.

Q11 - The characters were aware of each other's intentions. For this statement, version A had a mean of 3.94 ($SD = 1.526$); version B had a mean of 4.81 ($SD = 0.911$); and version C had a mean of 5.00 ($SD = 1.317$). Figure D.31 shows the distribution of the Likert scale values by version for this question. Similarly to the complete sample analysis, we found statistically significant differences between the versions ($\chi^2(2) = 8.341$, $p = 0.015$). Thus, we performed the post hoc analysis. No significant differences were found between versions C and B ($Z = -0.690, p = 0.490$) and between

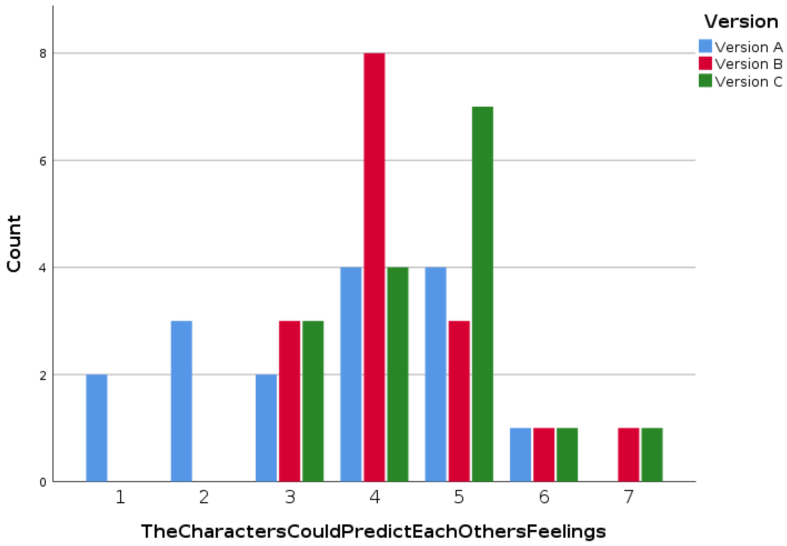


Figure D.30: Bar graph displaying the number of occurrences of each value of the Likert scale for the *Q10 - The characters could predict each others' feelings* - by played version in the subgroup analysis.

versions A and C ($Z = -2.201, p = 0.028$). We did find statistically significant differences between versions A and B, $Z = -2.488, p = 0.013$.

Although these results align with what we found in the complete sample analysis, here version C ranks highest, followed by version B, and then C. In these results, we find that version C is where the characters were perceived to be more aware of each other's intentions. Version B is a bit behind in score, but is almost as good. For version A, it was unclear to the participants if the characters were aware of each other's intentions or not.

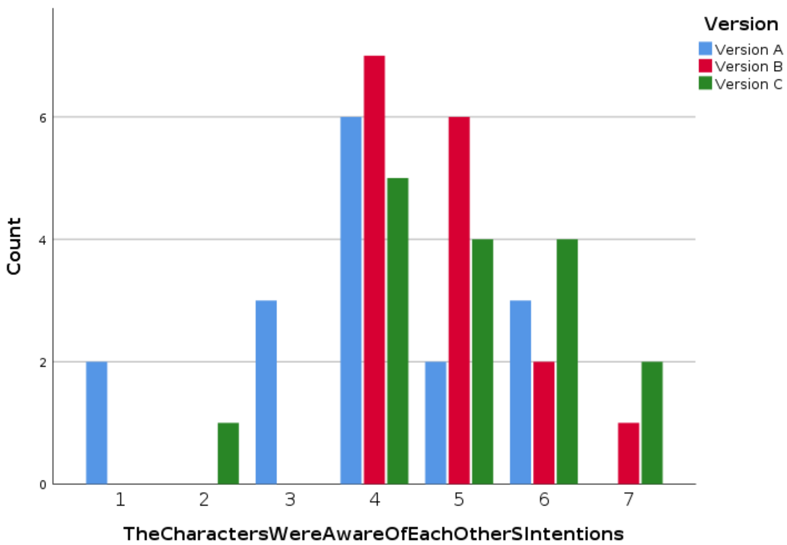


Figure D.31: Bar graph displaying the number of occurrences of each value of the Likert scale for the *Q11 - The characters were aware of each other's intentions* - by played version in the subgroup analysis.

Q12 - The interaction between characters in this scene was believable. For this statement, version A had a mean of 5.75 ($SD = 1.000$); version B had a mean of 6.00 ($SD = 0.894$); and version

C had a mean of 6.00 ($SD = 1.265$). Figure D.32 shows the distribution of the Likert scale values by version for this question. We found no statistically significant differences between the versions ($\chi^2(2) = 2.227$, $p = 0.328$). Overall, as in the complete sample analysis, participants seemed to find the interaction believable regardless of version. But unlike the former analysis, here versions B and C tied in averages.

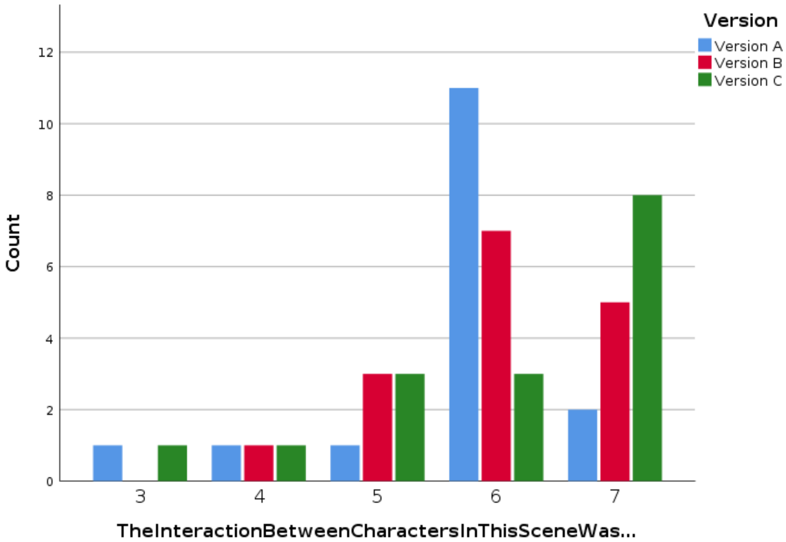


Figure D.32: Bar graph displaying the number of occurrences of each value of the Likert scale for the *Q12 - The interaction between characters in this scene was believable* - by played version in the subgroup analysis.

Q13 - The characters displayed different personalities. For this statement, version A had a mean of 4.00 ($SD = 2.160$); version B had a mean of 4.19 ($SD = 1.974$); and version C had a mean of 4.75 ($SD = 2.236$). Figure D.33 shows the distribution of the Likert scale values by version for this question. We found no statistically significant differences between the versions ($\chi^2(2) = 3.957$, $p = 0.138$). Overall, we did not find major changes compared to the complete sample analysis, the participants seemed to be unable to distinguish if the characters show or not different personalities. There is a positive tendency for versions B and C, whereas, in this subgroup, there is no negative tendency for version A.

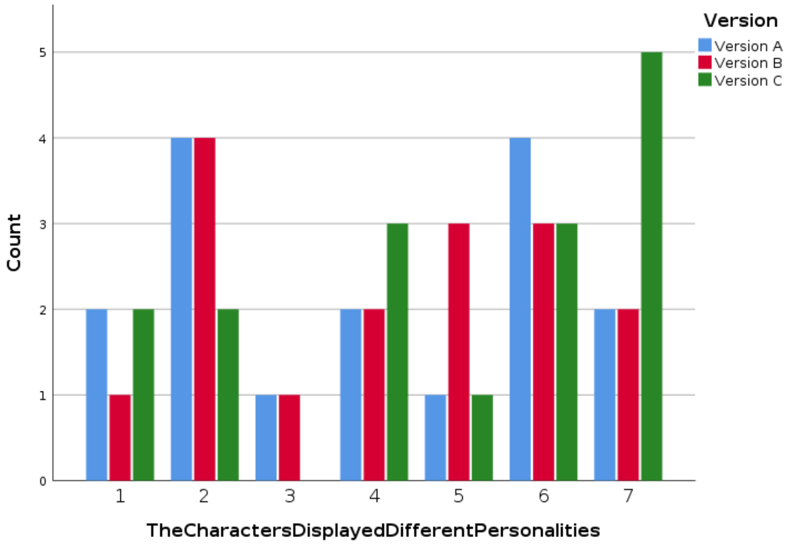


Figure D.33: Bar graph displaying the number of occurrences of each value of the Likert scale for the *Q13 - The characters displayed different personalities* - by played version in the subgroup analysis.