



Com@Viver: Using affective AI agents to encourage prosocial activity

Diogo Miguel Pimpista Antunes

Thesis to obtain the Master of Science Degree in
Information Systems and Computer Engineering

Supervisors: Prof. Ana Maria Severino de Almeida e Paiva
Prof. Paula Alexandra Nunes Da Costa Ferreira

Examination Committee

Chairperson: Prof. António Manuel Ferreira Rito da Silva
Supervisor: Prof. Ana Maria Severino de Almeida e Paiva
Member of the Committee: Prof. Sandra Pereira Gama

October 2020

Acknowledgments

I would like to thank my family for all their caring and support during my academic years, specially my parents, that gave me the opportunity to pursue my dreams, and without whom none of this would be possible.

I would also like to acknowledge the two teaching institutions that helped me make this thesis, as well as it's members. Firstly, I thank Instituto Superior Técnico, my dissertation supervisor Ana Paiva, and advisors Samuel Mascarenhas, Carlos Martinho and Rui Prada, for their insight, support and sharing of technical knowledge. Secondly, I thank Faculdade de Psicologia da Universidade de Lisboa, my dissertation supervisor Paula Ferreira, Sofia Francisco, Sónia Gomes and all the remaining members of their team, for their help both in knowledge and in all the logistics and processes regarding this study.

Last but not least, I'd like to thank all my close friends and colleagues, that helped me grow as a person, were always there for me during the good and bad times in my life, and helped me both inside and outside my school life.

All of these people made this thesis possible. To each and every one of you – Thank you.

Abstract

This study aims to gather information about how Artificial Intelligence (AI) agents can influence user behaviour, by encouraging empathy and prosocial activity, and gather data regarding players' choices and stands when dealing with cyberbullying scenarios. By exploring the capabilities of affective and prosocial computing, we changed the way the narrative of Com@Viver is managed throughout the play sessions, and adapted the narrative according to the users actions. Com@Viver is a serious educational game about cyberbullying, developed with the objective of analysing and improving bystander behaviour. The game, coupled with AI, aims to create believable cyberbullying scenarios where users play as the bystander. Bystanders play a major role in cyberbullying incidents, heavily impacting the outcome of these events depending on how they react to them, and AI has proved itself as being of great aid when dealing with psychological trauma and harmful behaviours. Com@Viver specifically targets teenagers, since these are typically the main victims and witnesses of cyberbullying incidents. Finally, we analysed the results with the data gathered from the play sessions, to conclude about the impact that our game had in the experience. The game showed positive results in changing player behaviour and increasing player self-efficacy beliefs through the usage of AI Agents.

Keywords

Serious game, artificial intelligence, affective computing, cyberbullying, agents, school study, pro-social activity, empathy, self-efficacy.

Resumo

Este estudo procura recolher informação sobre como podem agentes de Inteligência Artificial (IA) influenciar o comportamento de utilizadores, ao encorajar empatia e atividade pró-social, e recolher dados relacionados com as escolhas e atitudes dos jogadores ao lidarem com cenários de cyberbullying. Ao explorar as capacidades da computação afetiva e pró-social, mudámos a maneira como a narrativa do Com@Viver é gerida ao longo das sessões de jogo, e adaptámo-la de acordo com as ações dos utilizadores. Com@Viver é um jogo sério educacional sobre cyberbullying, desenvolvido com o objetivo de analisar e melhorar o comportamento de observador. O jogo, acompanhado de IA, procura criar cenários de cyberbullying credíveis, onde os utilizadores jogam como observadores. Os observadores desempenham um papel importantíssimo em incidentes de cyberbullying, impactando fortemente o desenrolar destes eventos dependendo de como reagem aos mesmos, e a IA tem-se comprovado como sendo uma grande ajuda ao lidar com o trauma psicológico e comportamentos prejudiciais. Com@Viver tem como público alvo adolescentes, visto que estes são tipicamente as principais vítimas e testemunhas de incidentes de cyberbullying. Por último, analisámos os resultados com os dados recolhidos das sessões de jogo para tirar conclusões sobre o impacto que o nosso jogo teve na experiência. O jogo demonstrou resultados positivos em mudar o comportamento dos jogadores e aumentar a sua noção de autoeficácia através do uso de agentes com IA.

Palavras Chave

Jogo sério, inteligência artificial, computação afetiva, cyberbullying, agentes, estudo escolar, actividade pró-social, empatia, auto-eficácia

Contents

1	Introduction	1
1.1	Problem	5
2	Related Work	6
2.1	Architectures and Models of AI Agents with Emotions	7
2.1.1	EMA	7
2.1.2	FAtiMA	10
2.1.3	Tardis	12
2.1.4	Comme Il Faut	14
2.2	Serious games	15
2.2.1	FearNot!	15
2.2.2	Façade	17
2.2.3	PromWeek	19
3	Solution	22
3.1	The Com@Viver game, agents and scoring system	23
3.1.1	Game setting and interface	23
3.1.2	Game goal and scoring system	25
3.2	Social Importance Dynamics Model	26
3.2.1	Computational Algorithm Explanation	26
3.3	Implementation	28
3.3.1	Dialogue Editor	29
3.3.2	World Model and Role Play Character Editor	29
3.3.3	Web API	31
3.4	Procedure and example	32
3.4.1	Belief Revision	32
3.4.2	Selecting the action to perform	34
3.4.3	Executing the dialog	34
3.5	Game Session Schematic	36

4	User Study	38
4.1	Study design	39
4.2	Participants	39
4.3	Instruments	39
4.3.1	Game Adaptation and gameplay questions	39
4.3.2	Interacting with Social Agents for Prosociality	40
4.4	Procedures	41
4.5	Data Analysis	42
4.5.1	Data gathered during gameplay	42
4.5.2	Pre-post group comparisons	42
4.5.3	Qualitative analysis of players' endgame responses	42
4.6	Results and Discussion	43
4.6.1	Process data gathered during gameplay	43
4.6.2	Pre-post group comparisons of players' perceptions of their interaction with the social agents	47
4.6.3	Qualitative analysis of players' endgame responses	48
4.6.4	Discussion	49
5	Conclusion	52
5.1	Future Work	53
A	Appendix	57
B	Additional Tables	65

List of Figures

2.1	FATiMA Core Architecture	11
2.2	Tardis Affective Model Architecture	13
2.3	Bullying scenario in FearNot!	16
2.4	Interaction with an Agent in FearNot!	17
2.5	User asking a question to Grace in Façade	18
2.6	Prom Week's user interface.	19
2.7	Prom Week's social exchange.	20
3.1	Com@Viver game interface.	23
3.2	Comments from 2 different players from the same group.	24
3.3	Winners endgame screen.	25
3.4	Losers endgame screen.	25
3.5	FATiMA Authoring Tool	29
3.6	FATiMA Authoring Tool - Dialogue Editor	30
3.7	FATiMA Authoring Tool - World Model	31
3.8	FATiMA Authoring Tool - Role Play Character Editor, Decision Making Action 1.	32
3.9	FATiMA Authoring Tool - Role Play Character Editor, Decision Making Action 2.	32
3.10	FATiMA Authoring Tool - Role Play Character Editor, SI Attribution rules.	33
3.11	FATiMA Authoring Tool - Web API	34
3.12	FATiMA Alternative Lines from cyberbullying case 3	35
3.13	FATiMA Server running locally on port 8080. A game session correspondent to session 3, with the identifier 34, is starting, and the necessary changes are being made to the beliefs of the first session's bullies.	35
3.14	The Decide and Execute commands shown on the FATiMA server console.	35
3.15	Session Scheme	37
4.1	Winner distribution total	43

4.2	Winner distribution per session	43
4.3	Percentage of positive comments chosen by players during cyberbullying cases 2 and 3 (Winners/Losers)	44
4.4	Percentage of positive comments chosen by players during conflict dialogues with alternative lines in sessions 2 and 3 (Winners/Losers)	44
4.5	Percentage of players that did not comment or ignored the post in case 2 (Winners/Losers)	45
4.6	Percentage of players that did not comment or ignored the post in case 3 (Winners/Losers)	45
4.7	Self Efficacy (Total)	46
4.8	Self efficacy (Winners/Losers)	46
4.9	Possibility of making choices	46
4.10	Aggressor's Behaviour Influence	46
4.11	Aggressor's Behaviour Influence (Winners/Losers)	47
4.12	CG EG comparison	48
A.1	The Com@Viver Network.	61
A.2	Com@Viver cyberbullying cases.	62
A.3	Case 3 - Nando Sapina showing prosocial behaviour due to winning groups.	63
A.4	Case 3 - Nando Sapina showing aggressive behaviour due to losing groups.	64

List of Tables

B.1	Item descriptive statistics, exploratory factor analysis parameters, reliability and correlations	66
B.2	Character dialogues with alternative roles during session 2. Roles in bold are different for the respective character, depending on the player's scores on previous sessions.	67
B.3	Character dialogues with alternative roles during session 3. Roles in bold are different for the respective character, depending on the player's scores on previous sessions.	67
B.4	Character dialogues with alternative roles during session 4. Roles in bold are different for the respective character, depending on the player's scores on previous sessions.	68
B.5	Qualitative Analysis (Part 1)	69
B.6	Qualitative Analysis (Part 2)	70
B.7	Qualitative Analysis - Category Operationalization	70

Acronyms

AI	Artificial Intelligence
AIC	Akaike Information Criterion
API	Application Programming Interface
BDI	Beliefs, Desires, Intentions
CIF	Comme il Faut
CFI	Comparative Fit Indices
CG	Control Group
CKB	Cultural Knowledge Base
EMA	EMotion and Adaptation
EG	Experimental Group
IFI	Incremental Fit Index
IT	Information Technology
IAT	Integrated Authoring Tool
ISAP	<i>Interacting with Social Agents for Prosociality</i>
KMO	Kaiser-Meyer-Olkin
RMSEA	Root Mean Square Error of Approximation
SRMSR	Standardized Root Mean Square Residual
SID	Social Importance Dynamics
SI	Social Importance
SNS	Social Network Site
ULS	Unweighted Least Squares

1

Introduction

Contents

1.1 Problem	5
-----------------------	---

With the increasing influence of Information Technology (IT) in our lives, children and adolescents spend growing amounts of time online using the internet for a multitude of purposes, like watching videos, playing games and maintaining social contact with each other through social media. As society adapted to the heavy presence of IT in daily life, education itself had to adapt as well. The usage of IT for the purpose of educating lead to the development of Serious Games, games that are designed with a purpose besides entertainment.

These games have been mostly used for teaching and learning purposes, as well as promoting behavioural change and anti-bullying, and have been proved to have an impact on affect, motivation, perception, acquiring knowledge and better understanding. [1] Serious games, and video games in general, have the ability of having an effect on player's emotions, either intentionally or unintentionally, in the same way that a movie or a song can.

Games have also been a subject of particular interest to the Artificial Intelligence (AI) field over the years, being it card games, board games, strategy or even role-playing games, with the goal of creating rational agents capable of evaluating the game and achieving the best possible outcome. A possible definition of AI is "the field that studies the synthesis and analysis of computational agents that act intelligently" [2]. While an Agent is something that acts in an environment, an Intelligent Agent needs to show some signs of intelligence in its course of action [2]. We can say that an Agent acts intelligently when [2]:

1. What it does is appropriate for its circumstances and its goals,
2. It is flexible to changing environments and changing goals,
3. It learns from experience,
4. It makes appropriate choices given its perceptual and computational limitations.

The engineering goal of AI is the design and synthesis of useful intelligent artefacts, like agents. The building of agents that act intelligently is useful in many applications. Some examples include giving emotional support [3], social support to children with autism [4], and, for instance, serious games. AI and AI Agents can have an impact in the overall effect and credibility of these games, as well as the user experience.

To better understand the way serious games and AI relate, it is crucial to note that an important aspect in educational games that aim to provoke behaviour change in the user is the necessity of stimulating empathy [5].

Empathy has been defined as "a cognitive ability which facilitates the understanding of the emotions of another person" [6]. By making AI Agents able to act according to simulated emotions, and making that behaviour credible enough, we can establish a connection with the player. The creation of the

empathic relationship between player and character is a way of making the player care about the situation at hand, giving him a sense of responsibility, and make him feel like the actions taken actually matter for the outcome. This “emotional behaviour” in Agents can be expressed through a change in the situation or facial expression. It is also important to note that humans feel more empathic towards people of similar traits or with whom they have a close relationship [5], so having character diversity in certain scenarios is important for preventing a biased effect on collected data.

As a result of the increasing influence of IT, particularly the social aspect, cyberbullying has emerged. It can be defined as ‘any behaviour performed through electronic or digital media by individuals or groups that repeatedly communicate hostile or aggressive messages intended to inflict harm or discomfort on others’ [7]. It is a complex issue, with potential lifelong consequences, that does not have a simple fix, and it can happen via e-mail, instant-messaging, Internet chat rooms, multiplayer online games, social websites, etc. This kind of behaviour involves 3 types of individuals: victims, bullies and bystanders [5].

Being a bystander can be defined as “those who are present but do not interact” [5], meaning that they are only observers. In bullying acts, bystanders’ decision to intervene is linked to reinforcement of the cyberbullying behaviour, and positive behaviour, like acting against the bully, asking for adult help, or defending the victim, can break the cyberbullying cycle [5]. On the opposite, negative bystander behaviour leads to the repetition of the cycle [5] and reinforces the bullying. Bystanders’ inability to take action is also considered negative bystander behaviour since victims and bully may consider it as a sign of approval with regards to the bully’s actions [5]. This lack of action is known as the Bystander Effect.

This concept is directly related to the lack of pro-social behaviour. Pro-social behaviour occurs when someone does something to help another [8], for example, a person helping a stranger, donating blood, doing volunteer work, etc. The development of technology has lead to situations where an individual may feel less responsible or accountable for their actions, leading simultaneously to an increase in bystander behaviour, and lack of pro-social behaviour. An example of this situation would be social media, where the natural distancing and anonymity is a key factor.

The necessity to fight against this effect has lead to the concept of Pro-social Computing, defined as “computing directed at supporting and promoting actions that benefit the society and others at the cost of one’s own” [8]. This could be directly applied to the virtual bystander effect and cyberbullying, due to the technological connection.

In this context, the empathic relationship is even more important, since making a person “care for” a situation increases the likelihood of said person acting upon something they disagree with, helping others, breaking the bystander effect and increasing prosociality. The notion of self-efficacy is also important, and is described as the belief an individual has regarding his/her own ability to accomplish something. This relates to the bystander effect in the sense that observers may feel hesitant to influence the ongoing bullying, due to their self-perceived inability to do so. Studies have shown that Education,

both of children and adults, and increasing awareness, are effective in tackling this problem. Also, considering IT is the main culprit, one of the main ways of tackling this problem is through the use of IT.

With all the previous statements in mind, the game Com@Viver [5] was created. A serious game that targets bystander behaviour and provides youngsters with an opportunity to learn about cyberbullying, while monitoring their actions to further understand their decision making process. The game Com@Viver was designed as a representation of a Social Network Site (SNS), where the player is presented with multiple cyberbullying scenarios, and can make various decisions regarding his stand on said scenarios. It is divided in 5 sessions (1 diagnostic session and 4 normal sessions) and has a save and load mechanism so the users can continue their progress.

Players, which are children from the 7th through 9th school years, form groups of 3 to play the game. They are placed in a virtual class of 15 children, where the remaining 12 users of the SNS are AI agents. The scenario consists of a social network platform with the user's name and profile picture, as well as posts and comments created by the other users (AI Agents). Players then interact with the game by liking or disliking posts and commenting below them. Once during every session there will be a cyberbullying case, where one of the AI Agents, the aggressor, will make a public post on the social network, directed towards one or more victims, which are also always AI Agents, since the users always play in the position of a bystander. The user is presented with multiple possibilities of reaction he can post in the comment section, and that directly affect his score, based on their negative or positive nature. There is also a messenger online chat where the user will be asked what he thinks about the whole situation, having once again multiple response options.

After all posts appear, the session ends, and the user must answer a post-session questionnaire about how he felt during the experiment. It is important to note that players should and will never know explicitly which response options will affect their score. The players must deduct that, upon acting in a pro-social manner towards the victims, those actions will affect their score in a positive manner. It is important to not present this information explicitly so that the children do not engage in a "goal-chasing" while playing the game.

Although the game Com@Viver has shown positive results in improving the players behaviour, both in encouraging pro-social behaviour amongst the users, and raising awareness towards cyberbullying and the bystander effect, we believed some aspects regarding the overall structure, narrative and player interaction could be better. The previous study, based on the previous iteration of the game, did not evaluate the extent to which the players' behaviour, during the experiment, could impact the game agents, and if the player could perceive this impact. The game made it seem as if the player's choices and actions did not have an actual impact on the narrative. Also, we believe further display of positive impact could help players improve their self-efficacy beliefs towards changing the behaviour of others. That is the motivation behind this project.

1.1 Problem

According to previous studies and investigations [9], the majority of serious games created for tackling the thematic of bullying to not engage in the theme of cyberbullying. Furthermore, the games centred around the cyberbullying theme mostly aim at raising awareness for the problem and teaching strategies to help students face the problem, leaving aside the importance of empathy, or trying to change the players behaviour during the experiments. It is also important to note that nearly all of the experiments done with video games in this field of research are played in a single session [9].

Although the usage of different scenarios and the simulation of cyberbullying situations is nothing new, during our research we have not found any studies focusing specifically on serious games for the purpose of verifying the relation between behavioural change in game characters, and behavioural change in players. Considering that allowing players to experience the consequences of different in-game choices seems to be an important way of promoting empathy, awareness, and constructive behaviour [9], the aim of this thesis is to implement AI in the previously developed game Com@Viver, giving the existing characters the decision making capabilities and rationale akin to those of an intelligent social agent, and therefore achieving a bigger impact on the players and greater effectiveness in incentivizing pro-social behaviour.

Given this necessity, our main objectives with this study were to implement an AI agent architecture into the game Com@Viver, measure the impact of AI agents in promoting pro-social behaviour, and measure the impact of AI agents in influencing the perception of players regarding their self-efficacy in behavioural change. Considering these objectives, and utilizing the Com@Viver platform with our implemented solution, we propose to answer the following research questions:

1. Can using affective AI agents encourage prosocial behaviour?
2. Do players perceive that their actions can affect the behaviour of AI agents in a game?
3. Can players' self-efficacy to promote pro-social behaviour improve by interacting with AI agents?
4. Does players' perception of the behaviour of AI agents change through interaction in a game?
5. Can players' behaviour change in the game specifically with regards to the changes they observe in the agents' behaviour?

For this purpose, we created a multitude of questions that were asked prior and after the game sessions, and also after the 2nd, 3rd and 4th game sessions. The study divided the players into 2 groups, were one played the game, while the other group "played" a paper version of the game. We also analysed the actions taken by players during the experiment. The results will be discussed in Chapter 4.

2

Related Work

Contents

2.1 Architectures and Models of AI Agents with Emotions	7
2.2 Serious games	15

In order to correctly develop our intended solution, we first had to analyse previous examples of similar implementations, and fully understand the background of our problem, being it the creation of more believable AI Agents for the purpose of educating children towards pro-social behaviour. With this goal in mind, it is important to note that emotions have a major impact on essential cognitive processes, and play a role not only in human creativity and intelligence, but also in rational thinking and decision-making.

We have all heard of the concept of “acting too emotionally”, but decision-making without emotion can be just as impairing as decision-making with too much emotion [10], and is popularly described as “acting like a robot”. Thus, one may deduct that to build computers that make more intelligent decisions, they may be required to “have emotions.” This knowledge led to the rise of a field of science called Affective Computing, which means “computing that relates to, arises from, or influences emotions” [10].

Affective computers, that have the main goal of interacting more naturally and intelligently with humans, need the ability to express or at least recognize affect, the behavioural expression of one’s mood. By recognizing affect, new applications become possible, that may be applied in areas like learning, communications, health, and human interaction. The primary way of computationally process affection is through the recognition and synthesis of facial expression, body language, and voice inflection (pitch and tone), but it is also possible to demonstrate emotional behaviour through text [10].

By making a computer simulate the expression of an emotion, the users may share or experience that emotion momentarily, and making it take the users affective state into account when going through its decision making process can make it perform its function better.

Ultimately, the goal is using affecting computing to perceive and express affective states, creating a more immersive and realistic experience for the user. Considering the referred factors, we investigated AI Agent Architectures, which are computational structures and processes utilized for the creation of believable AI Agents, specifically ones that enable said AI Agents the capability of simulating emotions. We also investigated serious games created around the cyberbullying theme.

2.1 Architectures and Models of AI Agents with Emotions

Agent architectures in computer science work as a blueprint for AI agents, depicting the structure arrangement of its components. In this section we focus on some of the existing agent architectures and models that are relevant for this project.

2.1.1 EMA

EMotion and Adaptation (EMA) [11] is a model based on the theory of appraisal, which states that emotions arise from a person’s interpretation of their relationship with the surrounding environment, and such interpretation can be mediated by cognitive processes that can be described through appraisal

variables (desirability, accountability, likelihood, etc). The model realizes a set of theoretical assumptions and requirements in order to achieve computational modelling of emotions.

Regarding the theoretical requirements, the model defends that certain inferences are necessary to distinguish between emotions:

1. **Relevance, valence and intensity:** a computational model must represent events, actions and their immediate consequences, as well as the valence (positive or negative) and intensity (importance) of these consequences for the agent;
2. **Future Implications:** a computational model must represent future goals and expectations, and must include mechanisms for assessing the likelihood of events, actions and their consequences, including interactions between possible outcomes (if achieving one goal interferes with achieving another);
3. **Blame and responsibility:** the model must represent some notion of causality and agency, as well as other actors' motivational and knowledge states such as intention (did they intend to hurt me?) and awareness (did they know this was going to happen?);
4. **Power and coping potential:** the model must be open to subjective reinterpretation (e.g. represent subjective rather than "true" beliefs), be able of representing the extent to which events can be controlled (e.g. how robust is my plan?), have some representation of relationships between agents such as representing different agents' spheres of influence or organizational hierarchies, and consider not only the individual's external power (over the world and other individuals) but also his or her internal power (e.g. one's ability to abandon a wanted goal);
5. **Coping strategies:** the model must provide mechanisms for translating patterns of appraisal into appropriate external actions or changes to the current configuration of beliefs, desires, intentions and plans.

Due to the ambiguities and conflicts between theorists of appraisal theory, a number of process assumptions were also adopted:

1. **Appraisal causes emotion:** appraisal processes are the cause of emotional responses
2. **Cycle of appraisal and re-appraisal:** it is assumed a cyclical relationship between appraisal, coping and re-appraisal, where the agent's initial appraisals of a situation causes a variety of cognitive and behaviour responses that change the person's relationship to the environment, and the process repeats itself.
3. **Appraisal is shallow and quick:** appraisal processes are always fast (reactive), parallel, and unique (single-level).

EMA realizes the assumptions and requirements stated previously. In regards to representing knowledge, the “agent-environment” relationship is illustrated with a mixture of symbolic and numeric representations. States and actions represent the world as a conjunction of propositions. For example, for representing a state where an Agent would be waiting and sat on a bus stop while the bus approaches.

WAITING $\wedge$ $\neg$ STANDING $\wedge$ TICKET-HAVE $\wedge$ BUS-APPROACH $\wedge$ CATCHING-DISTANCE

Actions are represented with preconditions and effects. For example, a CATCHING-THE-BUS action has the preconditions that the agent is standing, has the ticket, and the bus is within catching distance, and with the effects that the ticket will be given away and the bus will no longer be approaching.

States and actions are annotated with epistemic variables representing the beliefs, desires and intentions of agents in the situation. The model allows the agent to distinguish between act intention (agent X intends action A) and outcome-intention (agent X intends effect E to occur). This allows the model to represent unintended effects of actions. The model also represents probabilities, to represent uncertainty over these intentions.

In addition to states and actions, the model represents several relationships between actions and states. The effect of some action may establish a precondition of some other action, or negate it. “Establishment relations” establish a pre-condition, while “Threat relations” block them. For instance, the RAISE action has the effect STANDING, which is a pre-condition of the action CATCHING-THE-BUS. This same action has the effect $\neg$ TICKET-HAVE, which disables the precondition of catching the bus again, since you can’t use the same ticket to catch another bus and would have to buy a new one.

The concept of probability and utility is also important. Utilities represent the preference agents’ have for states. For example, the agent may assign a high value of utility to catching the bus, but while waiting for it he might not give a high value of utility to sitting down, since he may prefer to wait while standing up. Probabilities over states represent the agent’s certainty in the truth-value of the state at some point in time, for example knowing for sure that a bus is coming. Probabilities over actions represent the likelihood that an agent intends to execute an action, and the probability that the action can be executed. The action CATCHING-THE-BUS may be fully intended by the agent, but may not be achievable with 100% certainty, since the bus can be full on a busy day and will not accept more passengers.

EMA needs to support a set of cognitive operations in order to allow perceptual changes, form new inferences, initiate/terminate actions, etc. Some of these operations are “Listen”, “Speak”, “Wait”, “Update Plan”, “Initiate-Action”, amongst others. In addition, every proposition, either in the past, present or future, is appraised, and a continuously updated set of appraisal values are associated with each proposition. These include Relevance (if it has utility for the agent), Perspective (i.e., from the perspective of the agent catching the bus), Desirability, Likelihood, Expectedness (i.e. if the agent is catching the bus, and the bus is stopping to pick up the agent, it is highly expected that the bus is not full and the agent will be able to enter the bus), Causal attribution (who deserves the credit or blame), Controllability (can

the outcome be changed by the agent) and Changeability (can the outcome be changed by external factors).

The model also supports a notion of mood. The appraisal determines the agent's responses, but this is biased by an overall mood state. Multiple appraisal frames are maintained, labelled with a specific emotion type and intensity, and each competing to determine the agent's response. For example, an undesirable and uncontrollable future state (e.g., it looks like the bus is not slowing down and it might not stop), would be labelled as fear-eliciting. The same frame might generate multiple emotion labels. For example, an unexpected and beneficial outcome could elicit both joy and surprise, like the bus arriving earlier than expected. Agent's mood represents a summary of various appraised events, it is not intentional and tends to change over time as appraisal frames are added or removed due to changes in interpretation. The mood state is represented as a set of emotion labels (e.g., Hope, Joy, Fear, etc.) with an $[0..1]$ intensity that is based on all appraisal frames. For example, if EMA has several appraisal frames labelled with hope, the intensity of these frames are added and mapped into the range of zero to one, and serve as the hope component of the mood state. The mood state has an indirect effect on appraisal since EMA applies a mood adjustment to appraisal. For example, if an appraisal frame is labelled with hope and has an intensity of X , the mood-adjusted intensity of this frame is $X + \text{Mood}(\text{hope})$.

In sum, this model is based on decision-theoretic planning, with explicit representation of intentions and beliefs, the ability to detect future benefits and threats (and the causes and agents associated with them), as well as notions of probability, utility, desirability, mood, amongst others.

The relation between EMA and our solution are the concepts in common. The model utilized in our solution (explained in chapter 3) also utilizes forms of representation of actions with consequences, preconditions and effects, as well as subjective beliefs, desires, intentions, goals and cyclical agent functionality. It also features the notion of importance, causality, agency, utility, interactions between outcomes, relationships between agents, and cognitive operations. From a functional architecture standpoint, it is very closely related to our implementation.

2.1.2 FAtiMA

FAtiMA (Fearnot Affective Mind Architecture) [12] is an Agent Architecture that uses emotions and personality to influence the agent's behaviour. It is composed of a core layer (named FAtiMA Core) on which components are added to add functionality. The Core is a template that defines how the Agent Architecture works. Added components can provide specific implementations for the generic functions defined in the Core. Figure 2.1 shows a diagram of FAtiMA Core with the basic functionalities for an emotional agent architecture.

The Core architecture divides appraisal into two separate processes. Appraisal derivation is responsible for evaluating the relevance of an event to the agent and determine a set of appraisal variables

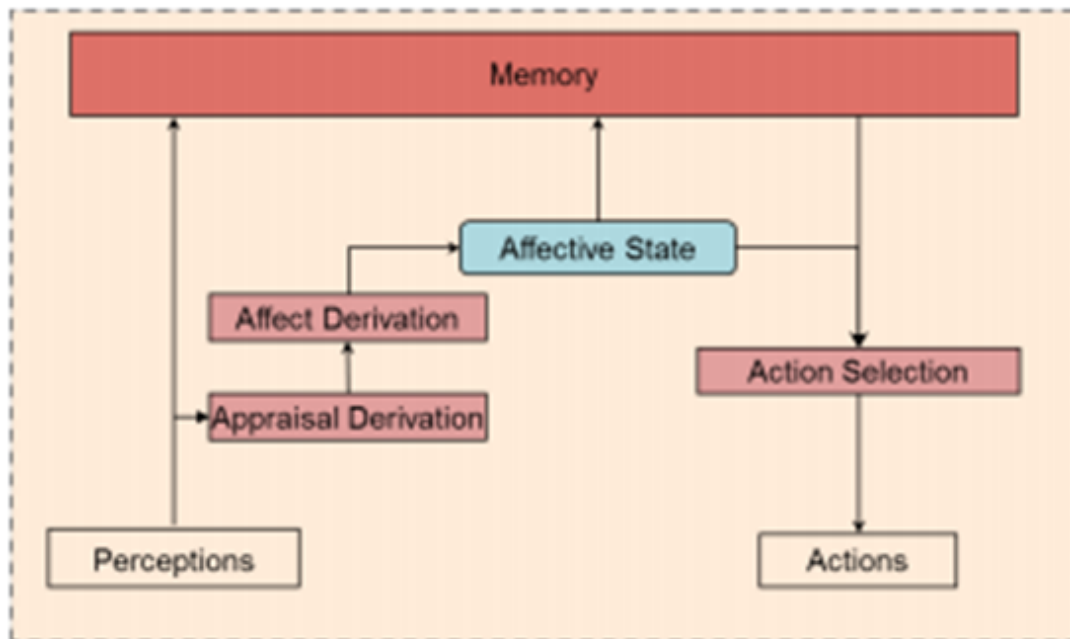


Figure 2.1: FAtiMA Core Architecture

(e.g. desirability and desirability for others). Affect derivation takes the appraisal variables as input and generates the resulting affective states (emotions or mood).

An agent is able to receive perceptions from the environment (events) which are used to update its memory (or internal state) and to trigger the appraisal process. The result of the appraisal process is then stored in the affective state, and used to influence the action selection processes, making it seem like the agent acts upon the environment.

However, a FAtiMA agent that only has a Core will not do anything. Behaviour is added by adding components that implement one or multiple of the functionalities mentioned in Figure 2.1.

To differentiate components when adding them to the Core, they are categorized according to the functionality. For instance, an AffectDerivation Component will have to implement an Affect Derivation process, and a Behaviour Component will have to implement an action selection function.

For example, the architecture can be defined with the following components:

Reactive Component - this component uses predefined emotional reaction rules to determine the value of the appraisal variables (e.g. Desirability, DesirabilityForOthers, Praiseworthiness and Like). When an event is perceived, the reactive appraisal matches the event against a set of emotional rules. A rule may define a particular value for each of the appraisal variables and can then target a specific event (e.g. the agent finds it desirable whenever it receives a compliment from agent B) or it can be more general (e.g. the agent finds it undesirable whenever the action cry is performed).

OCCAffectDerivation Component - generates emotions from the appraisal variables. For instance, an event with a positive desirability value for the agent will generate a Joy emotion if it surpasses the agent's predefined threshold for Joy.

Motivational Component - component that models basic human drives, such as energy and integrity and uses them to help select between competing goals in the deliberative component. The more a certain need is low/high, the higher/lower the utility of a goal that contributes positively for that need. When an event lowers/raises the agent's needs, it is evaluated as desirable/undesirable for that agent.

Theory of Mind Component - creates a model of the internal states of other agents. This component determines the desirability of an event for others by simulating their own appraisal processes.

Understanding FAtiMA is an important aspect to take into consideration since our own solution revolves around utilizing its agent architecture to modify our agents behaviour, as can be seen in Chapter 3.

2.1.3 Tardis

The TARDIS project [13] is a scenario-based serious game, that consists on the simulation of a platform for social training and coaching, in the context of job interviews. The platform has as its target audience young people aged 18-25, and job-inclusion associations. Users can practice and improve their social skills in a variety of possible job interview situations, and interact with virtual agents, which are designed to act as recruiters and deliver realistic socio-emotional interaction. The user will play himself applying for a fictional job and is expected to behave as he would in real life. The input device used is also the player himself. TARDIS can detect its users' emotions and social attitudes in real-time, through body language, voice, and facial expression recognition. The game's progress and the virtual agent's behaviour is then adapted based on that information and the agent's goals (making the applicant at ease, or trying to put him under pressure). The TARDIS architecture for social coaching involves three actors: the user (the youngster), the interlocutor (which is replaced by an intelligent virtual agent) and the tutor. It includes four main components: Scenario, Social Signal, Affective, and Animation modules, described below:

- The Scenario module controls the dialog during the interview. It tells the virtual agent the expectation in terms of emotions and attitudes expressed by the user.
- The Social Signal modules provide the affective model with information about the user's emotions and social attitudes (mental states) that are detected by the system.
- The Affective Module is responsible for building a model of beliefs and intentions for the virtual agent, about the mental states of the user and about the course of actions.
- The Animation module is responsible for expressing the virtual recruiter's affective state, through verbal and non-verbal behaviour (facial animation and body movements).

For the purpose of this project, and considering our interest in the affective nature of AI agents, we are mainly interested in the Affective Module.

The Affective Module for the virtual agent has two main computational functions:

The first one, the Affective model, periodically computes the new affective states of the agent, and provides a reactive model based on the perceptions from the user, scenario expectations, and its current affective states. It allows a computation of the agent emotions, moods and social attitudes.

The second one, the Decision module, will select actions (i.e. different forms of acting and specific branching in the scenario) according to the Agent's intentions, following the user's mental state, with the goal of building a representation of the user's beliefs. The virtual agent considers the user's answers in a particular context to derive the positive or negative attitude the user has to the considered topic. For instance, if the user re-acts with high detachment to a question about the technical requirements for the job, then the agent might deduce that the user is not very confident on the topic "skill", and will lower the value of $B_{User(skill)}$. The model will, in turn, influence the next actions of the virtual agents in function of the intentions. The agent will select in the scenario subjects where the youngster is at ease (helpful/friendly) or not (confrontational/provocative). Therefore, the computational model of agent social attitudes ultimately relies on its personality and its current mood. Figure 2.2 illustrates the different elements of the Affective module as well as the links to the other modules.

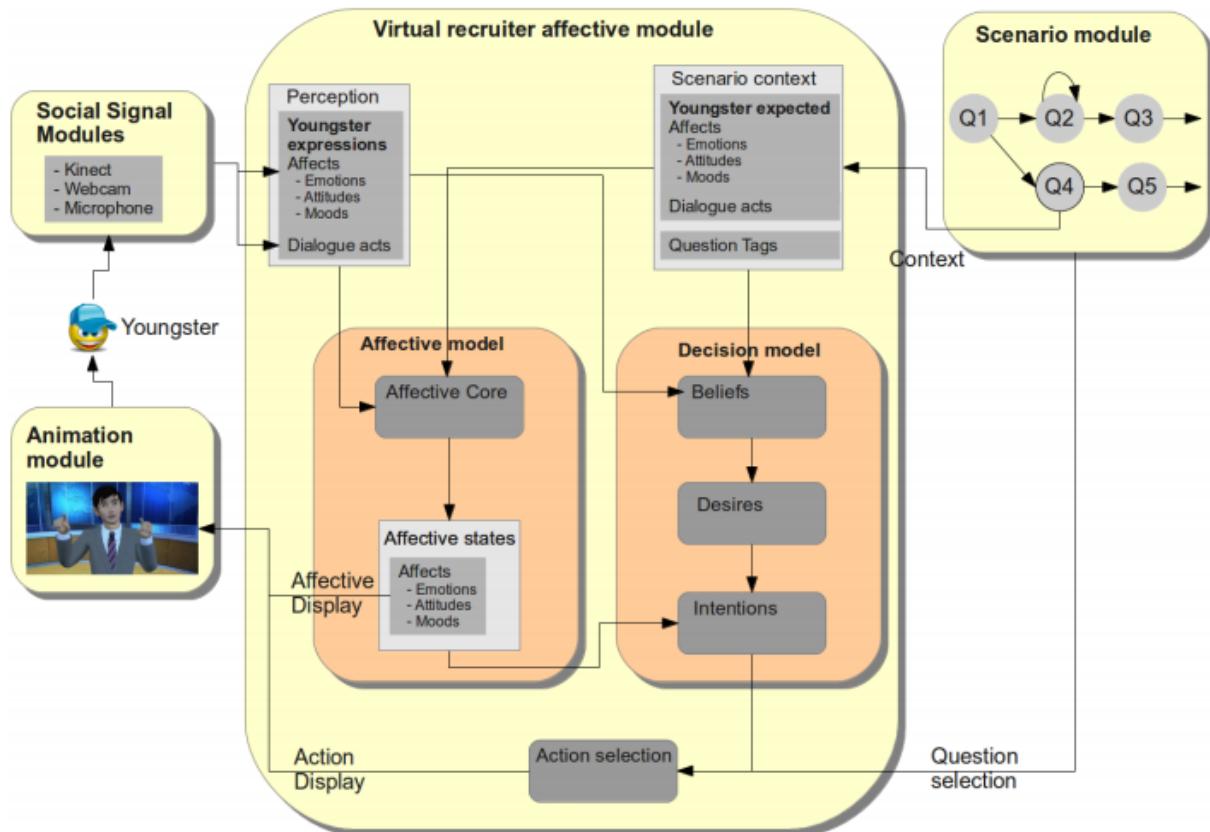


Figure 2.2: Tardis Affective Model Architecture

The TARDIS project is a great example of how proper computation of affective states, and subsequent

expression of said state, is of huge importance towards creating a believable AI agent and a more immersive user experience. The same logic is applied to our project: the constant input of the user has an effect on what is perceived by our agents, consequently changing their behaviour, although in our case said behaviour is expressed through text and appropriate context.

2.1.4 Comme Il Faut

Comme il Faut (CiF) [14] is a model of social state, a collection of processes which can reason over that social state, and a framework for defining actions which can alter the social state and ways for those actions to be performed. It is a tool for social reasoning, intended to be used as a component for games which wish to show social dynamics.

We will not analyse the entire CiF architecture, and will instead look at specific parts of it that are important and relatable towards our project, namely the definition of characters, and managing of social state, social exchanges, and rules.

Characters are defined through four primary sets of characteristics, which are part of the social states: traits and statuses, relationships and social networks, cultural knowledge base, and rules.

Traits and statuses are permanent or temporary binary properties, respectively, of a character, which impact how that character performs in the social space. For example, a character might always have traits like 'pretty' or 'smart', but have statuses, such as 'sad', for a short period of time. Traits are immutable, while statuses expire when the conditions that triggered them no longer hold, and can be directional, so a character might temporarily be, for example, 'angry' at another character.

Relationships are binary states that specify a significant social connection between a pair of characters. For example, two characters might have the relationship of friends or colleagues. Relationships are non-exclusive (a character can have multiple friends), non-restrictive (a character can have a friend who is also a colleagues) and work in conjunction with bidirectional social networks, showed by scalar values. This means one character may have an admiration network value of 20 towards a friend, who reciprocates with an admiration value of 60.

The Cultural Knowledge Base (CKB) is a way to define the world that characters inhabit, by providing them with a variety of concepts and objects from the story world's context. It can be specified both items and the ways characters can relate to them, such as desire or ownership. For example, the CKB for a CiF game might include "Fridays" and "work." Upon being defined the connection types "loves" and "hates", a starting state could be set in which a character "hates work" and "loves Fridays". Also, each item can be linked to a single adjective which defines an universal opinion. For instance, the author might say that Fridays are "awesome," a cultural construct that all characters agree exists, even if their personal opinion differs. A character might still "hate Fridays" in spite of the perception that they are awesome. This enables characters to operate within a cultural context.

CiF uses the above representations of characters and social state to determine how characters should interact with one another. A social exchange is an attempt by one character to change the social state between him and another character. For example, an unpopular character (the initiator) might try to start a friendship with a popular character (the responder). This desire, or volition, of the initiator results from CiF's evaluation of the current social state. The responder will choose to either accept or reject the proposed social exchange based on his own relation to the social state in a process of response determination. For example, some factors that might have influenced the unpopular character to seek this friendship could be a mutual interest for movies. Every social exchange authored for a CiF story world has a single primary intent, or intended change to the social state. Multiple social exchanges for the same intent define distinct ways of achieving the same social outcome. The intents, and thus the social exchanges, a character wants to pursue are recalculated after every social exchange in a process called "volition formation."

For each pair of characters, volition formation ranks all possible intents and exchanges based on a hand authored set of social influence rules. Each rule has a weight value which adjusts volition for either a specific social exchange or for an intent (a set of social exchanges), either positively or negatively. For example, one rule might give a higher weighting to accepting "being friends" if the responder has a high respect network value toward the initiator. Another rule might have a strong negative weight for the same intent if one of the two characters is friends with someone the other character hates.

CiF is an important reference for our work since it is utilized in the serious game PromWeek, which was also a reference. Furthermore, the concepts of character definitions, rules, desires, social exchanges, bidirectional scalar values to describe relationships, and the cultural knowledge base, are all related with our proposed solution.

2.2 Serious games

Serious games are games focused on a purpose, like for example, promoting behaviour change. The actual gameplay and fun aspects are directed towards that goal. There have been serious games created in the past related to the bullying thematic or managing of emotions, and some of which are discussed and analysed below.

2.2.1 FearNot!

FearNot! (Fun with Empathic Agents Reaching Novel Outcomes in Teaching) [5] [15] is a game for anti-bullying education. The serious game was created in order to fight bullying in schools and help educate children from 8 to 11 years of age. It allows children to explore what happens in bullying in a nonthreatening environment, in which they take responsibility for what happens to a victim, without

themselves feeling victimized, which is similar to our game's method. The game was developed as a virtual learning environment, and the action is placed in a virtual school, with characters representing the most significant roles in bullying: bullies, victims and bystanders. The environment is populated by intelligent virtual agents, characters that are autonomous agents capable of making their own decisions and acting out their own behaviours, due to a BDI architecture, presented as a mechanism for decision making. This leads to an unscripted narrative as the story line progresses, The session starts with an introduction to the school and the characters, and then a dramatic episode follows in which a bullying incident occurs (see Figure 2.3).



Figure 2.3: Bullying scenario in FearNot!

The victim then asks the user for advice in dealing with the situation, and the user suggests a coping behaviour in order to advise the victim on how to deal with the situation in the following episode. Victims are aware of themselves and decide if the advice given is to be followed or not. In this way, the user takes on the role of an invisible 'peer buddy', or friend, to the victim character, in different scenarios where the user is merely a spectator. This system allows children to try out various coping strategies without being directly involved themselves. The usefulness of a coping strategy can be learned safely through the victim character's experiences. It is important to provide youngsters with a safe-place that they can explore without the consequences of real life, so decisions only affect the characters and never the kids playing.

FearNot! is a serious game that was taken in consideration when developing our solution, since the way it works has similarities with our game. The game's context is related to bullying situations with school students, Agent's actions (in our case, posts and comments) are the main feedback mechanism,



Figure 2.4: Interaction with an Agent in FearNot!

and player input through dialog (in our case, comment options) is the main form of interactivity. It also provides different scenarios like Com@Viver. The most relevant feature is that the game has an adaptive narrative where the help provided by the player can make a difference, which is what we strove to achieve with the changes we made to Com@Viver.

2.2.2 Façade

Façade [16] [17] is a game that uses AI and natural language processing to simulate a married couple having a domestic dispute. The player role-plays as a close friend that arrived early to a meeting and accidentally interrupted the couple, Trip and Grace, while they were having a fight. The story can be unravelled through asking questions or answering the two characters.

The AI agents use vocal tones and facial expressions to indicate how well the conversation and player progression is going. The game can end abruptly if, by being offensive, the player gets an angry reaction from one of the characters, making them ask the player to leave the apartment and ending the



Figure 2.5: User asking a question to Grace in Façade

game.

The game was originally designed with the objective of engineering an AI system that responds to and integrates the player's moment-by-moment interactions, and providing the player with a high degree of responsiveness and narrative agency. However, the game can also lead to the creation of an empathic relationship between player and character. For instance, there is a point in the story that can generate a dilemma, where the player is faced in a situation where they must either talk to Grace to enquire about her personal well-being or to ignore her and ask Trip about his new cocktail kit. To progress and unfold Grace's story the player must ask her about her feelings, however, to unfold Trip's story the player must show interest in him. Both characters pull the player towards their conversation, and it's up to the player to make a choice about who to give attention to. The players can get caught between the two characters situations and can sometimes find themselves making decisions through concern, due to the characters reactions.

Regarding our solution, this phenomenon of creating a connection with the player through changing dialogue is relevant for our study. One of our goals was to motivate players to change their behaviour, making it gravitate towards pro-social decisions. We tried to achieve this by using AI Agents that change their dialogues according to player actions, similar to Façade. By stimulating a sense of empathy towards cyberbullying victims, we hoped players would feel motivated to stand up for the victims in the story.

2.2.3 PromWeek

Inspired by classic high school movies from the past few decades, Prom Week is a social simulation game where the player shapes the social lives of a group of high school students, in the week previous to their prom. It uses the artificial intelligence system CiF, and tries to combine the dynamic simulation of games like "the Sims" with the detailed characters and dialog of story driven games.



Figure 2.6: Prom Week's user interface.

The Gameplay in Prom Week revolves around the social lives of eighteen characters from the same school, and the game's goal is to create complex narratives that are unique for each player and adjusted to their actions and objectives. The characters keep track of every player action and respond dynamically.

The game features a story mode. In any given "story" or "campaign", the player is given a set of goals to potentially complete during the week leading up to the prom. For example, in Zack's story, goals can be getting him a prom date, or ending his war against a bully (a theme shared with our game). The game also features a free play mode, where the player can interact and manipulate characters at will. Goals can be satisfied through an open-ended set of solutions discovered through interaction

with the characters and their social state. For example, the player could have Zack form a friendship with a popular character over a shared interest. The player works this goals by choosing the top social exchanges that each character desires to play with each other character. This ordered list is provided by CiF, based on the outcome of the volition formation process.

Because the gameplay of Prom Week involves manipulating the "social space", which is the primary story content of the high school narrative, the gameplay ends up being the story itself. Every action the player takes advances the game's narrative and changes the internal social state, which in turn changes which actions are available in subsequent turns.

CiF gives the narrative meaning and shape, contrasting sandbox games in which gameplay may be the story, but the story is formed only in the mind of the player, and not understood or reasoned over by the system. While the game designers create the starting point and define the goals for each scenario, it is CiF that enables different emerging solutions to each social puzzle, making the story space highly dynamic and responsive to player action.



Figure 2.7: Prom Week's social exchange.

Every story's last level takes place at the prom. After the player runs out of turns, or decides to skip to the end of the night, a customized ending is presented that reflects the combination of goals achieved. Every story has many possible endings for various combinations of goals the player might have completed. For example, Zack's story might end with him becoming the prom king if the player was able to get him to date a popular person. As the player finds more endings, additional stories are unlocked.

Story goals are a good example of external rules. They are Prom Week specific, but reason over elements of CiF's social state. After every social exchange, Prom Week uses CiF's rule evaluation system to check to see if any of the story goals have become true (or, if any that used to be true have become false). If there has been a change, then Prom Week lets the player know that story goal progress has been made or lost.

The interest of this work in our study lies in the way interactions work, and the story behind it. The story's characters are eighteen students from the same school, similar to our own game that features 12 characters from the same classroom. The story is based on their social lives, and some aspects of the story change based on the players' interactions with the characters. The story's changes based in social states and exchanges, as well as the CiF rules applied to interactions between characters, are also similar to our solution (explained in chapter 3).

3

Solution

Contents

3.1	The Com@Viver game, agents and scoring system	23
3.2	Social Importance Dynamics Model	26
3.3	Implementation	28
3.4	Procedure and example	32
3.5	Game Session Schematic	36

3.1 The Com@Viver game, agents and scoring system

Before analysing the details of our solution, it is important to understand the setting in which the game is placed, as well as the current player dynamic and score attribution.

3.1.1 Game setting and interface

As briefly explained previously, Com@Viver is serious game designed as a representation of a SNS. It is populated by 12 Autonomous Agents (AA) that represent school students, and are able to comment on posts made in the Social Network, according to their roles. The AA fulfil the roles of aggressors, victims and bystanders. The players always have the role of bystanders. It is shown in fig. A.1 the composition of the virtual school class, as well as the relationships between the characters. In addition, fig. A.2 describes the characters' roles in the currently existing cyberbullying cases, and appendix A contains the AI Agents' backstory and profiles.

The game is played in groups of 3, where each group plays in a separate game instance. This means that players from different groups never interact with each other, and are limited to their own virtual classroom. Players interact with the game through two main components: the Feed, and the Chat (fig. 3.1).

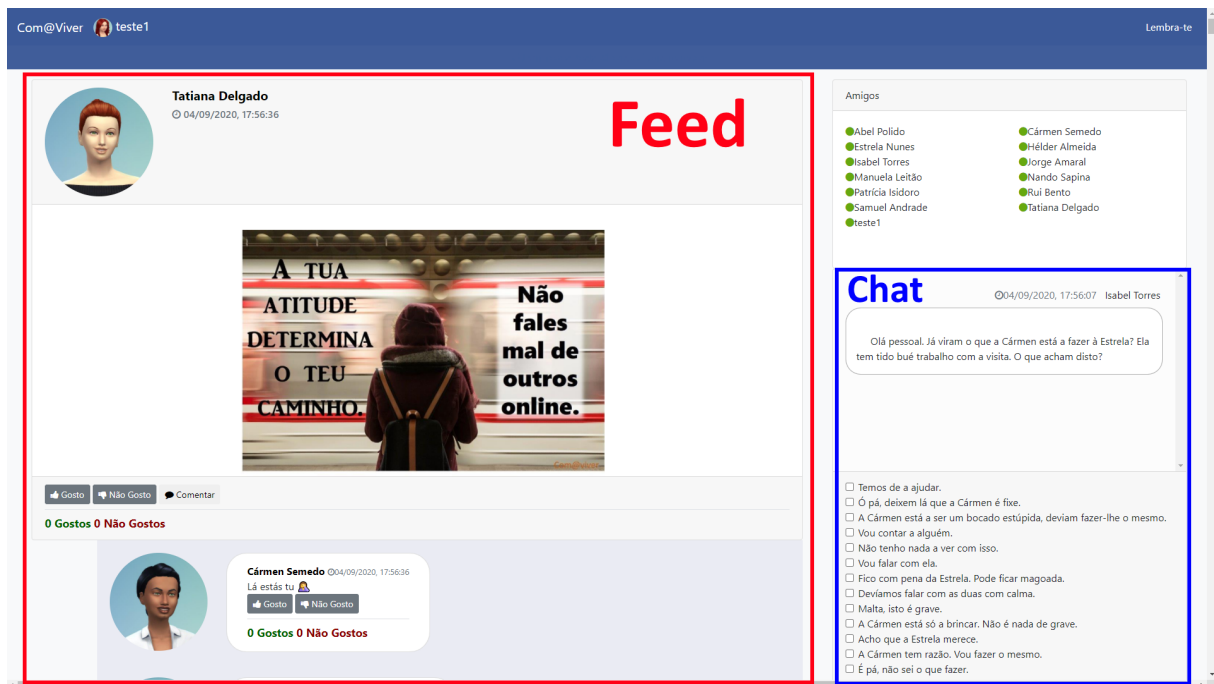


Figure 3.1: Com@Viver game interface.

The feed works as the main component, since it is where all the posts and comments appear. It is shared amongst all the players from the group, meaning the players can not only comment, like or dislike

any post and comment made by the AA, but they can also see what the other members of their group have chosen as comment options, and can also like or dislike comments from their own group (fig. 3.2). Posts and comments automatically appear from time to time, throughout the session.

The chat is a private component, where players answer questions about how they feel towards the cyberbullying case, and their answers are not shared with the rest of the group.

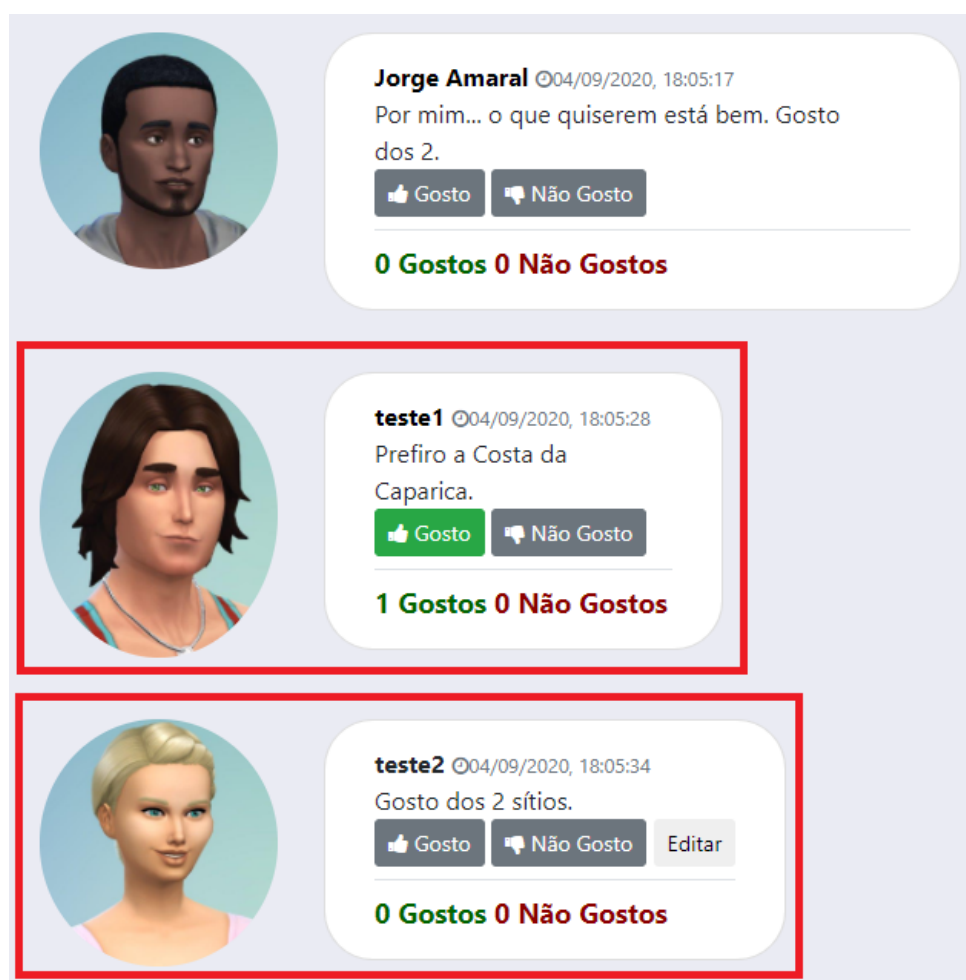


Figure 3.2: Comments from 2 different players from the same group.

3.1.2 Game goal and scoring system

The game's goal consists of being able to go on a school trip, with a limited number of students allowed. The 3 players form a group, and the 12 AA form the remaining 4 groups. Only the three most voted groups can go on the trip, meaning that if the players stay in the last 2 spots of the voting pool, they will not go on the trip, losing the game. At the end of each session there is a voting poll where players can vote on the group they think should go on the trip, followed by a set of questions, and an endgame score screen that simulates the groups position in the voting pool, to demonstrate the final score.

The final score of the game depends on the individual scores of each player from the group. The only actions that have an effect in a player's score are liking, disliking and commenting the cyberbullying posts. Liking or disliking comments, even comments from the cyberbullying post, will not change the score. Acting upon the post can have a positive or negative impact on the score depending on the action chosen. Disliking the post, or choosing a positive comment option, will be considered a positive behaviour, while liking the post or choosing a negative comment option will be considered a negative behaviour. The individual player score is described as follows:

1. Engaging in more than one negative behaviour actions will grant a score of 0,
2. Engaging in only one negative behaviour action will grant a score of 3,
3. Doing nothing (engaging in the bystander effect) will grant a score of 6,
4. Engaging in one or more positive behaviour actions will grant a score of 12,
5. Negative actions override positive actions. For example, one negative action and one positive action equals one negative action, resulting in a score of 3.

The final score is calculated as the group's score average. In order for the group of players to win, they must have a group average score of eight or higher. Figures fig. 3.3 and fig. 3.4 show the endgame screen for winning and losing groups, respectively.

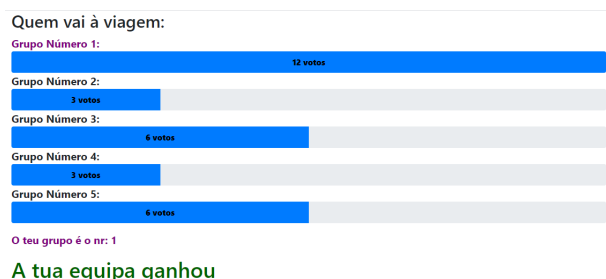


Figure 3.3: Winners endgame screen.

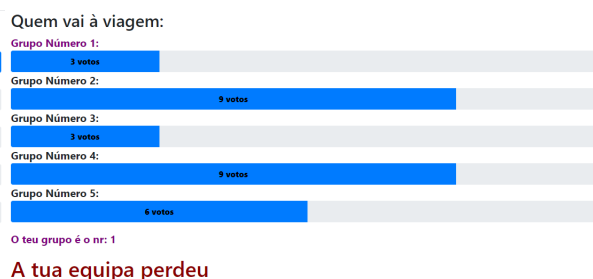


Figure 3.4: Losers endgame screen.

3.2 Social Importance Dynamics Model

Our proposed solution was to implement into our project a Social Importance Dynamics (SID) model developed by Mascarenhas and colleagues, in their work “Modeling culture in intelligent virtual agents” [18].

3.2.1 Computational Algorithm Explanation

The SID model is based on Kemper’s theory [12] that argues that all human social activity is motivated by two constructs: status and power. Both concepts are represented as a relational scalar between two social entities, and the amount is not necessarily reciprocal [12]. This notion of status is operationalized by making each agent attribute to other agents a value of Social Importance (SI). The SID model is a model for rational agents that has the following central concepts:

1. Beliefs - the knowledge the agent has of the environment, of itself or of other agents. Beliefs are represented as logical predicates, i.e “isFriend(Nando)”.
2. Desires - agent’s motivations, they represent states of the world the agent wants to achieve. Represented by G, it denotes the set of abstract goals the agent has. Such goals have the following properties:
 - i Name - a unique identifier;
 - ii Preconditions - a list of logical conditions that need to be grounded and verified in order for the goal to be instantiated;
 - iii SuccessConditions - a list of logical conditions that dictate when the goal is considered to have been achieved;
 - iv ImportanceOfSuccess - a numeric value that specifies how important the goal is for the agent.
3. Intentions - the states of the world the agent is committed to achieve at a given moment.

Further explaining the model’s language:, B_{self} represents the set of the agent’s beliefs B_x contains the agent’s beliefs about the beliefs of agent ‘x’. I denotes the set of intentions the agent currently has. Each intention in I is composed of an instantiated goal. As seen previously, the intention has an associated plan of actions to achieve the goal’s SuccessConditions, and each of these actions has a list of preconditions that are required to be verified in order for the action to be performed, as well as list of effects that are expected to become true if the action is performed successfully. This model, therefore, is known as a Beliefs, Desires, Intentions (BDI) model, and has 3 main cycles: Belief revision, Option generation, and Action selection.

In belief revision, agents start by "feeling" the environment in which they find themselves, and search for an event. Upon the occurrence of said event, their beliefs are revised, including the beliefs about SI relationships. The values of the SI relationships the agent has with others are stored in B_{self} in the form of $SI([x]) = [y]$, where $[x]$ is the name of the target agent, and $[y]$ is how much SI is attributed to that agent. This value is determined through SI Attribution rules, that are defined as a tuple $\langle T, A, V \rangle$ where:

1. T specifies the target of the rule.
2. A corresponds to a list of conditions that specify when the rule is activated.
3. V the amount of SI the target of the rule gains or loses

When the agent perceives a new agent in the simulation, it will create an initial belief of how much SI is attributed to said agent, and that value is determined by the sum of all SI Attribution rules that are activated when applied to that agent.

After the belief revision, a list of options is generated, composed of the abstract goals in G that can be instantiated according to the revised beliefs. A goal is considered active if a valid substitution is found for all the variables defined in the Preconditions and SuccessConditions of the goal. These generated options (active goals) represent the agents' desires, and are filtered regarding their expected utility. The utility of an option corresponds to the ImportanceOfSuccess of the associated goal, since that in this model, the agents have a general desire to perform acts that signify the amount of SI they give to other agents. These acts are called "Conferrals", and are defined as a tuple $\langle C, A, T, V \rangle$ where:

1. C is a set of preconditions that dictate the context in which the conferral is appropriate.
2. A is the name of the action that represents an SI conferral.
3. T corresponds to the target agent to which the conferral applies.
4. V specifies the amount of social importance conferred by the action.

So, for the purpose of option generation, an SI conferral is converted to an abstract goal G , where:

1. Preconditions of $G = C \cup \{ SI([T]) \geq [V] \}$
2. Success Conditions of $G = \text{perform } A$
3. Importance of Success of $G = V$

If the agent is not committed to an existing intention, the option with the most utility is chosen to become the current intention of the agent, and a plan is built, which is a list of actions, that must be done to achieve the successConditions of the intention.

After committing to the intention with the highest utility, the agent must build or choose a plan of actions to achieve the intention and then, from this plan, select the next action to perform. As we saw previously, options are filtered based on their expected utility, but cultural conventions establish what seems appropriate or inappropriate between people. To illustrate such conventions, SI Claims come into play, to mediate what is and isn't socially or culturally appropriate. SI Claims are defined as a tuple $\langle A, T, V \rangle$ where:

1. A is the name of the action that is perceived as a claim for social importance.
2. T is the target of the claim.
3. V is the amount of social importance the action is claiming.

Before a plan is set in motion, the agent determines whether any action has an associated SI Claim, and if such action is found, the agent checks the mental model of the claim's target T to infer whether the SI that T attributes to the agent is greater than or equal to the claims associated V. If it is lower, then the agent determines that it lacks enough SI to perform the action and an alternative plan must be developed. If no alternative is possible, then the agent considers that the intention is impossible to achieve in a socially appropriate way and drops it, searching for an alternative course of action. The implementation of this model, as well as an example of functionality, are explained in the following sections.

3.3 Implementation

In terms of structure, we used two separate servers, one running the Com@Viver platform while another ran the FATiMA system. We used Azure Cloud Services for the FATiMA server, and DigitalOcean for the Com@Viver server, and the servers communicated with each other sequentially during the sessions. This behaviour is explained in the last section of this chapter.

The implementation, in Com@Viver, of the previously described model and its rules, was made possible through the use of the FATiMA toolkit, a collection of tools and assets designed for the creation of characters and agents with social and emotional intelligence. By using the Integrated Authoring Tool (IAT) (fig. 3.5) we were able to create game scenarios, using the built in Dialogue Editor, Role Play Character Editor, World Model and Web API. The IAT uses its own file system, so a new "Scenario file" was created for each session.

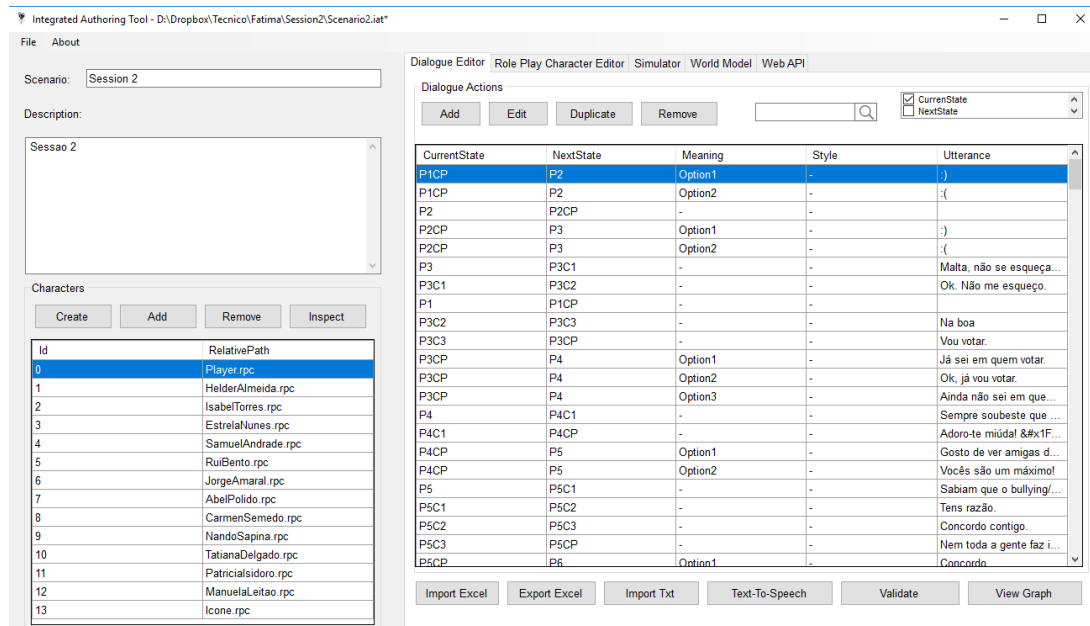


Figure 3.5: FAtiMA Authoring Tool

3.3.1 Dialogue Editor

The original game dialog was extracted to Excel files so it could be easily edited and imported into the authoring tool's dialogue editor. Each post content and comments were divided into multiple dialogue states (fig. 3.6) to allow for AI Agents to iterate through them.

The dialogues are defined by 5 parameters: The current state and next state define the order of dialogues, and followed a nomenclature logic. The first dialogue of a post was coded as "P + PostNumber", the dialogues of comments were coded as "P + PostNumber + C + CommentNumber", and the player options were coded as "P + PostNumber + CP"; The meaning field was used to distinguish between the different options the players had to answer posts, and the style field was used to distinguish between normal dialogues and positive dialogues. Finally, the utterance field contains the text that is written by the characters.

3.3.2 World Model and Role Play Character Editor

The world model allowed us to define what actions are available to characters, and the effects of those actions. As illustrated in figure fig. 3.7, the only action available to characters is too speak. This action has the effect of changing the property "DialogueState" of the target [t], to the dialogue's next state [ns]. This effect allows the characters to keep updating the dialogue state of each othe and keep the dialogue flow.

The role play character editor allowed us to create the game's characters inside the FAtiMA system,

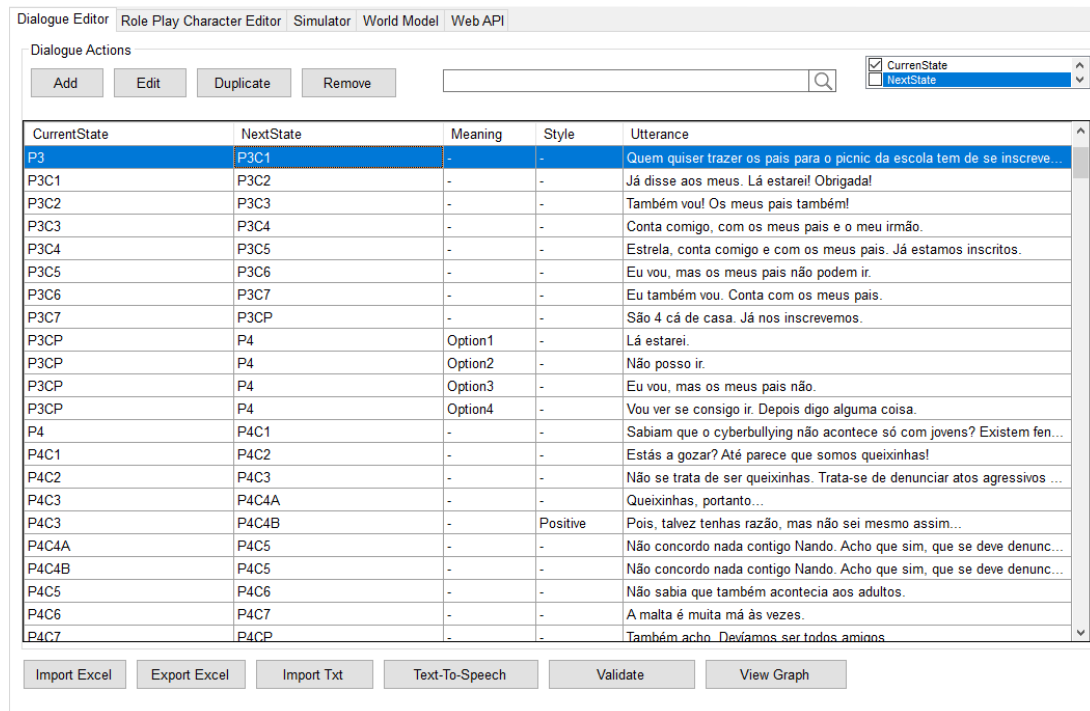


Figure 3.6: FAtiMA Authoring Tool - Dialogue Editor

and give them decision making and social importance rules to follow. The action rules are specified in the Decision-Making tab. Each rule had specified an action that it applies too, a target, a priority, and a set of preconditions that must be verified in order for the action to be selected. For example, in figure fig. 3.9 we can see that, in order for a character to decide for the action with the highest priority, the character must have a valid dialogue state, the target's dialogue state must be the current dialogue state, the selected line must be a positive line, and the SI value towards the player must be 10 or higher. The 2 rules defined allow the characters to choose between acting normally, or using the alternative positive lines we gave them.

The SI Attribution rules are specified in the social importance tab. Each rule has a name, a target to which the SI value is conferred, the SI Value conferred, and the preconditions that dictate when the rule is applied. This is observable in fig. 3.10.

Essentially, the characters from the game were created in the scenario file using the role play character editor, given proper decision making and social importance rules to follow, and their possible actions were specified in the world model.

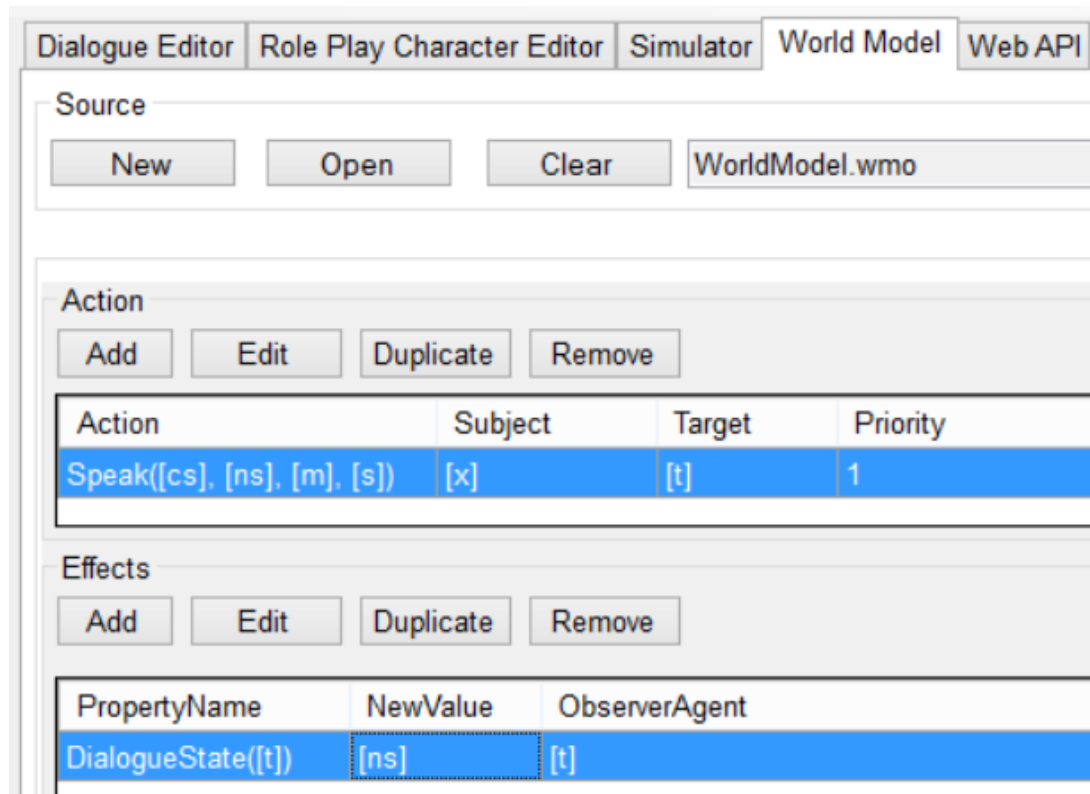


Figure 3.7: FAtiMA Authoring Tool - World Model

3.3.3 Web API

The Web Application Programming Interface (API) was the important piece that allowed the Com@Viver game to communicate with the FAtiMA system. The API consists of a series of HTTP Endpoints that allowed us to send HTTP requests and information between the game server and the FAtiMA server.

The used endpoints where:

1. Create, used to create a new scenario instance in the server,
2. Perceive, used to change characters beliefs and dialogue states,
3. Decide, used to make characters decide their next action, or in our case, their next dialogue,
4. Execute, used to make characters execute their action, in our case, write.

All the HTTP request URLs must follow a specific format, as follows:

- *FatimaServerIP:ServerPort/GroupInstanceIdentifier/Method*, with the required parameters.

In example: in order to allow the agent Nando Sapina to perceive an event, if the FAtiMA server is running locally on port 8080, and the group was playing in game instance 45, it would be necessary to send the FAtiMA server a HTTP request, by using the url "localhost:8080/45/perceive?c=NandoSapina".

Action Rules

Add Edit Duplicate

Action	Target	Priority
Speak([cs], [ns], [m], [s])	[t]	0
Speak([cs], [ns], [m], [s])	[t]	5

Conditions

Add Remove

Condition
- = [s]
ValidDialogue([cs], [ns], [m], [s]) = True
DialogueState([t]) = [cs]

Figure 3.8: FAtiMA Authoring Tool - Role Play Character Editor, Decision Making Action 1.

Action Rules

Add Edit Duplicate Re

Action	Target	Priority
Speak([cs], [ns], [m], [s])	[t]	0
Speak([cs], [ns], [m], [s])	[t]	5

Conditions

Add Remove

Condition
Sl(Player) >= 10
Positive = [s]
ValidDialogue([cs], [ns], [m], [s]) = True
DialogueState([t]) = [cs]

Figure 3.9: FAtiMA Authoring Tool - Role Play Character Editor, Decision Making Action 2.

3.4 Procedure and example

As explained previously, by playing the game, and through its choices, the players will inevitably end the game session with a positive or negative score, from which we can determine if the player defended the victim or not. Then, that information can be used by the characters to determine if they should improve their behaviour in the next session, or if they should behave the same.

Demonstrating how this implementation can be impactful, let us assume a group of players is about to play the third game session, with a FAtiMA server running locally. In Figure 3.12 we can see 2 alternative lines for the same dialog point in the story, created in the dialog editor. This choice of sentence is made by the character Nando Sapina during the third game session. Nando Sapina is the aggressor in the first cyberbullying case, but if the players demonstrate prosocial behaviour in the first session, Nando's behaviour in the third cyberbullying case will change. The order in which the characters speak is defined inside in the game.

3.4.1 Belief Revision

In fig. 3.13 we can observe the FAtiMA server console, where it is illustrated the beginning of the third session. At the start, one HTTP request is sent to the server, using the API method "Create", and with the GroupInstanceIdentifier associated with the group's game instance. Then, the AI agents Nando Sapina and Manuela Leitão, the aggressors of session 1, have their beliefs changed, to allow them

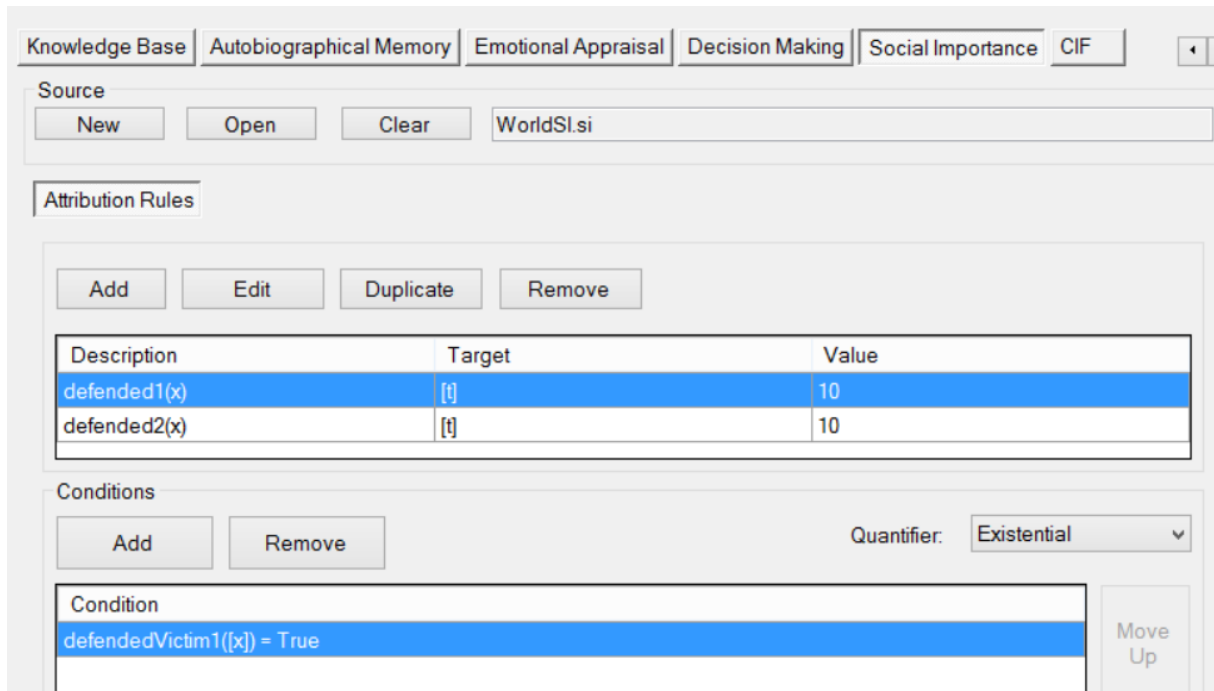


Figure 3.10: FATiMA Authoring Tool - Role Play Character Editor, SI Attribution rules.

to know if the players defended the victim in the first session. This is done through the usage of the API method "Perceive", and 2 HTTP requests are sent to the server. The requests follow the format "localhost:8080/34/perceive?c=NandoSapina", with the content "Event(Property-Change, NandoSapina, defendedVictim1(Player), True)". A similar request is sent to the character Manuela Leitão.

So, in the current session, Agent Nando Sapina is initialized with the following belief:

– $B_{self} = \{defendedVictim1(Player)\}$

After this belief is perceived, Nando will create an initial belief of how much SI is attributed to the player, and that value is determined by the sum of all SI Attribution rules that are activated when applied to the player. These rules are illustrated in fig. 3.10.

As illustrated, the following SI Attribution Rules are associated with agent Nando:

1. defended1: $\langle T = x, A = defendedVictim1(x), V = 10 \rangle$
2. defended2: $\langle T = x, A = defendedVictim2(x), V = 10 \rangle$

Agent Nando will determine the SI attributed to the user by checking if the rules above can be verified when substituting the variable 'x' with 'Player'. Given the initial beliefs perceived, Agent Nando will add $SI(Player) = 10$ to B_{self} by checking that 'defended1' is verified when the substitution 'x/Player' is made.

Dialogue Editor	Role Play Character Editor	Simulator	World Model	Web API
-----------------	----------------------------	-----------	-------------	---------

API Methods			
Name	Type	Params	Description
decide	GET	?c='char'	Returns the actions decided by the character to execute.
characters	GET		Returns all loaded characters and their emotional state.
ask	GET	?c='char...'	Returns the value of a belief for a given character
perceive	POST		Makes the characters perceive all the event strings passed in the body.
update	POST		Updates the characters' internal state for 'x' amount of ticks.
execute	POST		Executes a given action and trigger all the effects defined in the World Model.
reset	POST		Reloads the scenario and all the characters.

Figure 3.11: FAtiMA Authoring Tool - Web API

3.4.2 Selecting the action to perform

After the belief revision, whenever a character is supposed to act according to the flow of the game session, by using the API method "Decide", the characters will choose their next course of action. Let us assume the current dialog state is "CASEC4", which corresponds to the forth comment in the cyberbullying case post. It is Nando's turn to speak. By sending to the FAtiMA server the HTTP Request "localhost:8080/34/decide?c=NandoSapina", the FAtiMA system will return a list of options, that includes all actions available to Nando in that moment. By analysing fig. 3.7, we know the action will be "Speak", as it is the only action that exists. However, fig. 3.8 and fig. 3.9 show us that there are 2 variants of the speak action, with different priorities. Agent Nando will evaluate if the conditions of both paths are valid, and order them by the highest priority. Since there are valid dialogue states, the SI value towards the player is 10, and there is a dialogue with the style "Positive", all conditions are met, and Nando will pick the positive line as the priority of his next action.

3.4.3 Executing the dialog

In fig. A.3 and fig. A.4 we can see the final result of this process, with a visual difference in how the dialog is presented inside the game to winning and losing groups. The difference between this behaviour allow us to model a more human-like behaviour in our agents, that make intelligent decisions based on the user's actions.

After choosing the action to be executed, API method "Execute" allowed us to order the AI agent to execute an action and apply its effects. According to fig. 3.7, the action of speaking will have an effect on the selected target of the action (the agent being spoken to). The effect will make so that the dialogue state of the target will become the "Next state" specified in the dialogue editor, giving it access to the next lines in the story. In this case, the next character to speak will be Isabel Torres.

CASE	CASEC1	-	-	Pensava que eras panelheiro Samuel. Afinal apaixonaste-te pelo maior des...
CASEC1	CASEC2	-	-	Estás a ser um bocado parva Patrícia.
CASEC2	CASEC3	-	-	A patrícia está só a gozar. Já não há sentido de humor.
CASEC3	CASEC4	-	-	Não te metas Rui!
CASEC4	CASEC5	-	-	Ya, não te metas Rui.
CASEC4	CASEC5	-	Positive	O Rui só quer chamar-te à razão.
CASEC5	CASEC6	-	-	Não chateiem o Samuel. Que mal é que vos fez?
CASEC6	CASEC7	-	-	Isabel. Está calada! Queres levar na boca? Mete-te na tua vida!
CASEC7	CASEC8	-	-	Patrícia, daqui a pouco quem te dá na boca sou eu.
CASEC8	CASECP	-	-	Ó Hélder, ameaças a Patrícia e trato eu de ti!
CASEC8	CASECP	-	Positive	Não ameaces a Patrícia, isso só piora tudo!
CASECP	-	Option1	-	Ó Samuel, tens um gosto da treta!
CASECP	-	Option2	-	Ó Patrícia, estás parva ou quê?
CASECP	-	Option3	-	Acho que o resto da malta devia estar calada!
CASECP	-	Option4	-	O Samuel gosta dela. Temos de respeitar isso.
CASECP	-	Option5	-	O Samuel tem os seus gostos e sente-se bem assim
CASECP	-	Option5	-	Patrícia, percebo que estejas chateada com a Tatiana. Samuel, sei que e...
CASECP	-	Option6	-	O Samuel gosta de quem gosta. E se ele errou nalguma coisa não faz m...

Figure 3.12: FAtiMA Alternative Lines from cyberbullying case 3

```
Administrator: Command Prompt - .\FAtiMAHTTPServer.exe 8080 D:\Dropbox\Tecnico\Fatima\Session3\Scenario3.iat
D:\Dropbox\Tecnico\Fatima\Server>.\FAtiMAHTTPServer.exe 8080 D:\Dropbox\Tecnico\Fatima\Session3\Scenario3.iat
Server started on port '8080' on scenario 'Session 3'
POST request: /34/create [07-9-2020 5:33:33 PM]
Result: "Instance '34' created."
POST request: /34/perceive?c=NandoSapina [07-9-2020 5:33:34 PM]
Result: "1 event(s) perceived by nandosapina"
POST request: /34/perceive?c=ManuelaLeitao [07-9-2020 5:33:34 PM]
Result: "1 event(s) perceived by manuelaleitao"
```

Figure 3.13: FAtiMA Server running locally on port 8080. A game session correspondent to session 3, with the identifier 34, is starting, and the necessary changes are being made to the beliefs of the first session's bullies.

This is achieved by sending to the FAtiMA server the HTTP Request "localhost:8080/34/execute", with the content "(NandoSapina, Speak(CASEC4, CASEC5, -, Positive), IsabelTorres)". Upon receiving this request, the action will be executed. This process is illustrated in fig. 3.14

```
GET request: /34/decide?c=NandoSapina [08-9-2020 12:36:45 AM]
Result: [{"Action": "Speak(CASEC4, CASEC5, -, Positive)", "Target": "NandoSapina", "Utterance": "O Rui só quer chamar-te à razão.", "Utility": 6.0}, {"Action": "Speak(CASEC4, CASEC5, -, -)", "Target": "NandoSapina", "Utterance": "Ya, nao te metas Rui.", "Utility": 1.0}]
POST request: /34/execute [08-9-2020 12:36:45 AM]
Result: "1 actions(s) executed and 1 effects triggered"
GET request: /34/decide?c=IsabelTorres [08-9-2020 12:36:45 AM]
Result: [{"Action": "Speak(CASEC5, CASEC6, -, -)", "Target": "IsabelTorres", "Utterance": "Nao chateiem o Samuel. Que mal é que vos fez?", "Utility": 1.0}]
```

Figure 3.14: The Decide and Execute commands shown on the FAtiMA server console.

3.5 Game Session Schematic

In fig. 3.15 it's illustrated the process of a game session. The 3 players form a group and log in into the game on their web browsers. The login information is sent to the Com@Viver server, it is verified, and a game session is created for the group. The game server verifies the player's data regarding the session number and the previous scores, and sends it to the FAtiMA server. The FAtiMA server receives this information, and the game scenario correspondent to the session number is loaded. The agents SI values towards the player are loaded according to their previous sessions' results. After the necessary data is loaded, the FAtiMA server informs the Com@Viver server that it's ready to start, and the game session starts.

Upon starting, the game session enters the dialog cycle: The game server asks the FAtiMA server for the next dialog chain, and the FAtiMA server calculates the next dialog options based on the SI values and the current dialog state. Then, the FAtiMA server sends the dialog options to the game server, and the game server presents the new dialog to the players. The players then interact with the game by choosing the dialog options available. After this process, the game server evaluates if this was the final dialog. If it was, then the game state proceeds to the final stage. If not, the game repeats the dialog cycle until the final dialog is reached.

The dialog cycle is usually compressed of a first set of about 20 "normal" dialogues, followed by a special dialog that includes a cyberbullying case, and another set of 20 "normal" dialogues. In both sets of 20 dialogues there is conflict between the characters of the story. These conflicts will be adapted in key situations, according to the player's choices on previous sessions. In the cyberbullying case, the designated aggressor of the session posts something offensive towards the victim, and the players can react to it via the comment section and the web chat. Game sessions last around 35 to 40 minutes, with close to 5 minutes extra time (depending on the speed of the player answering the questions at the end).

After the end of the dialog and interaction part, the game presents the players with the end of session questions. The players respond to the questions presented to them and the results are registered to be used later in the data analysis. After answering these questions, the game session ends.

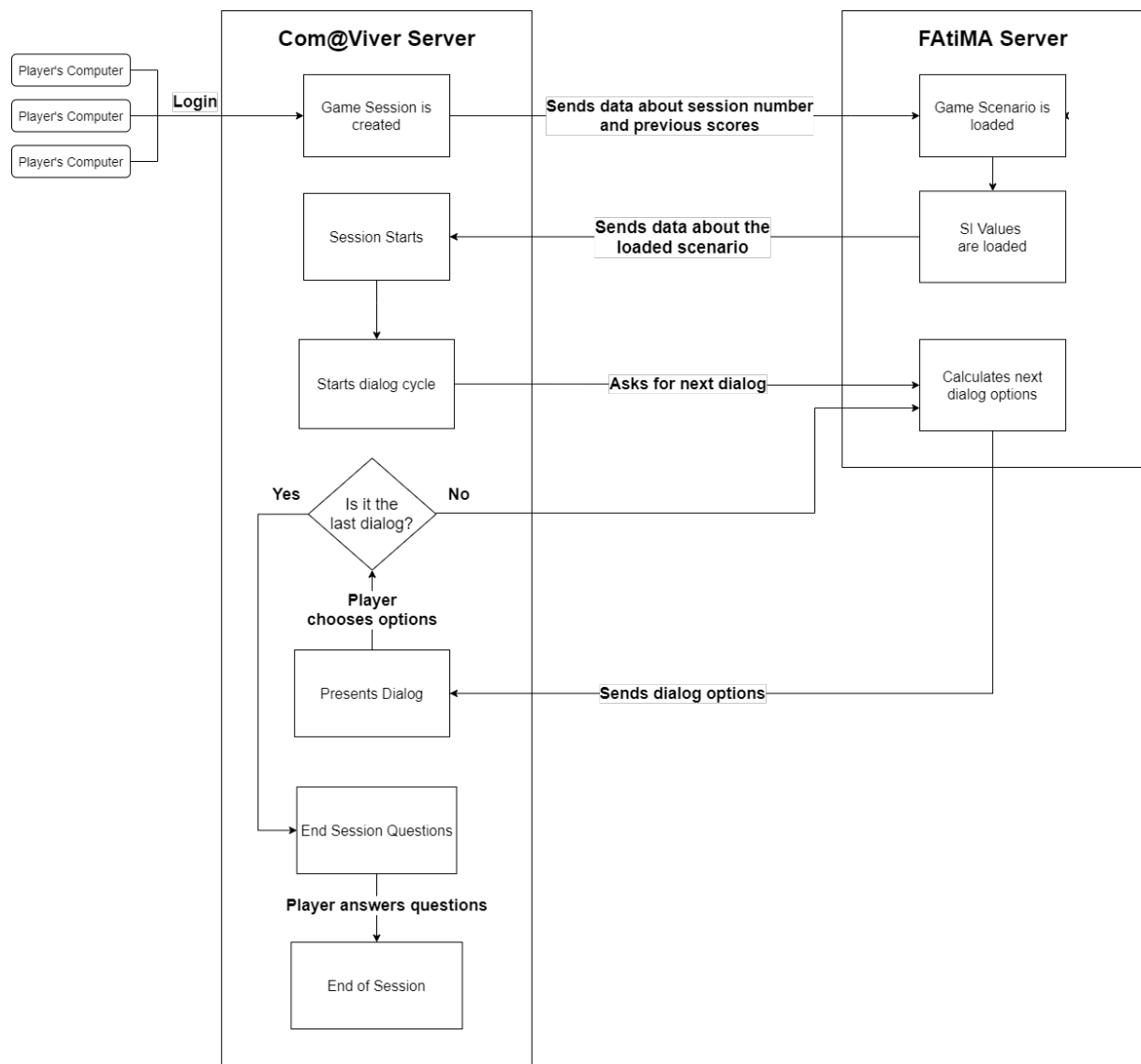


Figure 3.15: Session Scheme

4

User Study

Contents

4.1	Study design	39
4.2	Participants	39
4.3	Instruments	39
4.4	Procedures	41
4.5	Data Analysis	42
4.6	Results and Discussion	43

4.1 Study design

A quasi-experimental control-group design with repeated measurements (with pre and post test) was used with process data gathered during gameplay in a real-life context.

4.2 Participants

The convenience sample used in this study consisted of a total of 297 students (73,7% in 7th grade, 27,9% in 8th grade and 34,3% in 9th grade). In the exploratory analysis of the instrument we had 139 participants, whereas in the confirmatory analysis we had 158. The total sample was used in the main analysis of this study with 168 students in the Experimental Group (EG) and 129 students in the Control Group (CG).

4.3 Instruments

4.3.1 Game Adaptation and gameplay questions

As explained in chapter 3, and illustrated in fig. A.2 and fig. A.3, the game consists in a series of posts and comments made by the AI agents, to which the players respond to. These dialogues had to be adapted in order to create alternative dialogue lines in each session, to be presented to groups that had won or lost the session prior. These alternative lines led characters to change their ways, reflect upon their actions, or actually becoming a pro-social behaviour advocate and trying to shut down bully behaviour from other characters.

In table B.2, table B.3, and table B.4, we can observe the characters' variance of roles in each conflict, and the differences if the players had won or lost in the past. In the figure it is registered the occasions throughout the different sessions in which the AI Agents had alternative roles during a conflict. The agents could have the role of being the victims, the aggressors (through commenting or being the initial creator of the post), having positive dialogues, or being neutral.

With these alternative character lines available, we analysed the relative frequency of positive and negative comment options chosen by players during the experiment, in these specific dialogues, and also in the cyberbullying cases, and compared these frequencies between winning and losing groups. We also analysed the frequency of posts where players did not comment anything, but did at least like or dislike the post. Lastly, and the amount of players that completely ignored the cyberbullying cases.

After the posts and comments from their respective sessions come to an end, players had to respond to a series of endgame questions. Sessions 2, 3 and 4 had the following post-session questions at the end:

1. Do you think the aggressor's behaviour changed from the last session to this one?
2. What did you think of the aggressor's behaviour this session?
3. In what measure do you think the change in the aggressor's behaviour was due to your actions?
4. Why?
5. Do you think you had the possibility to make choices that influenced the story?
6. Do you think you will be able to change, for the better, the aggressor's behaviour in the future?

The first question had a simple binary (Yes or No) answer, followed up by the second question that had 3 possible answers: "Less appropriate", "Equal" and "More appropriate". The third question, as well as the 5th and 6th, were valued in a 5 point Likert scale. The 4th question, which was a follow-up to the third question, was open-ended. We analysed the frequency of these answers and compared them between the winning and losing groups through the use of relative frequency histograms, to see if there were frequency differences between these 2 groups of participants.

4.3.2 Interacting with Social Agents for Prosociality

To understand the players' perceptions of their interaction with the social agents in the game Com@Viver, we developed and validated the *Interacting with Social Agents for Prosociality* (ISAP) questionnaire with a Likert-type scale from 1 (not at all) to 5 (totally). An exploratory factor analysis using FACTOR 10.8.2 (Lorenzo-Seva & Ferrando, 2018) was computed to interpret the internal structure of the instrument. The table B.1 shows the correlations among all variables and the descriptive statistics. All variables revealed skewness values less than 2 and kurtosis values less than 5 as recommended by the literature (Bollen & Long, 1993). The data was tested with the Kaiser-Meyer-Olkin (KMO) and the Bartlett's Test of Sphericity for its underlying structure. The KMO measure of sampling adequacy was a good .84, whereas the Bartlett Sphericity was $X^2_{(36)} = 748.6$ ($p < .001$), revealing the variables as suitable for factor analyses. According to Bollen and Long (1993), there is multivariate normality if Mardia's coefficient is lower than $P(P + 2)$, where P is the number of observed variables. In this study, 13 observed variables were used with a Mardia's coefficient for skewness of $34.87 < 9(9 + 2) = 99$ and for kurtosis of $156.10 > 9(9 + 2) = 99$. Therefore, we used Unweighted Least Squares (ULS) as the method for factor extraction considering the kurtosis values. This estimation method is not dependent on distributional assumptions (Joreskog, 1977). To acquire the appropriate number of factors we used Horn Parallel analyses (Bandalos & Finney, 2010). Four items revealed loadings greater than .40 on two or more components, therefore, we considered the remaining items with structure coefficients values above .30 (Bandalos & Finney, 2010; Ford, MacCallum, & Tait, 1986). The analysis we performed presented a

four-factor model (*Perceived Influence of Social Agents on Players' Behaviour, Perceived Influence of Players' Behaviour on Social Agents' Behaviour, Self-efficacy Beliefs to Change Others' Behaviour, Perceived Social Agents' Behavioural Realism*) for the instrument with good reliability scores (See table B.1) according to the psychometric literature (Nunnally, 1978). Moreover, the values of goodness-of-fit ($GFI = 1.00$) and residuals statistics ($RMSR = .01$) were good according to the literature (McDonald, 1999; Nunnally, 1978; Velicer, 1976).

Confirmatory factor analyses were computed using IBM, SPSS AMOS 24.0 with estimation procedures of unweighted least squares, namely, fit indices such as chi-square, Root Mean Square Error of Approximation (RMSEA), Standardized Root Mean Square Residual (SRMSR), Comparative Fit Indices (CFI), Incremental Fit Index (IFI) and Akaike Information Criterion (AIC). CFI and IFI values close to 1 indicate a good statistical fit (Bentler, 1990), while RMSEA indicates a good fit if equal or less than .08 (Browne & Cudeck, 1993). As for the AIC, the lower value reveals the better fitting model. Moreover, the SRMR should be close to zero for a good fit. We tested the single-factor model to confirm the initial structure of the instrument from the exploratory factor analysis, which presented good reference values [$\chi^2_{(19)} = 20.43$, $p > .05$, $\chi^2/df = 1.07$, CFI = .99, IFI = .99, RMSEA = .02, LO=.00, HI=.07, SRMR = .02, AIC = 72.43] according to the literature (Hooper, Coughlan, & Mullen, 2008).

4.4 Procedures

This study was done in collaboration with Faculdade de Psicologia da Universidade de Lisboa (FPUL), that was responsible for getting proper authorization from students, parents, the school board, the Ministry of Education, and Comissão Nacional de Protecção de Dados (the Portuguese Data Protection Authority), to test Com@Viver in a portuguese school in the school year of 2019/2020.

The game was played by students inside the school classrooms, as opposed to another setting, so the experiment could be as relatable as possible. This was made in an attempt to maintain ecological validity, which means maintaining the integrity of the real-life situations of the experiment context. By maintaining high ecological validity, we can generalize the findings of our research to real-life settings. [19].

The study was split into 5 sessions. The first one, session 0, started with a facial validation where students tested the game so we could see if they understood the game and how to play it. The next 4 sessions, players were split in groups of 3 members and played one session per week. All data related to Player's scores, actions and choices during sessions were stored to be used for this study.

4.5 Data Analysis

4.5.1 Data gathered during gameplay

The gameplay data was analysed by verifying relative frequency distributions of quantitative variables. These variables consisted of the overall winning and losing groups distribution, winning and losing group distribution per session, the comparison of positive and negative comment options in dynamic conflict dialogues, chosen by players during the gameplay sessions, and the percentage of bystanders per cyberbullying case. The same process was utilized for the endgame questions associated with the variables "In what measure do you think the change in the aggressor's behaviour was due to your actions", "Do you think you had the possibility to make choices that influenced the story?" and "Do you think you will be able to change, for the better, the aggressor's behaviour in the future?".

4.5.2 Pre-post group comparisons

We performed an analysis of variance (ANOVA) for repeated measures for the variables Perceived Influence of Social Agents on Players' Behaviour and Perceived Influence of Players' Behaviour on Social Agents' Behaviour. Due to a non-normal distribution of the data we performed a Kruskal-Wallis test to understand if there were any significant differences between the EG and the CG regarding their Self-efficacy Beliefs to Change Others' Behaviour. Lastly, due to pretest differences, we computed an ANCOVA to examine any possible differences between the EG and the CG regarding their Perceived Social Agents' Behavioural Realism. All of these statistical procedures were performed with IBM SPSS 25.0.

4.5.3 Qualitative analysis of players' endgame responses

The data analysed regarding this topic was gathered through the qualitative question from the post-session questionnaire, with the objective of studying the player's reasoning behind their opinion on their self-efficacy in changing the game's characters behaviour, their perception of their behaviour and the AI agents' behaviour, and observing the most recurrent themes and trails of thought during the experiment. For these results we used content analysis, with a tenth of the analysis cross referenced with someone specialized in the theme of the study to check the Intraclass Correlation (ICC), that provided good results according to the literature $ICC(2, 2) = .89$, [20]. Specifically, 89% of the variance in the mean of both raters was true score variance.

The results are provided in table B.5, table B.6 and table B.7.

4.6 Results and Discussion

4.6.1 Process data gathered during gameplay

The following data was gathered from the gameplay sessions, and shows us the overall distribution of wins and losses, and quantitative trends noticed during the study, regarding the choices of actions taken during the gameplay, as well as the "closed answer" questions that were made to players at the end of each session.

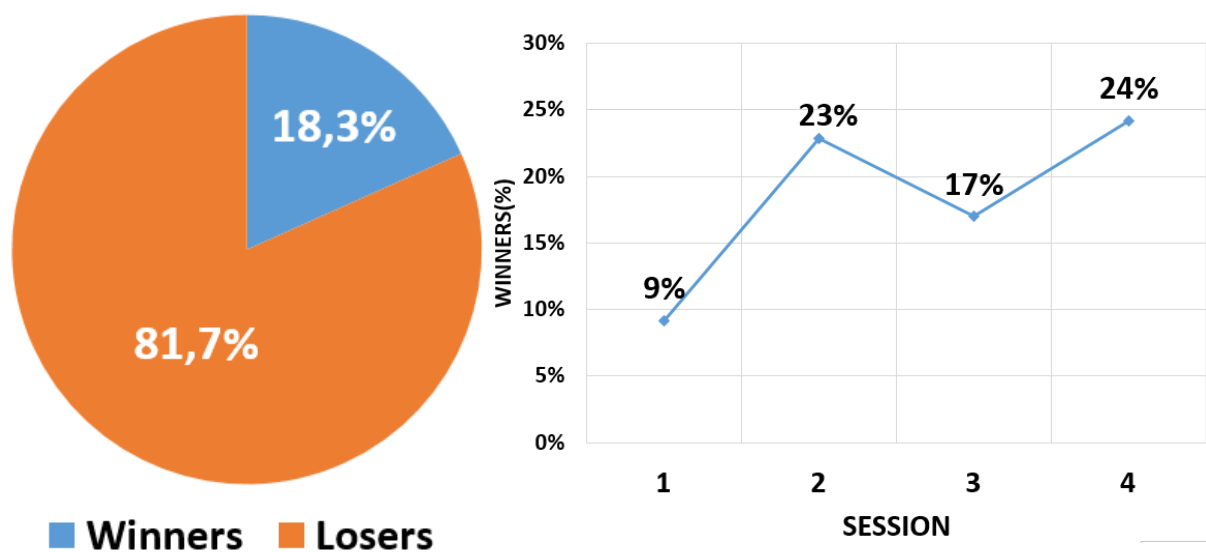


Figure 4.1: Winner distribution total

Figure 4.2: Winner distribution per session

Starting by analyzing the distribution between winners and losers, it is clearly shown by fig. 4.2 that the percentage of winning players increased dramatically over the course of the study. However, this increase was more noticeable between the first and second sessions of the study, with the number of winners jumping from 9% to 23%, topping at 24% in the last session. The total distribution of winners and losers across the 4 sessions was of 81.7% losers and 18.3% winners.

Secondly, fig. 4.3 shows, for sessions 2 and 3, the percentage of positive options that were chosen, by players, to comment during the cyberbullying case, relative to the total amount of times players commented something. The data is divided between winning and losing groups. In fig. 4.4 we can observe the same statistic, but for the average in conflict dialogues during gameplay in sessions 2 and 3.

It is intended as a "negative option" any option that engaged in aggressive behaviour, or that ignored the conflict at course. For example, if in a post about football, two AI Agents started arguing, and the player had the "positive option" to comment in a way that could improve the situation (i.e reprehend the characters behaviour or trying to calm things down), but instead chose an option that ignored the

situation (i.e talking about the football post itself), that option would be considered a negative option. It is also possible for a player to choose both positive and negative options in a post, since the player can choose all dialogue options to comment if he/she wants too.

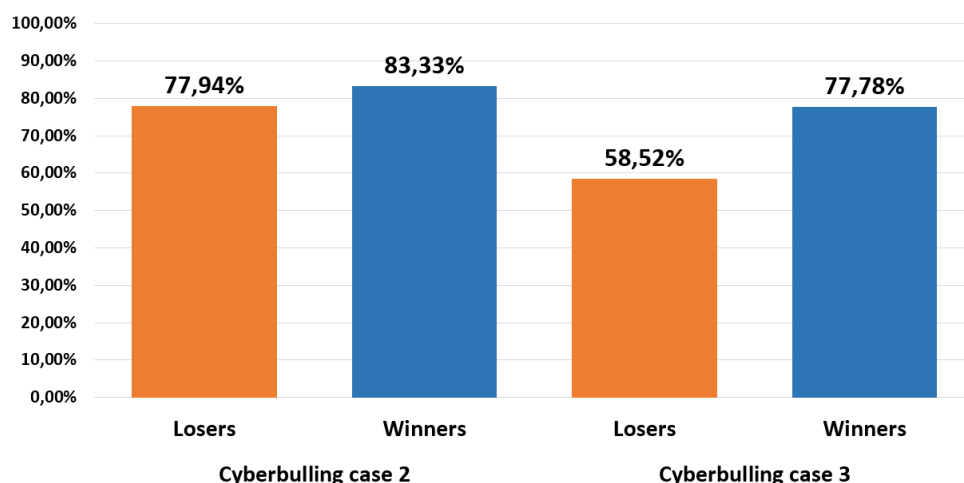


Figure 4.3: Percentage of positive comments chosen by players during cyberbullying cases 2 and 3 (Winners/Losers)

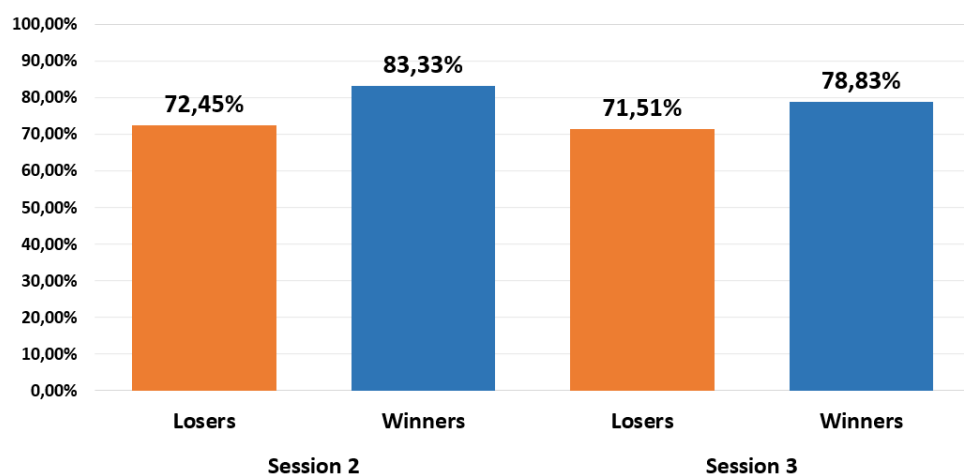


Figure 4.4: Percentage of positive comments chosen by players during conflict dialogues with alternative lines in sessions 2 and 3 (Winners/Losers)

These 2 specific sessions contain the majority of changes in dialogue in the experiment, meaning they are the most indicated to demonstrate changes in course of action taken by players. It is observable in both figures that players who had won the previous session, and by consequence, visualized changes in the behaviour of aggressors from the previous session, made positive comment choices more often than the players that did not see these changes.

If fig. 4.5 and fig. 4.6, it is visible the relative frequency of players that did not comment anything on the cyberbullying cases, and players that ignored the post. An important distinction to make is that

whenever a player does not comment anything on the post, but still reacts to it through liking or disliking, it is not considered as ignoring the post. Only players that did not react to the post, by not commenting nor liking or disliking, are counted as players that ignored the post.

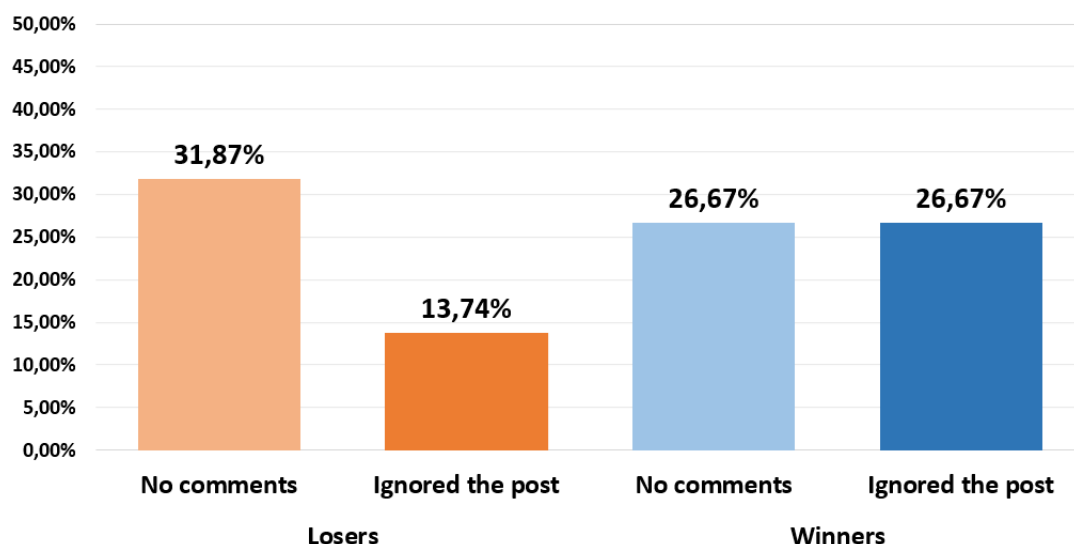


Figure 4.5: Percentage of players that did not comment or ignored the post in case 2 (Winners/Losers)

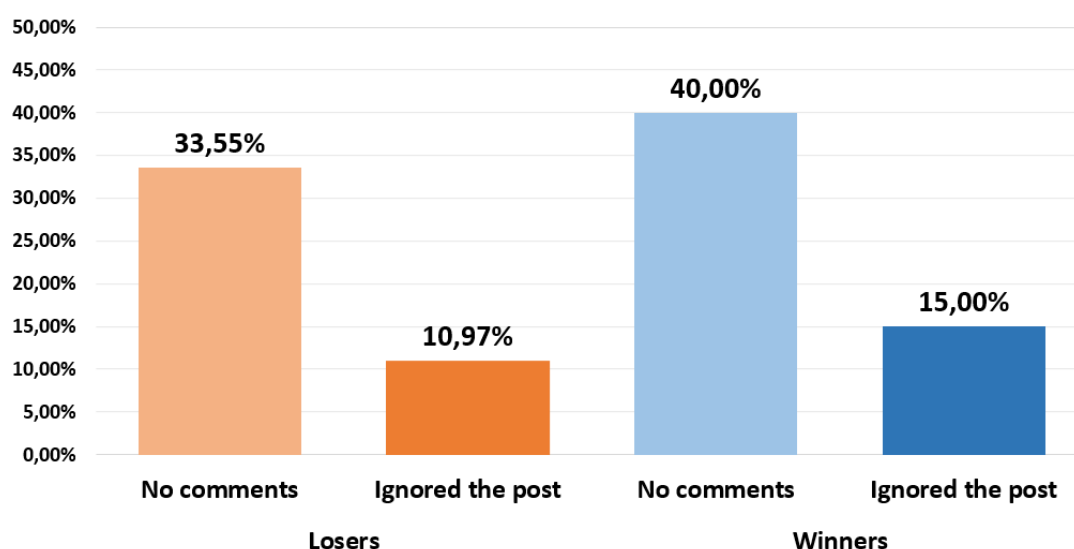


Figure 4.6: Percentage of players that did not comment or ignored the post in case 3 (Winners/Losers)

By analysing this figures, it is observable that the group of players that interacted more in positive ways during cyberbullying cases, also had a higher percentage of players that did not comment anything nor reacted to the posts. A possible explanation for this behaviour could be players liking positive comments of other group members, instead of commenting on their own, as a way of showing agreement with their team members' comments, or thinking their actions are enough to help the victim. This

behaviour would make sense from a pro-social perspective and what we see in social networks in real life scenarios, but it was not an evaluated form of behaviour in this experiment.

Continuing by analysing the endgame questions, regarding the question "In what measure do you think the change in the aggressor's behaviour was due to your actions", we can observe the histograms in fig. 4.7 and fig. 4.8.

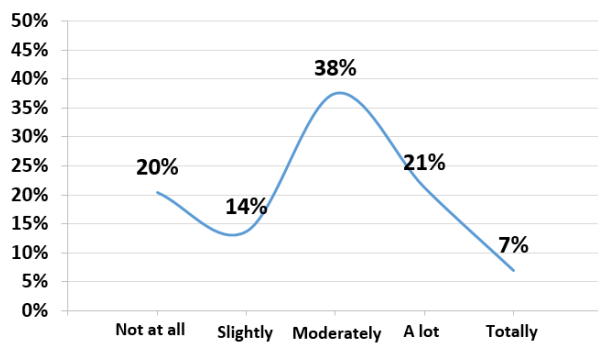


Figure 4.7: Self Efficacy (Total)

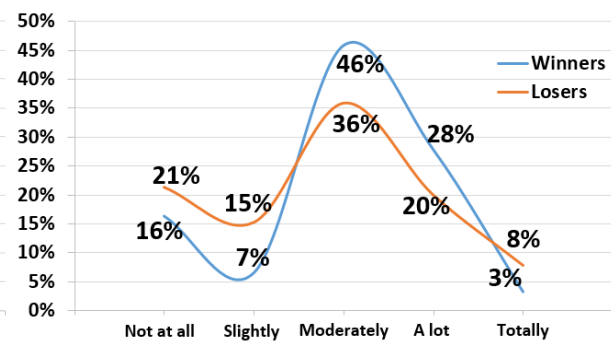


Figure 4.8: Self efficacy (Winners/Losers)

Through the comparison of the curves, we can see that there is a difference in the self-efficacy beliefs. The group of users that lost their gameplay sessions had a cumulative distribution of 36% in negative values (1 and 2) and 64% in the neutral and positive values (3 through 5), while the group of users that won their gameplay sessions had only 23% and 77% respectively, representing a 13% increase.

In respect to the questions "Do you think you had the possibility to make choices that influenced the story?" and "Do you think you will be able to change, for the better, the aggressor's behaviour in the future?", in fig. 4.9 and fig. 4.10 we can observe the answer distribution of these 2 questions.

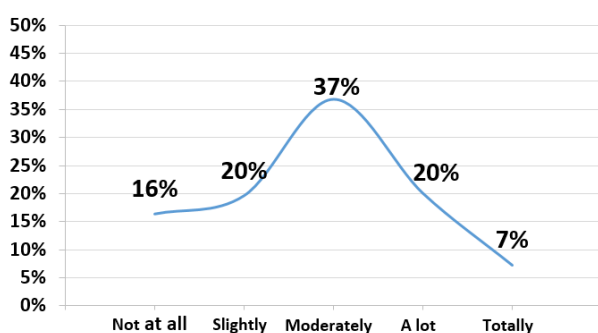


Figure 4.9: Possibility of making choices

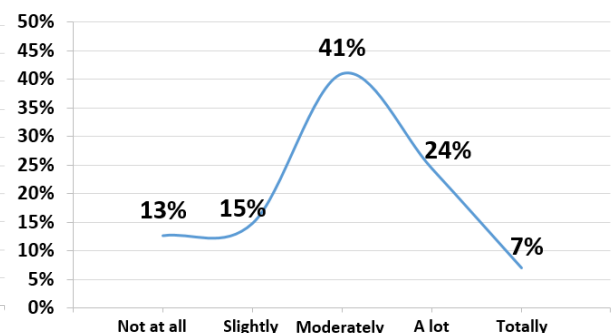


Figure 4.10: Aggressor's Behaviour Influence

In the first question, the negative values corresponded to 36% of the answers, against 27% in positive values, with 37% neutral, showing us that in the majority of cases, users felt unsure or believe they had little to no "free-will" during the play through. In the second question, the distribution was of 28% on negative values, with 31% positive and 41% neutral. These values, with just 3% difference between

positives and negatives, show us that users were divided or unsure about their possible future impact on the aggressor character. However, in fig. 4.11 we can observe that there is a slight deviation towards the positive and neutral values, of 14% of the answers, when comparing the players that had won the previous sessions versus the ones that did not. This means that the players who had won, and consequentially had seen a change in aggressors' behaviour, were 14% more likely to think they could have future impact on the aggressor's behaviour, relative to the players who lost.

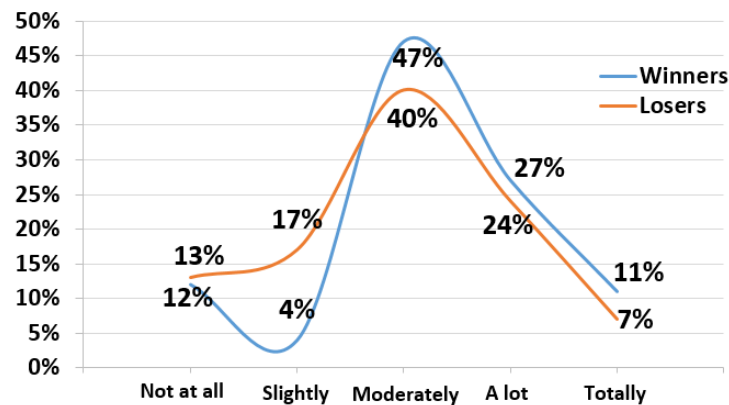


Figure 4.11: Aggressor's Behaviour Influence (Winners/Losers)

4.6.2 Pre-post group comparisons of players' perceptions of their interaction with the social agents

Due to a non-normal distribution of the data, we computed a Kruskal-Wallis which test showed that there was a statistically significant difference in self-efficacy beliefs to change others' behaviour score between the two groups, $\chi^2_{(1)} = 16,53$, $p = 0.00$, with a mean rank self-efficacy score of 166.25 for the EG and 126.54 for the CG.

Then, due to pre-test differences, we computed an ANCOVA for the players' perceived social agents' behavioural realism. There was a significant difference in mean concerning the social agents' behavioural realism [$F(1, 294) = 4,529$, <0.05 , $\eta^2_p = .02$] between conditions. In the ANCOVA analysis, the estimated mean for conditions concerning the social agents' behavioural realism in the pretest was 2.30. Post hoc tests showed there was a significant difference between the EG and CG ($p < 0.05$). The estimated marginal means showed that the EG revealed higher perceptions the social agents' behavioural realism at the end of the intervention with the game Com@Viver ($M = 2.47$) than the CG ($M = 2.183$).

There were no significant differences in the ANOVA for repeated measures for the variables "Perceived Influence of Social Agents on Players' Behaviour" and "Perceived Influence of Players' Behaviour on Social Agents' Behaviour".

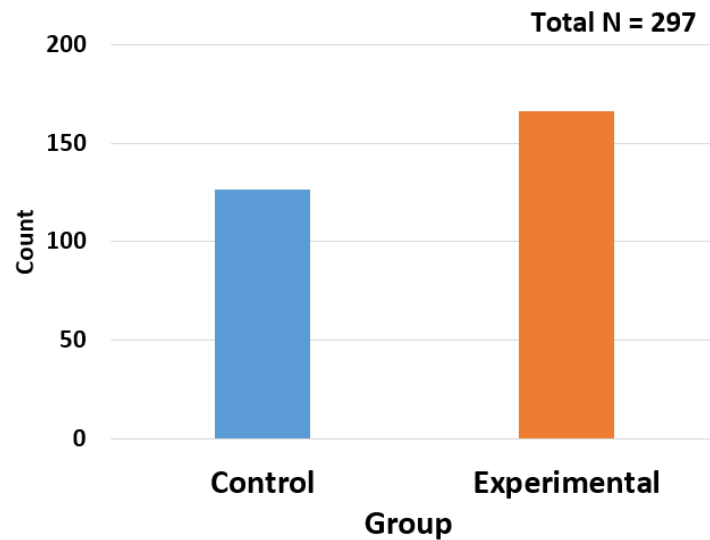


Figure 4.12: CG EG comparison

4.6.3 Qualitative analysis of players' endgame responses

As observable in tables table B.5 and table B.6, the answers were divided in a total of 9 categories. The most recurring theme was "Cyberbully Self-Improvement" (30.22%), that refers to the perceived changes that showed the cyberbully's behavioural improvement. This is probably due to a results-oriented thinking, where students based their opinion on the observable changes in cyberbully behaviour. This is directly related to our goal, since we aimed to create dynamic behaviour changes in our agents, so we could improve players' notion of self-efficacy.

This category was followed by "Bystander Effect / Lack of change" (16.20%), referring to observations that try to justify being a bystander instead of acting pro-socially, as well as doubting one's self-efficacy.

The third most common theme was "Bystander Prosocial Intervention" (11.84%), that refers to how bystanders should intervene in cyberbullying situations in a prosocial manner. In succession, "Negative emotions towards cyberbullying situations" (9.03%) was the fourth most recurrent theme, that refers to the emotions caused by, or that lead to, cyberbullying situations.

Next, the category "Internet / Social Network" (8.72%) refers to the usage of technology within the context of the game and cyberbullying. "Cyberbullying Behaviour" (7.79%) refers to behaviours associated with cyberbullying, "Positive concepts related to stopping bullying" (7.48%) refers to positive concepts associated with the changes that happened during the gameplay.

Finally, the two less frequent categories were "Ways of embracing differences" and "Motives behind cyberbullying behaviour" (both with 4.36%). These categories refer to how one can respect differences without engaging in bullying behaviour and what can lead to said behaviour, respectively. This occurrence may be caused by the students having a harder time identifying the main concepts that bullying

"feeds" on, and in contrast, find it easier to describe what they feel and/or see when being a bystander.

4.6.4 Discussion

As explained previously, we aimed to answer the following research questions:

1. Can using affective AI agents encourage prosocial behaviour?
2. Do players perceive that their actions can affect the behaviour of AI agents in a game?
3. Can players' self-efficacy to promote pro-social behaviour improve by interacting with AI agents?
4. Does players' perception of the behaviour of AI agents change through interaction in a game?
5. Can players' behaviour change in the game specifically with regards to the changes they observe in the agents' behaviour?

To answer these research questions, we analysed the data gathered in the previous sections.

Can using affective AI agents encourage prosocial behaviour?

Although fig. 4.1 shows that the number of prosocial groups was far outweighed by the number of non-prosocial groups, we can see in 4.2 that there was a clear evolution in the number of winning groups between the first session and the second session onwards.

Considering that the group scoring system inside the game is the main indicator of prosocial activity, since that in order to win the players must behave in a prosocial manner, we can conclude that this objective was achieved, and the usage of affective AI Agents in a game can encourage prosocial behaviour.

Do players perceive that their actions can affect the behaviour of AI agents in a game?

In our analysis of pre-post group comparisons of players' perceptions, there were no significant differences for the variable "Perceived Influence of Players Behaviour on Social Agents' Behaviour". In fig. 4.9 it is also observable that players felt unsure about their possibility of making choices during sessions, with more players thinking that they could not.

However, in our qualitative analysis, the category "Bystander Prosocial Intervention" was the third most referred category (11.84%), with the subcategory "Influence" making up for almost half the answers.

It is also observable that, in fig. 4.7, when asked if they believed the changes in the aggressor's behaviour was due to their actions, the majority of players answered "moderately". There were more players answering they were not at all or slightly convinced the changes in the behaviour of AI agents

had anything to do with them, than players answering they believed to have a lot or all responsibility for the fact. This statistic, however, had a 13% increase to positive values when comparing the answers of players from winning and losing groups. In this final comparison, the first 2 values of the Likert scale made up for 23% of answers, while the last 3 values had 77%.

The statistics referred allow us to conclude that, although there were no significant differences between the CG and the EG, endgame answers show that some players believed they had an impact in the behaviour of AI agents.

Can players' self-efficacy to promote pro-social behaviour improve by interacting with AI agents?

From our analysis of pre-post group comparisons of players' perceptions, there was a statistically significant difference in self-efficacy beliefs to change others' behaviour, with players from the EG showing a higher self-efficacy score.

As noted previously, in our qualitative analysis, the category "Bystander Prosocial Intervention" was the third most referred category. In this category, subcategories "Influence" and "Defend" make up for more than 50% of the answers.

The qualitative analysis of the endgame question "Do you think you will be able to change, for the better, the aggressor's behaviour in the future" (fig. 4.10, showed us there was only a 3% difference between players that believed they could influence the AI characters in the future, and players that doubted themselves on the topic, with the majority of players believing they could moderately influence the games' characters. This is a small difference, however, the winning players, that had visualized behavioural change in characters due to their previous actions, were 14% less likely to doubt their impact when answering the question (fig. 4.11).

This evidence suggests that the interaction with the AI Agents had a clear impact in increasing player self-efficacy beliefs regarding promoting pro-social behaviour.

Does players' perception of the behaviour of AI agents change through interaction in a game?

According to our analysis of pre-post group comparisons of players' perceptions, the players that played the computer version of the game with the FAtiMA system implemented had a higher perception of the Agents' behavioural realism.

It is also notable that during our qualitative analysis (table B.5) the category "Cyberbully Self-improvement", that refers to the perceived changes that showed the cyberbully's behavioural improvement, was the most frequent category (30.22%), and was directly tied with the perception of AI agents' behaviour. In the same analysis, the category "Internet / Social Network", that relates to the usage of technology within the context of the game and cyberbullying, only had 1 answer related to the concept of "game as a tool", where the player said our experience was just a game and the characters where fake. The

small percentage of these occurrences is a positive result, especially considering the amount of different answers obtained during the experiment.

These 2 occurrences confirm our theory that interacting with the game and its' agents will make players find virtual agents' behaviour more realistic.

Can players' behaviour change in the game specifically with regards to the changes they observe in the agents' behaviour?

As referred previously, the number of winning groups increased after the first session. For this topic, it is important to consider that players were only able to see differences in dialogue after the first session, since they first needed to win a session in order to change aggressors' behaviour in the next session.

Analysing fig. 4.3 and fig. 4.4 showed us that winning groups behaved better than losing groups during the cyberbullying cases and conflict dialogues of future sessions, the sessions in which they saw different dialogues than the losing groups.

In particular, during session 3, players that had won session 2 would see, right before the cyberbullying case, a dialogue where 2 of the aggressors of session 2 would behave pro-socially (Agents Carmen and Isabel, dialogue 19 table B.3). This seems to have had a major impact in the outcome of player actions, with the relative frequencies of positive versus negative interactions being much more inclined to the positive side.

These results suggest that player behaviour can be changed and improved by making players observe behavioural change in characters.

5

Conclusion

Contents

5.1 Future Work	53
---------------------------	----

Our aim with this project was to improve the previously created serious game "Com@Viver" by using AI agents paired with the FAtiMA system and theory of SI, so we could improve the user experience, perceived self-efficacy, player behaviour, and grant the users a better sense of feedback on actions taken during the play through.

We addressed the implementation of the FAtiMA Agent Architecture through the usage of the FAtiMA toolkit, in the pre-existing game Com@Viver, to improve the game effectiveness in incentivizing prosocial behaviour. By creating dynamic dialogues, we made players see different consequences of their actions, so we could observe if these changes in the behaviour of virtual agents inside a game could influence the behaviour of players.

In collaboration with psychologists from FPUL, we were able to improve the current Social Network and game characters' AI, to better illustrate and encourage behavioural change. We conducted a study to elaborate on the data gathered during the experiment, and deduct the impact that said changes in the game had on the players' perception of their self-efficacy, and the effect on bystander behaviour.

The results of this study have shown that, with the usage of AI Agents with dynamic interaction, that demonstrate changes in their behaviour according to player actions, we can improve players' behaviour, self-efficacy towards being pro-social, and self-efficacy towards changing others' behaviour.

Our works contribution to the field lies in important distinctions when compared to the other serious games and experiments related to the cyberbullying theme. According to the literature [9], changing behaviour, increasing awareness and increasing empathy are topics that have been explored successfully in serious games with themes other than cyberbullying, but researches inside the theme did not try to do the same. Most serious games are played through only one session, and are focused in raising awareness or teaching coping strategies. Experiments usually involve one single questionnaire after the intervention, or one pre-post questionnaire, but not both a pre-post questionnaire and questions during the experiment. There have been experiments where the players were asked if they liked the game or found it useful, but not if they found the characters to be realistic, nor if they believe they could influence, or be influenced by, said characters. Our work distinguishes itself in all these topics, and contributes directly against these flaws.

5.1 Future Work

One of the limitations of the project were the logistics involved. The game was played by students inside the school classrooms in an attempt to maintain ecological validity. This ultimately led to extra effort by our team in order to properly organise all the students' sessions and group distribution, as well as managing schedules and available classrooms. One aspect that could help minimizing this problem would be restructuring the game's interface, implementing a proper menu and saved sessions system, to

help keeping track of the gathered data throughout the experience, and help organizing game sessions.

Another problem was related to monotony and lack of excitement amongst the players during their interactions, mostly due to the low interactivity of being limited to comments on Agents' posts. The players could only choose from a set of predetermined options in each post when commenting, and the reasoning behind this limitation was to guarantee that a player could not engage in aggressive cyberbullying behaviour directed towards other players. Consequently, players also expressed their wishes on being allowed to write their own answers instead of picking options, which could be interesting to analyse as well, and constitutes a future goal. It would, however, require a different approach in answer processing and decoding, demanding text analysis during gameplay.

Considering that the literature reports gender differences in cyberbullying experiments, it could also be interesting to see if there are significant gender differences in Com@Viver gameplay, as well as age differences, to test if younger players are more easily influenced.

The narrative could be further adapted to incorporate more behavioural changes, and perhaps even roles changes between characters in the story. Currently, the characters roles are fixed and the game will always play in the same way. The aggressors and victims of each cyberbullying case will always be the same ones, and apart from the specific dialogues that are dynamic, everything stays the same. Dynamic roles and relationships could be implemented in order to fully utilize the capabilities of the FAtiMA architecture, create an ever-changing experience.

Creating more dynamic dialogues and more "points of influence" would also be beneficial. At the current state, the only way to have an effect on the characters dialogues is to win as a group. This creates limitations since the majority of players could not win, and the dialogue differences are only visible on the following session. The existence of more specific points in the story where players can influence characters would create the possibility for changing dialogues inside the session currently being played, and could remove the necessity of winning to see those changes.

Bibliography

- [1] T. M. Connolly, E. A. Boyle, E. MacArthur, T. Hainey, and J. M. Boyle, “A systematic literature review of empirical evidence on computer games and serious games,” *Computers Education*, vol. 59, no. 2, pp. 661 – 686, 2012. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0360131512000619>
- [2] D. L. Poole and A. K. Mackworth, *Artificial Intelligence: Foundations of Computational Agents*, 1st ed., C. U. Press, Ed. Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo, Delhi, Dubai, Tokyo: Cambridge University Press, 2010.
- [3] J. M. van der Zwaan, V. Dignum, and C. M. Jonker, “A conversation model enabling intelligent agents to give emotional support,” in *Modern Advances in Intelligent Systems and Tools*, W. Ding, H. Jiang, M. Ali, and M. Li, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 47–52.
- [4] S. Jain, B. Thiagarajan, Z. Shi, C. Clabaugh, and M. Mataric, “Modeling engagement in long-term, in-home socially assistive robot interventions for children with autism spectrum disorders,” *Science Robotics*, vol. 5, p. eaaz3791, 02 2020.
- [5] T. J. M. Vicente, “Com@Viver: A serious game to improve bystander behavior,” no. September, 2018.
- [6] D. J. Å and D. P. Farrington, “Examining the Relationship Between Low Empathy and Bullying,” vol. 32, no. June 2005, pp. 540–550, 2006.
- [7] J. van der Zwaan, V. Dignum, and C. Jonker, “Simulating peer support for victims of cyberbullying,” *Belgian/Netherlands Artificial Intelligence Conference*, 2010.
- [8] A. Paiva, F. Santos, and F. Santos, “Engineering pro-sociality with autonomous agents,” 2018. [Online]. Available: <https://aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/16799>
- [9] A. Calvo-Morata, C. Alonso-Fernandez, M. Freire, I. Martinez-Ortiz, and B. Fernández-Manjón, “Serious games to prevent and detect bullying and cyberbullying: A systematic serious games and literature review,” *Computers Education*, p. 103958, 07 2020.

- [10] R. W. Picard, "Affective Computing," Tech. Rep. 321, 1995.
- [11] S. C. Marsella and J. Gratch, "EMA : A process model of appraisal dynamics," no. January 2018, 2009.
- [12] J. Dias, S. Mascarenhas, and A. Paiva, "Fatima modular: Towards an agent architecture with a generic appraisal framework," vol. 8750, 11 2014.
- [13] K. Anderson, E. André, T. Baur, S. Bernardini, M. Chollet, E. Chryssafidou, I. Damian, C. Ennis, A. Egges, P. Gebhard, H. Jones, M. Ochs, C. Pelachaud, K. Porayska-Pomsta, P. Rizzo, and N. Sabouret, "The tardis framework: Intelligent virtual agents for social coaching in job interviews," in *Advances in Computer Entertainment*, D. Reidsma, H. Katayose, and A. Nijholt, Eds. Cham: Springer International Publishing, 2013, pp. 476–491.
- [14] J. McCoy, M. Treanor, B. Samuel, A. A. Reed, M. Mateas, and N. Wardrip-Fruin, "Social story worlds with Comme il Faut," *IEEE Transactions on Computational Intelligence and AI in Games*, vol. 6, no. 2, pp. 97–112, 2014.
- [15] R. S. Aylett, S. Louchart, J. Dias, A. Paiva, and M. Vala, "Fearnott! – an experiment in emergent narrative," in *Intelligent Virtual Agents*, T. Panayiotopoulos, J. Gratch, R. Aylett, D. Ballin, P. Olivier, and T. Rist, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 305–316.
- [16] K. P. Star and J. P. Allsopp, "D2.2 Prosocial Game Scenarios," ProSocialLearn, Tech. Rep., 2015.
- [17] M. Mateas and A. Stern, "Structuring content in the façade interactive drama architecture," in *Proceedings of the First AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, ser. AIIDE'05. AAAI Press, 2005, p. 93–98.
- [18] S. Mascarenhas, N. Degens, A. Paiva, R. Prada, G. J. Hofstede, A. Beulens, and R. Aylett, "Modeling culture in intelligent virtual agents," *Autonomous Agents and Multi-Agent Systems*, vol. 30, no. 5, pp. 931–962, Sep 2016. [Online]. Available: <https://doi.org/10.1007/s10458-015-9312-6>
- [19] M. Schmuckler, "What is ecological validity? a dimensional analysis," *Infancy*, vol. 2, 10 2001.
- [20] K. McGraw and S. Wong, "Forming inferences about some intraclass correlation coefficients," *Psychological Methods*, vol. 1, pp. 30–46, 03 1996.



Appendix

Characters Backstory

Tatiana Delgado

A Tatiana é uma jovem tímida e estudiosa. Tira sempre boas notas. A Tatiana não é muito popular na escola, mas tende a defender os colegas que ela pensa que precisam de ajuda. Há um grupo de 3 colegas (i.e. a Manuela, a Patrícia e a Cármén) na turma da Tatiana que costuma gozar repetidamente com ela na sala de aula, atirando-lhe por vezes objetos às costas dela sem que os professores se apercebam. Quando estes incidentes acontecem, a Tatiana fica muito magoada fisicamente e psicologicamente. Há uns tempos, a Tatiana criou um perfil falso (i.e. sob o usuário “João”) no Com@Viver e começou a conversar com a Patrícia, a colega que normalmente gozava mais com ela nas aulas. A Patrícia demonstrou bastante interesse pelo “João”, não só pela simpatia nas suas interações, como também pelas fotos atraentes que foram colocadas no seu perfil. Depois, a Tatiana, fingindo ser o “João”, disse à Patrícia que não gostava mais dela porque a achava feia. A Patrícia começou a receber constantes mensagens agressivas privadas do perfil do “João” (i.e. tanto em casa, como na escola), como por exemplo: “Não vales nada! És gorda e ninguém te quer!” Nessa altura, a Patrícia sentiu-se muito mal e teve de ir para o hospital. O namorado atual da Patrícia, o Nando, não gostou desta história. Em relação ao Samuel, a Tatiana não sabe que este gosta dela.

Nando Sapina

O Nando é um aluno popular na escola que tem boas notas e joga futebol. Gosta muito de animais e nos seus tempos livres, trabalha num canil como voluntário para ajudar os cães que não têm casa. O Nando é o namorado atual da Patrícia. Quando a Patrícia lhe contou o que se tinha passado com a Tatiana, o Nando não gostou e partiu em defesa da namorada, postando comentários e fotografias sobre a Tatiana. Ou seja, cada vez que vê a Tatiana ou há algo na Internet que envolva a Tatiana, o Nando faz questão que ela sinta o mesmo que a Patrícia sentiu. O Nando defende sempre a sua namorada.

Patrícia Isidoro

A Patrícia é uma jovem estudiosa e fofoqueira. Gosta de passar tempo com os seus amigos a estudar, a ouvir música e a navegar nas redes sociais. Há uns tempos atrás, começou a gozar com a Tatiana na escola, pelo que esta criou um perfil falso e começou a comunicar com a Patrícia. A Patrícia sentiu-se mal e teve de ir para o hospital porque o conhecido da Internet (o João, um perfil falso criado pela Tatiana) começou a enviar-lhe mensagens hostis passado uns tempos no Com@Viver. Esta situação tornou-se pública e todos os colegas da escola sabem o que a Tatiana fez. Apesar de ter gozado com a Tatiana nas aulas, a Patrícia sente-se injustiçada pelo que ela fez online. A Patrícia namora com o Nando e contou-lhe esta história. Também sabe que o Samuel gosta da Tatiana, e como forma de

ajuste de contas, a Patrícia começou a colocar comentários sobre o Samuel no Com@Viver.

Manuela Leitão

A Manuela é uma jovem extrovertida. Gosta de dançar ballet e de coscuvilhices com as colegas. Já ganhou um prémio europeu de ballet na categoria de juniores, sendo desta forma, uma boa representante da escola. Há uns tempos, a Manuela, em conjunto com a Patrícia e a Cármén, começou a gozar com a Tatiana na escola. A Manuela costumava ser amiga da Tatiana na escola primária, mas desde que conheceu a Patrícia e a Cármén, que se afastou na Tatiana. Por vezes a Manuela ficava triste por estar a gozar com a Tatiana, mas tinha receio que se ela não o fizesse, que a Patrícia e a Cármén se virassem contra ela e que comesçassem a gozar também com ela, especialmente porque costumava ser amiga da Tatiana. A Manuela não sabia como lidar com a situação e continua arrependida por ter feito o que fez à Tatiana. No entanto, também não gostou do que a Tatiana fez à Patrícia com o perfil falso. Neste momento, a Manuela está confusa com a situação toda e não sabe o que pensar.

Cármén Semedo

A Cármén é uma jovem alegre com um talento especial para cantar. Canta no coro da escola e tem uma pequena legião de fãs que são colegas da escola. A Cármén é amiga do seu amigo e não gosta de injustiças. É amiga da Patrícia desde a escola primária. Desde essa altura que costumam frequentar as casas de uma da outra. Como os pais da Cármén são muito protetores, não a deixam sair muito para além da casa da Patrícia. Como se conhecem bem, são amigas muito unidas. A Cármén nunca gostou da Tatiana porque nunca falava com ninguém e sempre a achou antipática. Agora gosta ainda menos dela desde que a Tatiana criou o perfil falso para comunicar com a Patrícia. Não gostou do que ela fez à amiga. Nos últimos meses, a Cármén tem escrito alguns posts agressivos à Estrela porque gostaria de ir à visita de estudo e preferia que o destino da visita fosse outro. Ela acha que os pais não a vão deixar ir porque o Algarve fica muito longe. A forma que a Cármén arranjou de lidar com a sua frustração, foi a de confrontar a Estrela Online por causa da visita de estudo. A Cármén pensa que a Estrela colocou o destino da visita como sendo o Algarve porque quis agradar à sua melhor amiga, a Isabel Torres, que sempre quis ir lá com os amigos.

Rui Bento

O Rui é um jovem estudioso, sendo o melhor da turma a matemática, e tendo ganho sempre os torneios nacionais dessa área. O Rui gosta de fazer voluntariado na Associação de Apoio à Vítima e por isso, é muito sensível a situações e comentários negativos por parte dos colegas. Quando o Rui lê algo na Internet que não lhe agrada, ele normalmente posicionase, defendendo a pessoa que está a ser alvo de comentários inapropriados.

Abel Polido

O Abel é um jovem muito atlético e bem-humorado. Está sempre a contar anedotas e a fazer a malta rir. Pertence ao clube de Judo e à equipa de futebol da escola. Quando joga nos torneios, ganha sempre. O Abel é amigo de infância do Nando e faz tudo pelo seu amigo. Quando percebe que algo está errado com ele, parte sempre em sua defesa. No entanto, tem um fraquinho pela Patrícia, mas o Nando não sabe. O Abel vive num bairro social e a sua família tem muitas dificuldades financeiras, pelo que nem sempre consegue acompanhar os seus amigos nas saídas. Por norma, o Nando até o ajuda a comprar alguns materiais da escola.

Estrela Nunes

A Estrela uma jovem responsável. É delegada da turma e está muito empenhada na organização da visita de estudo. Ela dedica muito do seu tempo a questões relacionadas com a escola e vai sempre em defesa dos colegas. A melhor amiga dela é a Isabel Torres, uma jovem muito extrovertida. No geral, os alunos da turma gostam da Estrela. No entanto, há uma aluna da turma, a Cármen, que a tem chateado nos últimos meses por causa da visita de estudo. A Estrela normalmente ignora-a, mas os posts da Cármen têm-se tornado cada vez mais agressivos.

Jorge Amaral

O Jorge é um jovem que tem muitas dificuldades com os estudos e apesar de não ter muitos amigos, tem um grupo de amigos próximos que lhe dão muito apoio. Gosta de estudar com os amigos, o Abel, o Rui e o Samuel porque o ajudam com a matéria. O Jorge é muito defensor deste grupo de amigos e normalmente apoia as suas decisões. Sabe que o Abel não gosta do Hélder e por isso também nunca engraçou muito com ele, apesar de não ter nada especificamente contra ele.

Isabel Torres

A Isabel é uma jovem muito extrovertida e popular. Tem algumas dificuldades na escola porque não se consegue concentrar muito nos estudos. Gosta de estar com os amigos e de viajar. A Isabel sempre quis ir ao Algarve com a sua melhor amiga, a Estrela, mas os pais até agora nunca a deixaram ir. Ela acredita que se for uma visita organizada pela escola, que os pais a vão deixar ir. A Isabel tem tentado convencer a sua amiga Estrela, que está a organizar a visita de estudo, a marcar a visita para o Algarve. Outros amigos da Isabel incluem o Samuel, o Hélder e a Tatiana.

Samuel Andrade

O Samuel é um aluno muito inteligente, mas pouco popular na escola. Gosta de jogar videojogos e de estudar sozinho na biblioteca. Por vezes, o Samuel é gozado pelos colegas por ele gostar de estar sozinho na biblioteca, o que o leva a ter poucos amigos. Tem um fraquinho pela Tatiana desde há uns meses, mas não sabe como se poderá aproximar dela. Ele pensa que talvez a visita de estudo seja uma boa oportunidade para se aproximar dela. Na semana passada, a Patrícia descobriu que o Samuel gosta da Tatiana e desde então, não o tem largado nas redes sociais.

Hélder Almeida

O Hélder Almeida é um aluno muito popular na escola. Gosta de falar com todas as pessoas e é o diretor da Secção de Solidariedade da Associação de Estudantes. Tem espírito de missionário e quando for adulto, quer viajar pelo mundo para ajudar os povos mais desfavorecidos. O Hélder não gosta de intrigas e quando vê alguém a falar mal de outra pessoa, fica muito agressivo. Por vezes coloca comentários na Internet que podem ser considerados ofensivos. No entanto, na visão do Hélder, só o faz para defender quem ele considera serem os mais fracos da escola.

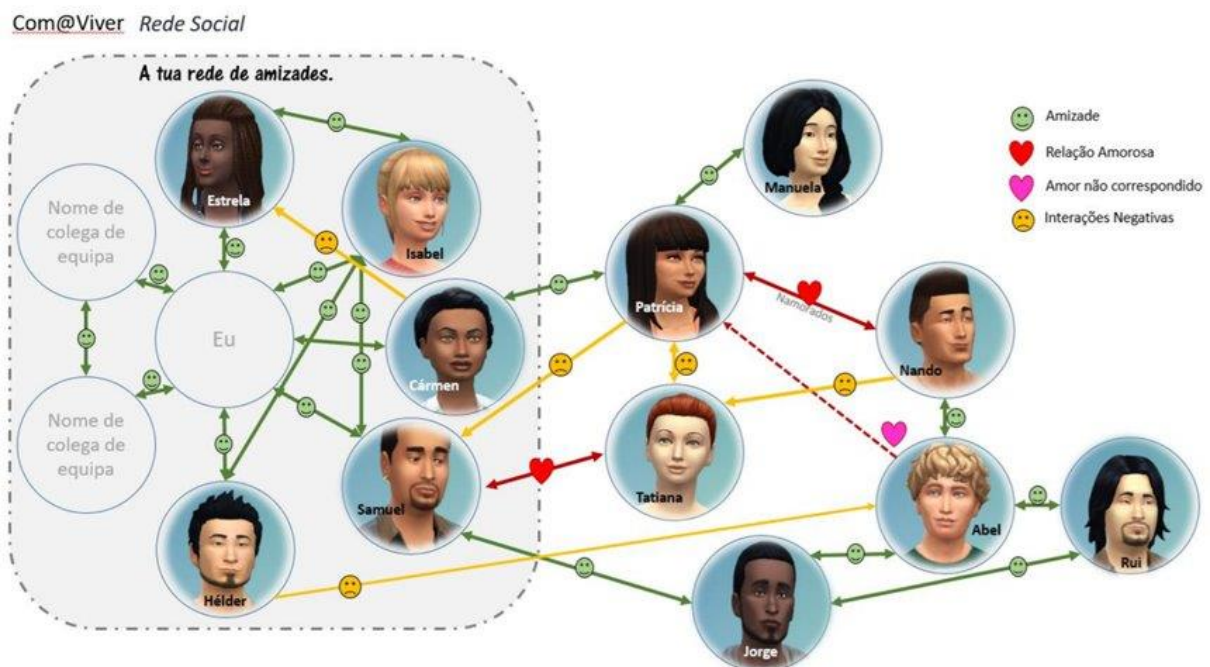


Figure A.1: The Com@Viver Network.

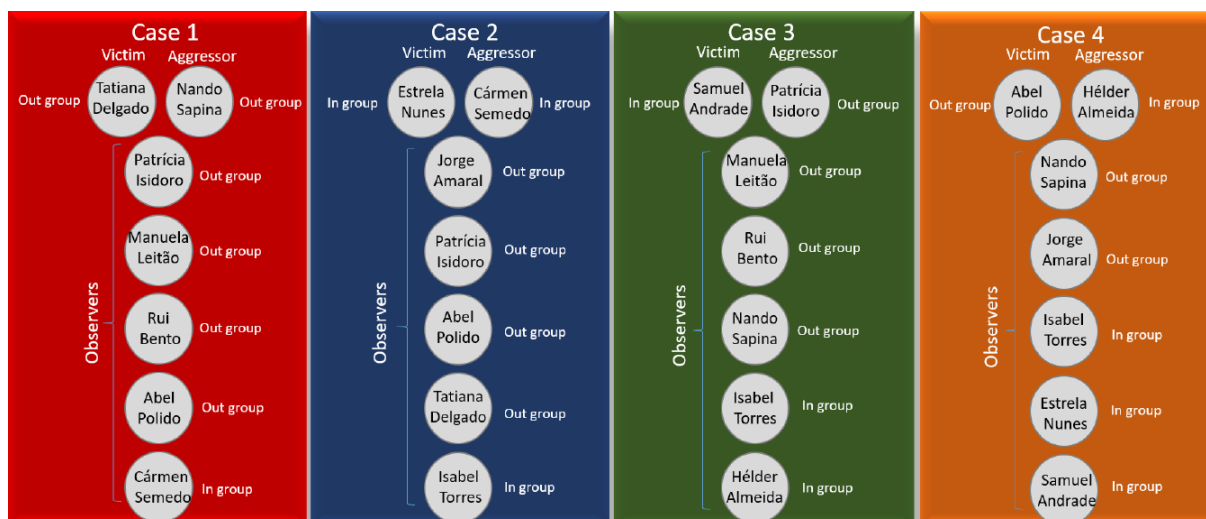


Figure A.2: Com@Viver cyberbullying cases.



Patrícia Isidoro
03/09/2020, 17:51:59

Pensava que eras paneleiro Samuel. Afinal apaixonaste-te pelo maior desperdício de oxigénio que anda por aí!

Gosto
Não Gosto
Comentar

0 Gostos 0 Não Gostos



Rui Bento 03/09/2020, 17:51:59
Estás a ser um bocado parva Patrícia.

Gosto
Não Gosto

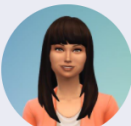
0 Gostos 0 Não Gostos



Abel Polido 03/09/2020, 17:51:59
A patrícia está só a gozar. Já não há sentido de humor.

Gosto
Não Gosto

0 Gostos 0 Não Gostos



Patrícia Isidoro 03/09/2020, 17:51:59
Não te metas Rui!

Gosto
Não Gosto

0 Gostos 0 Não Gostos



Nando Sapina 03/09/2020, 17:51:59
O Rui só quer chamar-te à razão.

Gosto
Não Gosto

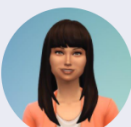
0 Gostos 0 Não Gostos



Isabel Torres 03/09/2020, 17:51:59
Não chateiem o Samuel. Que mal é que vos fez?

Gosto
Não Gosto

0 Gostos 0 Não Gostos



Patrícia Isidoro 03/09/2020, 17:52:00
Isabel. Está calada! Queres levar na boca? Mete-te na tua vida!

Gosto
Não Gosto

0 Gostos 0 Não Gostos



Hélder Almeida 03/09/2020, 17:52:00
Patrícia, daqui a pouco quem te dá na boca sou eu.

Gosto
Não Gosto

0 Gostos 0 Não Gostos

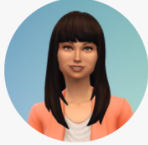


Nando Sapina 03/09/2020, 17:52:00
Não ameaces a Patrícia, isso só piora tudo!

Gosto
Não Gosto

0 Gostos 0 Não Gostos

Figure A.3: Case 3 - Nando Sapina showing prosocial behaviour due to winning groups.



Patrícia Isidoro
03/09/2020, 17:51:59

Pensava que eras paneleiro Samuel. Afinal apaixonaste-te pelo maior desperdício de oxigénio que anda por aí!

 Gosto

 Não Gosto

 Comentar

0 Gostos 0 Não Gostos



Rui Bento 03/09/2020, 17:51:59
Estás a ser um bocado parva Patrícia.

 Gosto

 Não Gosto

0 Gostos 0 Não Gostos



Abel Polido 03/09/2020, 17:51:59
A patrícia está só a gozar. Já não há sentido de humor.

 Gosto

 Não Gosto

0 Gostos 0 Não Gostos



Patrícia Isidoro 03/09/2020, 18:00:47
Não te metas Rui!

 Gosto

 Não Gosto

0 Gostos 0 Não Gostos



Nando Sapina 03/09/2020, 18:00:48
Ya, não te metas Rui.

 Gosto

 Não Gosto

0 Gostos 0 Não Gostos



Isabel Torres 03/09/2020, 18:00:48
Não chateiem o Samuel. Que mal é que vos fez?

 Gosto

 Não Gosto

0 Gostos 0 Não Gostos



Patrícia Isidoro 03/09/2020, 18:00:48
Isabel. Está calada! Queres levar na boca? Mete-te na tua vida!

 Gosto

 Não Gosto

0 Gostos 0 Não Gostos



Hélder Almeida 03/09/2020, 18:00:48
Patrícia, daqui a pouco quem te dá na boca sou eu.

 Gosto

 Não Gosto

0 Gostos 0 Não Gostos



Nando Sapina 03/09/2020, 18:00:48
Ó Hélder, ameaças a Patrícia e trato eu de ti!

 Gosto

 Não Gosto

0 Gostos 0 Não Gostos

64

Figure A.4: Case 3 - Nando Sapina showing aggressive behaviour due to losing groups.

B

Additional Tables

Table B.1: Item descriptive statistics, exploratory factor analysis parameters, reliability and correlations

	Items	Structure Coefficients	Cronbach's Alpha	Eigenvalues	M(SD)	1	2	3	4	5	6	7	8	
1	Can a character from a game influence your behaviour towards your classmates?	.52	.97	4.88	1.57(.97)									
2	Can a character from a game influence your behaviour online?	1.0			1.58(.94)	0.63**								
3	Can your actions change the course of a video game story?	.85	1.00	1.05	2.10(1.31)	0.35**	0.32**							
4	Can your actions contribute towards a more pro-social behaviour from a video game character?	.78			1.90(1.18)	0.39**	0.46**	0.68**						
5	Can you change the behaviour of a video game character, even if there are consequences towards you inside the game?	.69	.92	.99	2.12(1.32)	0.35**	0.37**	0.52**	0.63**					
6	Can you change the behaviour of someone else, even if there are consequences towards you?	.96			2.18(1.21)	0.41**	0.41**	0.51**	0.58**	0.77**				
7	Can you change the online behaviour of someone else, even if there are consequences towards you?	.87			2.07(1.20)	0.40**	0.43**	0.49**	0.59**	0.76**	0.86**			
8	Can a character from a video game have realistic behaviour?	1.0	1.00	.71	2.30(1.25)	0.39**	0.34**	0.37**	0.45**	0.48**	0.52**	0.61**		
9	Can a character from a video game present human-like behaviour?	.64			2.30(1.31)	0.27**	0.28**	0.33**	0.35**	0.36**	0.37**	0.46**	0.68**	
% of Explained Variance 84%														
Note: **p <0.01														

		Session 2							
Dialogue		12		24		36		Case 2	
Player Type		Winners	Losers	Winners	Losers	Winners	Losers		
AI Agents	Tatiana	Neutral	Victim					Neutral	
	Nando			Positive	Aggressor	Positive	Aggressor		
	Manuela	Positive	Aggressor						
	Abel			Neutral	Neutral			Neutral	
	Carmen							Aggressor	
	Rui	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral		
	Jorge			Positive	Neutral			Neutral	
	Helder			Neutral	Neutral				
	Patrícia					Neutral	Neutral		
	Isabel							Aggressor	
	Estrela					Positive	Neutral	Victim	
	Samuel					Neutral	Neutral		

Table B.2: Character dialogues with alternative roles during session 2. Roles in bold are different for the respective character, depending on the player's scores on previous sessions.

		Session 3											
Dialogue		4		19		27		28		41		Case 3	
Player Type		Winners	Losers	Winners	Losers	Winners	Losers	Winners	Losers	Winners	Losers	Winners	Losers
AI Agents	Tatiana	Neutral	Neutral	Victim	Victim	Positive		Victim	Victim	Victim	Victim		
	Nando	Positive	Aggressor									Positive	Aggressor
	Manuela												
	Abel											Neutral	Neutral
	Carmen			Positive	-	Positive	-	Positive	-	Positive	-		
	Rui	Neutral	Neutral	Neutral	Neutral							Neutral	Neutral
	Jorge												
	Helder					Neutral	Neutral					Aggressor	Aggressor
	Patrícia			Aggressor	Aggressor	Aggressor	Aggressor	Aggressor	Aggressor	Aggressor	Aggressor	Aggressor	Aggressor
	Isabel	Positive	Positive	Positive	-					Positive	-	Positive	Positive
	Estrela			Positive	Positive	Positive	Positive	Positive	Positive				
	Samuel	Positive	Positive	Positive	Positive	Positive	Aggressor	Positive	Positive			Victim	Victim

Table B.3: Character dialogues with alternative roles during session 3. Roles in bold are different for the respective character, depending on the player's scores on previous sessions.

		Session 4				
Dialogue		23		34		Case 4
Player Type		Winners	Losers	Winners	Losers	
AI Agents	Tatiana					
	Nando			Positive	-	Neutral
	Manuela	Positive	-			
	Abel	Victim	Victim	Neutral	Neutral	Victim
	Carmen					
	Rui					
	Jorge					Positive
	Helder	Aggressor	Aggressor	Aggressor	Aggressor	Aggressor
	Patrícia					Victim
	Isabel					Positive
	Estrela					Positive
	Samuel					Positive

Table B.4: Character dialogues with alternative roles during session 4. Roles in bold are different for the respective character, depending on the player's scores on previous sessions.

Theme	Category	Answers / Category	Percentage / Category	Code	Answers / Code	Percentage / Code
Most recurring themes when answering to "Why do you think the changes in the aggressor's behaviour were/were not due to your actions?"	Cyberbully Self-improvement	97	30.22%	Improve / having good behaviour	25	25.77%
				Stop misbehaving	18	18.56%
				Apologizing	14	14.43%
				Reflect and comprehend	10	10.31%
				Demonstrate change	8	8.25%
				Being good	7	7.22%
				Being calm	5	5.15%
				Making up	5	5.15%
				Being adequate	2	2.06%
				Being modest	1	1.03%
				Being respectful	1	1.03%
				Learning a lesson	1	1.03%
				Total	97	100.00%
	Bystander Effect / Lack of change	52	16.20%	Passiveness / Silence / Not interfering	21	40.38%
				Being neutral / normal	13	25.00%
				Ineffective / Non-important	9	17.31%
				Indifference	6	11.54%
				Unchanged behaviour	3	5.77%
				Total	52	100.00%
	Bystander Prosocial Intervention	38	11.84%	Influence	16	42.11%
				Defend	5	13.16%
				Communicate	3	7.89%
				Co-exist	3	7.89%
				Act upon	2	5.26%
				Help	2	5.26%
				Doing what's right	2	5.26%
				Advice	1	2.63%
				Being kind	1	2.63%
				Worry about others	1	2.63%
				Straightforwardness	1	2.63%
				Listen	1	2.63%
				Total	38	100.00%
	Negative emotions towards cyberbullying situations	29	9.03%	Wrong	7	24.14%
				Bad / Evil	6	20.69%
				Aggressive	5	17.24%
				Dislike	3	10.34%
				Unjust	3	10.34%
				Inadequate	1	3.45%
				Unnecessary	1	3.45%
				Weird	1	3.45%
				Horrible	1	3.45%
				Offensive	1	3.45%
				Total	29	100.00%
	Internet / Social Network	28	8.72%	Comments / Posts / Chat	15	53.57%
				Online / Internet	5	17.86%
				Social Network	3	10.71%
				Cyberbullying	2	7.14%
				Personal matters	2	7.14%
				Game	1	3.57%
				Total	28	100.00%
	Cyberbullying Behaviour	25	7.79%	Being rude	7	28.00%
				Make fun of others	5	20.00%
				Insult	3	12.00%
				Provoke	3	12.00%
				Threat	2	8.00%
				Attack	1	4.00%
				Acuse	1	4.00%
				Brutality	1	4.00%
				Criticize	1	4.00%
				Judge	1	4.00%
				Total	25	100.00%

Table B.5: Qualitative Analysis (Part 1)

Theme	Category	Answers / Category	Percentage / Category	Code	Answers / Code	Percentage / Code
Most recurring themes when answering to "Why do you think the changes in the aggressor's behaviour were/were not due to your actions?"	Positive concepts related to stopping bullying	24	7.48%	Friendship	9	37.50%
				Sympathy	7	29.17%
				Education	4	16.67%
				Control	1	4.17%
				Funny	1	4.17%
				Respect	1	4.17%
				Solidarity	1	4.17%
				Total	24	100.00%
	Ways of embracing differences	14	4.36%	Independence / Being independent	5	35.71%
				Respect different opinions	4	28.57%
				Respecting other people's feelings	3	21.43%
				Feeling good in your own body	1	7.14%
				Ignore what other people say	1	7.14%
				Total	14	100.00%
	Motives behind cyberbullying behaviour	14	4.36%	Interpersonal conflicts	4	28.57%
				Non-intentional / Unconscious	2	14.29%
				Lack of Improvement (Maintaining bad attitudes)	2	14.29%
				Discussion	1	7.14%
				Impatience	1	7.14%
				Selfishness	1	7.14%
				Impulsiveness	1	7.14%
				Indecency	1	7.14%
				Anger	1	7.14%
				Total	14	100.00%
	Total	321				

Table B.6: Qualitative Analysis (Part 2)

Category	Meaning
Cyberbully Self-improvement	Refers to the perceived changes that showed the cyberbully's behavioural improvement
Bystander Effect / Lack of change	Refers to observations that try to justify being a bystander instead of acting pro-socially, as well as doubting self-efficacy
Bystander Prosocial Intervention	Refers to how bystanders should intervene in cyberbullying situations in a prosocial manner
Negative emotions towards cyberbullying situations	Refers to the emotions caused by, or that lead to, cyberbullying situations
Internet / Social Network	Refers to the usage of technology within the context of the game and cyberbullying
Cyberbullying Behaviour	Refers to behaviours associated with cyberbullying
Positive concepts related to stopping bullying	Refers to positive concepts associated with the changes that happened in the scenario
Ways of embracing differences	Refers to how one can respect differences without engaging in bullying behaviour
Motives behind cyberbullying behaviour	Refers to what can lead to cyberbullying behaviour

Table B.7: Qualitative Analysis - Category Operationalization