



HAL
open science

Virtual agents' perceived intelligence and believability in video games: exploring the relation between design factors and players' expectations

Rémi Poivet

► To cite this version:

Rémi Poivet. Virtual agents' perceived intelligence and believability in video games: exploring the relation between design factors and players' expectations. Human-Computer Interaction [cs.HC]. Sorbonne Université, 2023. English. NNT : 2023SORUS524 . tel-04478017

HAL Id: tel-04478017

<https://theses.hal.science/tel-04478017v1>

Submitted on 26 Feb 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Ecole Doctorale 158 - Cerveau, Cognition, Comportement

Virtual agents' perceived intelligence and believability in video games: Exploring the relation between design factors and players' expectations

by

Remi Poivet

A Doctoral Thesis in Cognitive Science

under the supervision of Dr. Malika Auvray, Dr. Catherine Pelachaud and Siva Veloumourougane

Approved Dissertation Committee

Dr. Christophe Lopez Reviewer	Aix-Marseille University Senior Researcher
Dr. Rachel McDonnell Reviewer	Trinity College Dublin Associate Professor
Dr. Jean-Marie Burkhardt Examiner	Gustave Eiffel University Senior Researcher
Dr. Ludovic Hoyet Examiner	INRIA Rennes Researcher
Dr. Rui Prada Examiner	University of Lisbon Associate Professor

Date of Defense: December 7, 2023

Acknowledgements

This dissertation represents over three years of academic and personal effort devoted to the significance of designing video game characters that consider players' perspectives. The work presented here has been made possible through the collaboration between Sorbonne Université, ISIR – Institut des Systèmes Intelligents et de Robotiques, and Ubisoft.

First and foremost, I am grateful to the committee members for agreeing to evaluate my PhD work and for dedicating their time to this process. The discussions I had with them were enriching and interesting. Therefore, I am thankful for their valuable input during the PhD defense and for their warm feedback regarding this manuscript.

Next, I wish to express my gratitude to my three supervisors: Dr. Malika Auvray and Dr. Catherine Pelachaud from Sorbonne Université, and Siva Veloumourougane from Ubisoft. Pursuing a PhD is a challenging journey, especially in an industry-applied context, compounded by the additional difficulties posed by the 2020 pandemic. However, the manuscript stands as evidence that we successfully overcame these challenges. Throughout my journey, I have learned alongside you, and I am immensely grateful. Whether in drafting, writing, or implementing the methodology for conducting experiments, I extend my thanks to Malika and Catherine for their patience and assistance during these stages. The practical implementation of the experiments, facilitated by discussions with various teams at Ubisoft, would not have been as seamless without Siva's guidance. Together, through these works and projects, we aimed to apply a user-centered approach that combine qualitative and quantitative data to support production teams, which I believe was a brilliant idea.

I am also grateful to everyone at ISIR, particularly the members of the PIROS team, who have been part of my journey over the years. The team's organized meetings significantly contributed to my scientific and academic development, and I am certain of their positive impact on my work. Additionally, I express my sincere gratitude to all the employees at Ubisoft who closely or remotely followed and supported my work, both within the French studios and in Canada. I hold a special

appreciation for the analytical teams, particularly for their essential support in assisting production teams and their unwavering help in successfully executing my research projects.

In conclusion, I wish to express my gratitude to my friends and family (and my dog). Thank you for believing in me and this project over the years, despite the difficulties I might have faced. I appreciated your presence, support, and companionship throughout this journey. I am grateful for your interest in my research topic, even if some of you are not gamers. Your attention and assistance have been invaluable, from proofreading my pending articles for submission to attempting to grasp the concepts of 'believability' and 'perceived intelligence.' I am thankful for your willingness to listen to my methodology presentations, even without having requested them. Your unwavering support meant a lot to me.

Résumé

Les personnages non joueurs (PNJ) sont une sous-catégorie d'agents virtuels qui peuplent les jeux vidéo. La manière dont ils sont perçus par les joueurs est essentielle pour l'expérience de jeu. En particulier, la perception d'intelligence et de crédibilité des PNJ contribue grandement à créer une expérience engageante et immersive. Ces deux aspects sont influencés par les attentes des joueurs lors de leur interaction avec les PNJ. Cependant, la connaissance de la manière dont les joueurs évaluent les PNJ est insuffisante, ce qui pourrait expliquer l'écart entre les attentes des joueurs et leur évaluation selon les facteurs de la conception des PNJ. L'objectif de cette thèse est d'avancer dans la compréhension de l'évaluation des PNJ par les joueurs en fonction des facteurs de conception des PNJ et des attentes des joueurs. Notre hypothèse suppose que renforcer la cohérence entre la conception des PNJ et les attentes des joueurs conduirait à une amélioration de leur intelligence et de leur crédibilité. Trois études ont été menées pour examiner l'évaluation par les joueurs de l'intelligence et de la crédibilité des PNJ. La première étude portait sur l'évaluation par les joueurs de l'hostilité des PNJ dans un jeu de tir militaire, examinant les paramètres de conception de l'apparence, des comportements et de leur interaction. Les résultats ont indiqué que les comportements des PNJ étaient le seul facteur significatif pour transmettre l'hostilité, quelle que soit leur apparence. La deuxième étude a manipulé l'apparence et les comportements des PNJ dans un jeu de tir militaire pour étudier l'impact de la violation des attentes des joueurs sur l'expérience de jeu, l'évaluation de l'intelligence perçue et de la crédibilité des PNJ. Cette recherche a impliqué des mesures comportementales et physiologiques et a souligné l'importance des facteurs de conception cohérents des PNJ dans l'expérience de jeu. De plus, la manipulation des facteurs de conception des PNJ a induit des changements significatifs dans les réponses physiologiques et les comportements des joueurs dans le jeu, suggérant l'émergence de nouvelles formes d'attentes. La troisième étude a porté sur l'expérience des joueurs, l'évaluation de l'intelligence perçue et de la crédibilité des PNJ lors d'interactions verbales. La recherche a impliqué la manipulation du rôle explicite des PNJ dans le récit et du style de communication à l'aide d'une méthode du «Wizard of Oz». L'évaluation de l'intelligence et de la crédibilité des joueurs a été principalement influencée par le rôle explicite des PNJ, tandis que le style de communication des PNJ a eu un impact sur les comportements des joueurs lors des conversations.

Tout au long de ces études, les facteurs de conception des PNJ ont été manipulés pour étudier les expériences et les évaluations des joueurs en matière d'intelligence perçue et de crédibilité. Les différentes découvertes de la recherche ont souligné l'impact significatif de la cohérence entre les facteurs de conception et les attentes des joueurs sur l'expérience de jeu. De plus, les résultats ont souligné l'importance des attentes des joueurs et l'influence conséquente des facteurs de conception en tant que rétroaction sur l'évaluation par les joueurs de l'intelligence perçue et de la crédibilité. La thèse se conclut par des recommandations pour les concepteurs de jeux et la manière dont ils peuvent utiliser des expériences comportementales pour créer des PNJ engageants dans leurs jeux.

Abstract

Non-player characters (NPCs) are a subtype of virtual agents that populate video games. How they are perceived by players is central to gaming experience. In particular, NPCs' perceived intelligence and believability highly contribute to creating an engaging and immersive experience. These two aspects are influenced by players' expectations during their interaction with NPCs. However, the knowledge of how players evaluate NPCs is insufficient, which might explain the actual gap between players' expectations and their actual evaluation of NPCs' design. The aim of this thesis is to advance knowledge on players' evaluation of NPCs as a function of their design factors and players' expectations. Our hypothesis is that enhancing coherence between NPCs' design and players' expectation would lead to enhanced perceived intelligence and believability. Three studies were conducted to investigate players' evaluation of NPCs' intelligence and believability. The first study focused on players' evaluation of NPCs' hostility in a military shooter game, examining the design parameters of appearance and behaviors and their relationship. The results indicated NPCs' behaviors were the only significant factor in conveying hostility, regardless of appearance. The second study manipulated NPCs' appearance and behaviors in a military shooter game to investigate the impact of violating players' expectations on players' gaming experience and evaluation of NPCs' intelligence and believability. This research involved behavioral and physiological measures and highlighted the importance of coherent NPCs' design factors in shaping gaming experience. Additionally, the manipulation of NPCs' design factors induced significant changes in players' physiological responses and behaviors in the game, suggesting the emergence of new forms of expectations. The verbal modality was not addressed in the first two studies. The third study was conducted on players' experience and evaluation of NPCs' perceived intelligence and believability during verbal interactions. The research involved the manipulation of NPCs' explicit role in the narrative and the communication style using a Wizard of Oz method. Players' evaluation of intelligence and believability was primarily influenced by the explicit role of NPCs while NPCs' communication style impacted players' behaviors during conversations. Throughout these studies, the design factors of NPCs were manipulated to investigate players' experiences and evaluations of intelligence and believability. The different

research findings highlighted the significant impact of coherence between the design factors and players' expectations on the gaming experience. Furthermore, the results underlined the importance of players' expectations and the consequential influence of design factors as feedback on players' evaluation of perceived intelligence and believability. The thesis concludes with recommendations for game designers and how those can leverage behavioral experiments to create impactful NPCs in their games.

Contents

CHAPTER 1: General Introduction	12
1. Players' interactions with virtual agents	15
1.1 NPCs as a subtype of virtual agents: definition and characteristics	15
1.1.1 NPCs' appearance	17
1.1.2 NPCs' behaviors	20
1.2 Players' evaluation of NPCs	21
1.2.1 Human assessment of virtual agents	22
1.2.2 Players' assessment of NPCs	27
1.3 Players' expectations about NPCs	29
1.3.1 Definition of players' expectations	29
1.3.2 State of the art on players' expectations about NPCs	31
1.4 Violation of players' expectations	32
1.4.1 Definition of violation of players' expectations	33
2. Believability	35
2.1 Definition of NPCs' believability	35
2.2 State of the art on believable NPCs	38
3. Perceived Intelligence	41
3.1 Definition of NPCs' perceived intelligence	41
3.2 State of the art on perceived intelligence	43
4. Research Questions and Hypotheses	44
CHAPTER 2: Evaluation of virtual agents' hostility in video games	46
1. Introduction	46
2. Background	48
2.1 Definition of hostility and its characteristics	48
2.2 Threatful appearance	49
2.3 Aggressive behaviors	50
3. Personality traits	51
3.1 Morality	52

3.2 Likability	53
3.3 Perceived intelligence.....	53
4. Experiments on hostile NPCs	54
5. General Methods	55
Participants	55
Procedure and materials	55
6. Experiment 1: Threat of NPCs’ appearances.....	57
Participants, stimuli, and procedure	57
Results and discussion.....	57
7. Experiment 2: Aggressiveness of NPCs’ behaviors.....	60
Participants, stimuli, and procedure	60
Results and discussion.....	61
8. Experiment 3: Contribution of appearance and behavior to the perception of hostility ..	63
Participants, stimuli, and procedure	63
Results and discussion.....	65
9. General Discussion	66
9.1 Summary	66
9.2 Threat of the appearance	67
9.3 Aggressiveness of the behavior	68
9.4 Hostility within NPCs	70
10. Limitations and future research	71
11. Conclusion.....	72
References	73
CHAPTER 3: Breaking expectations in video games: the role of coherence and consistency in players’ experience.....	77
1. Introduction	78
2. Background.....	79
Appearance.....	80
Behaviors.....	80
2.2 Breaking players’ expectations.....	81
How do we form and break expectations?	81
The impact of breaking expectations.....	82

3. How to measure perceived intelligence and believability of NPCs?83

4. Experiments85

5. General Methods85

 Materials and Procedure 85

 Creation of NPCs’ design..... 86

 Videos for participants’ evaluation 87

6. Experiment 1.....88

 Participants 88

 Procedure and materials 88

 Results and discussion..... 88

7. Experiment 2.....90

 Participants 90

 Procedure and materials 91

 Game sessions 91

 Questionnaires 92

 Behavioral measures..... 93

 Physiological measures 93

 Results and discussion..... 93

 Questionnaires’ analysis..... 93

 Behavioral analysis..... 94

 Physiological analysis 95

 Videos’ evaluation..... 97

8. General Discussion100

 8.1 Summary 100

 8.2 Coherence manipulation..... 100

 8.3 Consistency manipulation 101

 Game experience when manipulating coherence and consistency 101

 Players’ expectation of NPCs’ intelligence and believability 103

9. Limits and future research104

10. Conclusion.....105

References106

CHAPTER 4: The influence of conversational agents’ role and communication style on user experience.....	109
1. Introduction	110
2. Background.....	110
3. Scales of perceived intelligence and believability	113
4. General methods and procedure.....	115
Participants	116
5. Materials	116
6. Results	122
7. Discussion.....	124
8. Conclusion.....	128
13. References	130
CHAPTER 5: General Discussion	133
Insights on players’ evaluation of NPCs’ hostility.....	133
Insights on players’ evaluation of unexpected NPCs	136
Insights on players’ evaluation of Conversational Agents	140
Conclusion and opportunities for future works	141
REFERENCES	143

CHAPTER 1: General Introduction

Non-player characters (NPCs) are a subtype of virtual agents that participate in the video game experience. This experience relies on the interactions between the main character of a video game and the virtual environment, including its agents. As such, human players controlling the main character engage with NPCs that exhibit behaviors aligned with their narrative roles. NPCs' roles, like being friendly or hostile, are defined by their social standing in the storyline, which subsequently shapes their design. The term «design of NPCs» encompasses the parameters manipulated by game designers to craft these characters. NPCs' parameters involve appearance and behaviors. Appearance refers to the virtual embodiment perceived by players in the game environment, while behaviors involve predefined sequences of actions enabling interactions with the main character and their surroundings. Hence, game designers' choices regarding NPCs' parameters significantly influence the outcomes and quality of players' interactions, underscoring the importance of considering players' evaluations when crafting these characters.

During their interactions with NPCs, players evaluate both their appearance and behaviors, leading to the attribution of human traits like morality and likability (Lee & Heeter, 2012; Pradantyo et al., 2021). This attribution aims to shape players' comprehension of the narration, thereby establishing the NPCs' social position within the storyline. Thus, players' evaluation of NPCs always occurs and contributes to their engagement in the game experience. Two crucial dimensions influencing players' engagement with NPCs are their perceived intelligence and believability. A positive assessment along these dimensions can enhance players' willingness to interact and overall enjoyment of the gaming experience (Warpefelt, 2016; Moussawi et al., 2021). Perceived intelligence refers to players' evaluation of NPCs' proficiency in fulfilling their roles within the narrative (Bartneck et al., 2009). This evaluation entails understanding NPCs' actions accurately and assessing the outcomes of their actions. On the other hand, NPCs' believability pertains to the disparity between players' expectations and their observations, with greater disparities leading to lower believability evaluations (Lee & Heeter, 2008). Both dimensions heavily depend on players' expectations, which are shaped by their stereotypes and knowledge about video games (Warpefelt, 2015). Thus, exploring how players' expectations can be influenced and their connection to NPCs' design would enhance our understanding of human interaction with virtual agents and their

application in video games. The overall goal of the thesis is to develop a comprehensive understanding of the relationship between players' expectations and NPCs' design. Since perceived intelligence and believability significantly affect player engagement and depend on their expectations of NPCs, maximizing alignment between player expectations and NPCs' design is essential. However, within the context of video games as narrative experiences, game designers can strategically use NPCs' design as a tool to introduce surprises and rhythm into the storyline by manipulating their parameters to disrupt players expectations. Yet, the potential impact of such design manipulation on players' evaluations of NPCs remains unclear. The thesis hypothesis posits that NPCs designed to correspond to players' expectations will result in heightened perceived intelligence and believability. In other words, when NPCs' design aligns coherently with their expectations, players are expected to perceive them as more intelligent and believable, ultimately enhancing their overall engagement with the game.

To this aim, three studies were conducted focusing on players' evaluation of NPCs. The primary focus of the first study was to investigate how players evaluate NPCs' hostility within the context of video games, as interactions of this nature often constitute a fundamental aspect of the gaming experience. This study entails three experiments that measured players' evaluation of the threat of NPCs' appearance, the aggressiveness of their behaviors and the hostility of NPCs and their impact on personality traits ascription. This investigation shed light on the role of NPCs' behaviors as a primary factor to convey hostility and the various personality traits that are conveyed by their design's parameters. The second study analyzed how players' expectations can be influenced by their interactions with NPCs in a military shooter game and how violating players' expectations influences their evaluations of NPCs' perceived intelligence and believability. The results highlighted the pivotal role played by the consistency of NPCs' design parameters in enabling players to develop accurate expectations about their behavior. Additionally, it revealed that players evaluate NPCs more positively if their designs' parameters align to players' initial expectations. This experiment provided a deeper insight into the intricate relationship between NPCs' design and players' expectations, shedding light not only on their combined effect on the game experience but also on the influential role of players' expectations in evaluating NPCs' perceived intelligence and believability.

The third study explored another modality of interaction: verbal interaction with conversational

agents in narrative experience, considering that NPCs could potentially engage in such interactions using advanced language models. The aim of this third study was to analyze how players' evaluation of perceived intelligence and believability is influenced by the design's parameters during a verbal interaction using a Wizard of Oz method. The results emphasized the significant impact of the conversational agent's role on players' assessments of perceived intelligence and believability. Simultaneously, the communication style influenced players' verbal behaviors during the conversation. This experiment provided insights into the potential effects of players' expectations regarding NPCs during verbal interactions and suggested considering these expectations when determining NPCs' parameters.

In conclusion, the three studies collectively provide a deeper understanding of the intricate connection between players' expectations and NPCs' design, specifically in relation to their assessments of perceived intelligence and believability.

The manuscript is structured into five chapters. The first chapter is the general introduction of the topic, providing definitions of believability and perceived intelligence and raising the main research questions. Chapters 2, 3 and 4 are the experimental studies mentioned above: Chapters 2 and 4 are submitted papers under review, chapter 3 is a paper in preparation. Finally, chapter 5 engages in an in-depth discussion of the primary findings, as well as address limitations and propose future directions for video game design.

Paper references:

Poivet, R., de Lagarde, A., Pelachaud, C., & Auvray, M. (under review). Evaluation of virtual agents' hostility in video games. *IEEE Transaction on affective Computing*.

Poivet, R., Pelachaud, C., & Auvray, M. (in preparation). Breaking expectations: the role of coherence and consistency on players' experience.

Poivet, R., Lopez Malet, M., Pelachaud, C., & Auvray, M. (under review). The influence of conversational agents' role and communication style on user experience. *Frontiers in Psychology Cognitive Science*.

1. Players' interactions with virtual agents

In this section, virtual agents in video games are introduced, and a definition of a subset of them, namely non-player characters (NPCs), is provided. In Section 1.1, the role and function of NPCs are presented, followed by an examination of their design parameters in the subsequent subsections. Sections 1.2 explore the assessment of virtual agents' parameters by human users and its impact on the attribution of human qualities to NPCs. Sections 1.3 and 1.4 define and analyze players' expectations and their potential violations.

1.1 NPCs as a subtype of virtual agents: definition and characteristics

Virtual agents are computer-generated entities or characters, programmed to simulate life-like interactions, behavior, and responses within virtual environments. They encompass a wide range of forms, from simple conversational agents (chatbots) to complex characters in video games (Lugrin, 2021). In video games, the virtual character controlled by the human player, referred to as their avatar or the main character of the narrative, interacts with virtual agents that populate the virtual environment. This subset of virtual agents is known as NPCs and assumes social roles and functions to contribute to the narrative and gaming experience (Johansson et al., 2014). For example, in role-playing games (RPG), the main character is often on a journey that involves encountering both friendly and hostile NPCs that contribute to the narrative (see figure 1). The NPCs that are encountered play distinct roles, such as companions on the main character's journey, shopkeepers, quest givers, or even enemies (Warpefelt, 2016). Consequently, these NPCs' social roles are defined by their functions within the storyline, and these roles shape how they interact with the main character, thereby delivering the intended game experience. Therefore, hostile and friendly NPCs have different functions that shape their interactions and their outcomes (Giusti et al., 2012). One of their functions is to influence the game's level of challenge through their interaction with the main character. For example, a shopkeeper would not exhibit aggressive behavior towards the main character, as their primary function in the game experience is to provide items that help the main character in pursuing their journey. On the other hand, the enemies' function is to be at the core of the game experience as they contribute the most to the interactions. For instance, in a military shooter game, players anticipate immersing themselves in a war simulation, where the essence of the game experience revolves around defeating NPCs. In this

scenario, NPCs adopt aggressive behaviors towards the main character to convey the desired game experience.



Figure 1: NPCs' roles in the game experience as illustrated in the RPG *Skyrim* (Bethesda, 2011); on the left, an enemy is depicted, while on the right, a companion is shown.

When crafting NPCs for a specific role, game designers configure parameters to establish a recognizable design in the game experience. This process encompasses defining appearances and behaviors to guide players in recognizing the NPCs' role within the narrative. Thus, the association between the two parameters remains consistent throughout the gaming experience to convey a cohesive NPCs' design (Giusti et al., 2012). For instance, crafting enemies' roles involve linking a specific appearance to aggressive behaviors to allow players to anticipate that NPCs with a similar appearance will likely endorse a similar role in future interactions (Orhberg, 2019). Furthermore, within a specific role, game designers create various subtypes to elevate the game experience. For enemies, these types are characterized by their distinct appearances and unique behaviors, aiming to elevate the game's challenge by introducing diverse forms of hostile interaction (Giusti et al., 2012). These roles' subtypes are referred to as archetypes, contributing to a multifaceted and captivating gaming experience.

In summary, NPCs' roles are pivotal for how they are portrayed in the narrative and contribute to players' immersion (Warpefelt, 2015; Zibrek et al., 2018). These roles establish their position in the story and their functions, which collectively define their purpose in the gaming experience. As an integral part of both the game and narrative, NPCs play a fundamental role in engaging players within the video game (Hamdy & King, 2017). Consequently, the design of NPCs significantly shapes the overall gaming experience, prompting the need to consider its impact on players' expectations and assessment of NPCs. The following sections will delve into presenting NPCs' appearances and behaviors, as well as exploring the interaction between these two parameters.

1.1.1 NPCs' appearance

Virtual agents' appearance refers to their virtual embodiment, which users perceive within the virtual environment (McDonnell & Mutlu, 2021). The appearance of virtual agents is a crucial element of their design, as it effectively communicates information to users (Cafaro et al., 2016) and influences users' engagement (Shiban et al., 2015). For instance, embodied conversational agents with the purpose of providing medical advice have a stronger influence on users' inclination to follow suggestions when they adopt the appearance of a doctor (Dai & Macdorman, 2018). In this situation, developers manipulate the visual cues of the virtual agent's appearance to activate stereotypes associated with medicine in the user's mind. These visual cues can pertain to agents' shapes (Fleming et al., 2016), facial features (Ferstl & McDonnell, 2018), or items associated with stereotypes with the aim of enhancing users' understanding and engagement towards the agents (Veletsianos, 2010).

In video games, the appearance of NPCs spans a wide spectrum, encompassing everything from pixel art to realistic shapes (see figure 2). These appearances can exist in either two or three dimensions, depending on the type of video game in which they are featured. Designing NPCs' appearance follows a similar approach to virtual agents, as developers manipulate visual cues to convey NPCs' social role and function in the narrative (Warpefelt, 2015). These cues encompass NPCs' shape, their clothes, face and items they exhibit. Warpefelt (2015) conducted an analysis of players' expectations of NPCs' role in the game based on their appearances.



Figure 2: Examples of the various possibilities of rendering in video games. From top left to bottom right: *Celeste* (Extremely OK Games, 2018); *Cyberpunk 2077* (CD Projekt RED, 2020); *Call of duty: Modern Warfare 3* (Activision, 2011); *Have A Nice Death* (Gearbox Software, 2022)

Notably, players rely on knowledge and stereotypes associated with items and clothes to identify NPCs' roles in the game. For instance, a shopkeeper is typically positioned behind a counter, while a guard is expected to wear a uniform and carry a weapon. Furthermore, in role-playing games, players encounter multiple archetypes of shopkeepers, each defined by the items they have for sale (see figure 3). This often results in them sharing highly similar appearances, facilitating players' ability to identify them even in bustling environments. However, developers incorporate variations in stereotypical elements of their appearances, signifying the unique items they offer. Take, for example, the *Assassin's Creed* series (Ubisoft), where shopkeepers are consistently placed behind counters and differentiated through appearances that align with the items they have on offer. In *Assassin's Creed II* (Ubisoft, 2009), medical vendors adopt an appearance reminiscent of doctors from the plague era, efficiently conveying their wares to players by activating their historical knowledge (see figure 4).



Figure 3: NPCs with the role of shopkeepers with different appearances to convey the items or services provided in the Western action-adventure game *Red dead redemption 2* (2018); On the left is a barber, while on the right is a General Store seller.



Figure 4: A medical vendor from *Assassin's Creed II* (Ubisoft, 2009)

In the context of enemies, their appearance impacts players' expectations of the interaction, as certain visual cues provide valuable information about their abilities. For example, in military shooter games, the size of their weapons or their level of protection can offer insights into their level of threat or resilience during interactions. Additionally, players' recognition of enemies' weapons allows them to predict NPCs' mobility capabilities. For instance, players can anticipate that enemies armed with shotguns will attempt to get closer to them, unlike those equipped with rifles, given that shotguns are designed for close-range combat. Overall, NPCs' appearance shapes

players' expectations about their interaction with them. Subsequently, NPCs' behaviors enable these interactions while also confirming or challenging these expectations.

1.1.2 NPCs' behaviors

The behaviors displayed by virtual agents encompass both verbal and non-verbal actions aimed at conveying a sense of animacy during their interactions with users. The purpose of these behaviors is to enhance users' engagement with the agents through a sense of lifelikeness (Brusk, 2014). Verbal behaviors aim at replicating natural human language and include activities such as text generation and comprehension (Adamopoulou & Moussiades, 2020). To assess the impact of verbal behaviors, certain studies manipulated linguistic parameters in the conversation style during interactions with virtual agents, and some used the Wizard of Oz method (Bell et al., 2019; Mitchell & Mamykina, 2021). In this approach, researchers informed participants that they are conversing with a conversational agent, while it is actually a human operator generating the text. This method allows for precise control over text generation and the evaluation of the impact of specific linguistic parameters on users' experiences. Non-verbal behaviors encompass a broad spectrum of actions, from hand and facial gestures (Bavelas et al., 2014) to more intricate actions such as moving and grasping objects within the virtual environment (Vosinakis & Panayiotopoulos, 2001). Studies focusing on nonverbal behaviors have highlighted their impact on users' attention and their role in enhancing the believability of interactions (Niewiadomski et al., 2010). NPCs possess intricate behaviors structured as sequences of predefined actions, enabling them to engage with their surroundings. As was previously mentioned, NPCs' roles dictate their interactions and consequently, the sequences of actions available to them. The selection of sequences of actions is designed to align with players' expectations influenced by NPCs' appearances. Hence, the range of behaviors NPCs can exhibit is constrained by their roles, meaning friendly NPCs cannot act aggressively towards the main character, as their purpose is to provide assistance. In the case of enemies, their sequences of actions contribute to the game's challenge as they act aggressively towards the main character. Therefore, the selection of the sequence of action delivers negative outcomes during and after the interaction (i.e., trying to defeat the main character). In shooter games, enemies are hierarchized by their archetypes, which are characterized by their physical abilities, closely tied to the weapons they carry (see figure 5). For example, enemies armed with shotguns exhibit distinct sequences of actions compared to those with sniper rifles when facing the

main character. In the former case, the weapon requires reducing the distance between them to reach the main character, while snipers follow the opposite strategy. Furthermore, game developers manipulate other parameters that can impact the challenge of the gaming experience, such as the speed of action execution or the occurrence of specific actions (e.g., seeking cover during conflicts with the main character in shooter games). The archetypes serve to make the game interaction more complex by introducing aggressive NPCs with diverse action sequences. The limitations on enemies' actions ensure consistent interaction throughout the game. For instance, players might expect NPCs carrying shotguns to approach them, given the restriction to that action. Therefore, through NPCs' design, players can accurately predict NPCs' behavior by linking their appearance—marked by visual cues that trigger stereotypes—to corresponding actions.



Figure 5: An example of two archetypes from the action-RPG *The Division* (Ubisoft, 2015); on the left is the archetype named ‘Crab,’ characterized by its approach of reaching the main character while using a shotgun, and on the right is the archetype called ‘Sniper,’ defined by its use of long-range weapons and tendency to stay far from the main character.

1.2 Players’ evaluation of NPCs

Given that humans naturally function as social beings, we have a tendency to project social qualities onto virtual agents (Reeves & Nass, 1996). This tendency is rooted in our innate social behaviors, where we engage with others and form expectations based on social cues. Understanding how users

evaluate virtual agents through this lens is crucial. It allows researchers to gain insights into how various parameters influence users' engagement with these virtual entities. By comprehending the factors that shape user perceptions of and interactions with virtual agents, their design and functionality can effectively be enhanced (Drennan et al., 2004). Users' engagement refers to the degree of involvement, interest, and interaction that they exhibit while interacting with a virtual agent (Oertel et al., 2020). It encompasses the extent to which users are absorbed, attentive, and emotionally involved, often resulting in sustained attention, active participation, and an overall positive experience (Peters et al., 2009). Engaged users are more likely to invest time, effort, and attention, leading to a deeper connection with the agent (Szafir & Mutlu, 2012). In video games, players can be captivated by the narrative through the storyline and characters, but the game experience relies on the main character interacting with NPCs, as they dynamically shape the virtual world, provide challenges, and contribute to immersive environments. Consequently, crafting NPCs that effectively engage players during their interaction stands as a crucial factor in delivering an immersive gaming experience. The subsequent sections will delve into human evaluation of virtual agents, followed by a discussion on players' evaluation of NPCs.

1.2.1 Human assessment of virtual agents

Designing virtual agents involves considering their appearance and the modality of their interaction, encompassing both verbal and nonverbal communication (Pelachaud et al., 2021). Consequently, research has been conducted to comprehend how design parameters impact users' evaluation and engagement with virtual agents. This subsection explores the effects of virtual agents' appearance and behaviors on user evaluation.

User evaluation of virtual agents' appearance

To ensure users' engagement, designers strive for appealing appearances, as it's widely acknowledged that aesthetics influence audience attention (McDonnell & Mutlu, 2021). Additionally, similar to human interactions, virtual agents are regarded as social entities, triggering stereotypes through visual cues in their appearance (Nass et al., 1997). This tendency leads users to attribute human traits such as warmth and competence to these agents (Dai & MacDorman, 2018).

Crafting virtual agents' appearance involves considering their intended role during user interactions. As was previously mentioned, designers use a metaphorical design approach where appearance aligns with stereotypes and common knowledge related to their function. For instance, embodied conversational agents offering medical advice have a greater impact on persuading users to follow health suggestions when resembling doctors (Dai & MacDorman, 2018). Thus, the context of interaction significantly influences virtual agents' appearance choices. Rendering decisions for virtual agents' appearance should correspond to their user's expectations. For instance, a cartoonish appearance might be less suitable for serious contexts like business meetings (Junuzovic et al., 2012) or medical training (Volante et al., 2016).

Exploring stylization's effects on user evaluation, McDonnell et al. (2012) conducted an experiment on virtual agents' appeal and trustworthiness (see figure 6). By adjusting the rendering style while maintaining consistent shape and motion in a virtual agent, they observed that the most realistic character had comparable appeal and trustworthiness ratings to the cartoon characters. However, characters with intermediate appearances encountered reduced appeal due to categorization difficulties (Saygin et al., 2012).

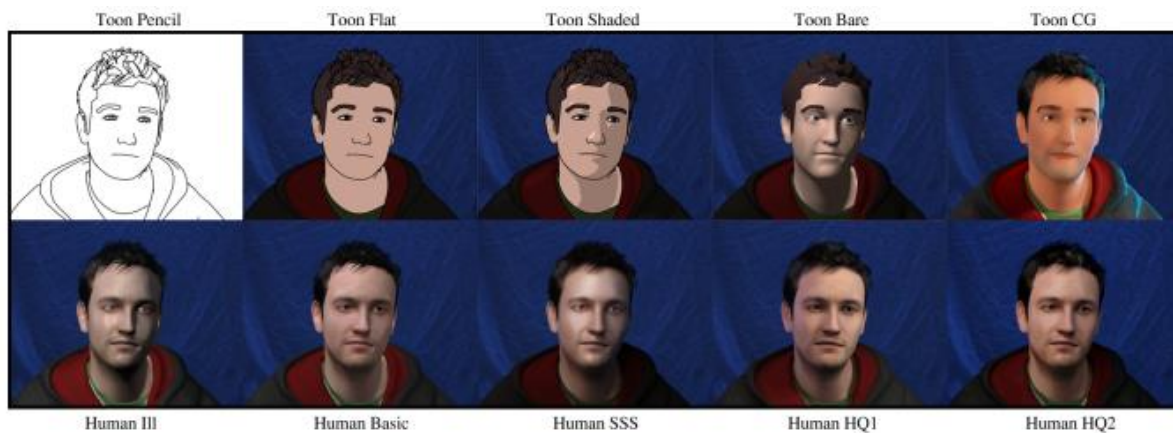


Figure 6: The different rendering style manipulated by McDonnell et al. (2012).

Moving beyond virtual embodiment stylization, facial and body features impact user evaluation. Since faces are the most attention-grabbing parameter in social interaction (Walker-Smith et al., 2013), the manipulation of facial features requires an assessment of their impact on the user

experience. Analyzing facial features and their influence reveals distinct effects of their size and localization. Notably, studies on facial proportions indicate that narrow eyes, in both abstract creatures (Ferstl et al., 2017) and realistic portrayals of virtual humans (Ferstl & McDonnell, 2018), are associated with heightened aggression and diminished trustworthiness in user judgment. Interestingly, wider faces don't decrease trustworthiness and are perceived as less aggressive than narrow faces, regardless of whether characters are realistic (Wang et al., 2013) or abstract (Ferstl et al., 2017). Accurate placement of facial features plays a significant role in perceived appeal. Green et al. (2008) reveals that not only proportions but also the positioning of facial parts can negatively affect perceived attractiveness, particularly for humanlike faces (see figure 7).



Figure 7: The manipulation of eye placements on 3D female virtual agents by Green et al. (2008). From left to right, the eye placements are stretched to display different spatial sizes and measure their effects on users' perception.

Additionally, ensuring consistency of realism across facial components is pivotal to maintaining appeal, as demonstrated by studies highlighting that mismatches of realism adversely affect perceived attractiveness (Burleigh et al., 2013; MacDorman et al., 2009). However, the maintenance of natural eye size deviations contributes to virtual agents' pleasantness evaluation (Seyama & Nagayama, 2007). Regarding their body representation, studies have explored how various styles ranging from minimalistic figures to photorealistic or anthropomorphized forms, influence users' evaluation of their actions. Research into body shape's impact reveals that virtual

agents' body shapes, especially female ones, significantly influence perceptions of gender and emotions, as seen in studies evaluating appeal, realism, and gender perception through morphs between realistic and stylized versions (McDonnell et al., 2009; Fleming et al., 2016) (see figure 8 and 9).

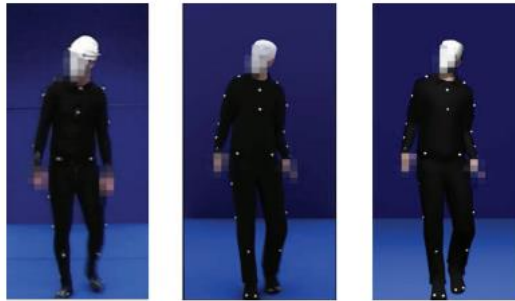


Fig. 1. Image taken from real video, high-resolution virtual male, and low-resolution virtual male.



Fig. 2. Image taken from wooden mannequin, toon, zombie, and points.

Figure 8: Investigation of motion perception based on different appearances by McDonnell et al. (2009).

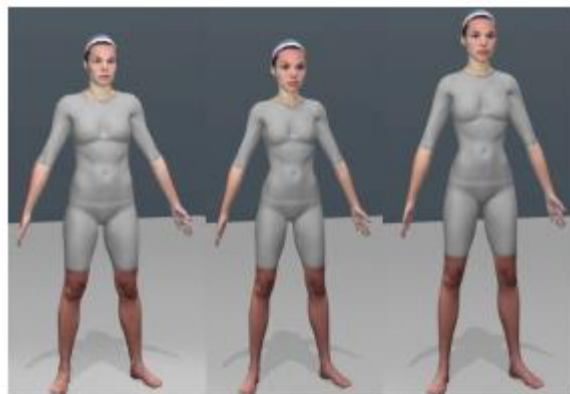


Figure 9: Analysis of appearance' stylization on users' ratings of female avatar's appeal by Fleming et al. (2016).

Consideration of users' specificities, such as gender stereotypes or cultural affiliations, was introduced in the design of virtual agents' appearances, as it can impact their evaluation. For example, Zibrek et al. (2015) examined gender bias in emotion perception concerning male and female virtual agents and observed a significant influence of established gender stereotypes on attributing emotions to virtual agents based on human interactions. Anger tends to be more associated with male virtual agents, while fear and sadness are less commonly attributed to them. A pronounced contrast effect was noted for anger, demonstrating heightened masculinity even when conveyed through a female agent. Additionally, Koda et al. (2010) explored cultural differences in emotion identification strategies between Japanese and Hungarian participants using virtual agents. Their findings revealed cultural disparities in emotion identification across various facial expressions. Japanese participants tended to focus more on the eye region when perceiving facial expressions, while Hungarian participants leaned towards interpreting emotions through the mouth region. Hence, gender stereotypes and cultural distinctions must be considered when designing virtual agents' appearance, as variations in interpreting facial expressions can differ among cultures, impacting the selection and manipulation of facial features.

User evaluation of virtual agents' behaviors

Virtual agents use nonverbal behaviors to capture attention and enhance engagement during interactions (Wang & Ruiz, 2021). These behaviors encompass motion abilities and facial expressions, deepening interactions, and establishing profound connections. Subsequently, the facial and bodily movements of virtual agents convey additional information beyond verbal communication, including emotional expressions (Ekman, 2004). Even subtle interface cues, such as virtual agents' gaze patterns, contribute to enhancing interactions. For instance, Fukayama et al. (2002) demonstrated that gaze patterns exhibited by agents' eyes evoke distinct impressions, revealing nuances through parameters like gaze duration and intensity. These simple patterns convey emotions such as likability, warmth, and potency. Building on this, Heylen et al. (2005) added the importance of appropriate gaze patterns, highlighting their impact on users' evaluation of agents' task performance, perceived trustworthiness, and warmth. These cues can be combined as multimodal behaviors to establish joint attention, as observed in pedagogical agents employing gestures and gaze cues to guide users' focus (Andre et al., 1996; Lester et al., 2000).

Moreover, a combination of various nonverbal behaviors enhances users' intimacy with virtual agents, as demonstrated by Kang et al. (2011). In their experiments, participants were interviewed by virtual agents expressing different nonverbal behaviors (e.g., head nods and tilts, pauses.). Researchers evaluated participants' level of intimacy based on the nature of the information disclosed in the interview (i.e., low, medium, and high intimacy). This evaluation revealed that certain nonverbal behaviors used by virtual agents, such as head tilts and pauses, efficiently convey a strong sense of high intimacy in users. Consequently, generating nonverbal behaviors can be optimized using strategies such as adaptation and imitation.

Virtual agents may exhibit imitative behaviors by reflecting users' actions and expressions during their interactions, responding with similar body postures, head movements, smiles, or laughter (Burgoon et al., 2007). This imitation contributes to building rapport, fostering affiliation, promoting engagement, and cultivating empathy, ultimately enhancing social interaction (Lakin & Chartrand, 2003). Additionally, adaptation mechanisms can enhance virtual agents' interactions (Biancardi et al., 2019b; Dermouche & Pelachaud, 2019). In that sense, machine learning can be used to create virtual agents that exhibit reactive behaviors to convey specific evaluation. In pursuit of this goal, researchers have employed a reinforcement learning approach to identify conversational strategies and behaviors that improve users' perception of warmth. Virtual agents that dynamically adapt their conversational and nonverbal strategies to elicit warmer evaluations during interactions optimize user engagement and their overall impression of the agents.

1.2.2 Players' assessment of NPCs

Players' evaluation of NPCs in video games constitutes a multifaceted process encompassing appearance and behaviors. Numerous studies investigated this dynamic, revealing player assessment of NPCs based on cues from their appearance and the outcome of their behavior. Concerning NPCs' appearance, Pradantyo et al. (2021) examined game designers' portrayal of video game antagonists and players' perception of their morality (see figure 10). Their research underscores that players discern moral attributes from visual cues, highlighting the link between NPCs' design and player perception. Additionally, Headleand et al. (2016) probed how NPCs' perceived identity influences players' protective behaviors towards them in a shooter game. Their experiment had players safeguard various NPCs from hostile entities while manipulating their appearances to convey human or robot forms. They also altered NPCs' gender within the human

appearance. Results indicated that players interact distinctively with human versus robot companions, favoring protection for humans.



Figure 10: One of the antagonists' appearance used by Pradantyo et al. (2021) to assess the effect of visual cues on players' morality ascription.

Additionally, players exhibited a preference for protecting female versions of the human appearance, suggesting gender stereotypes' influence on their decisions. Regarding gender's effect, Desai et al. (2016) compared players' perception of NPCs in story-based games. The results showed that there was no gender difference in the identification of NPCs' emotions. However, their ability to distinguish between emotions such as fear, anger, and happiness in NPCs of different genders varied. Consequently, NPCs' appearance impacts players' evaluations of NPCs and their behaviors. Focusing on NPCs' behaviors, research underscores their role in enhancing player evaluations and suggests the positive impact of adaptive behaviors. Paradedda et al. (2019) investigated the impact of nonverbal behaviors on players' evaluation of assertiveness and its influence on their game experience. In the experiments, participants assumed the role of a country's leader confronting an enemy threat. To defend their country, they made critical decisions with the

assistance of two NPCs endorsing the role of advisors. Researchers manipulated various parameters, including NPCs' voice pitch, speech rate, gaze direction, and posture, to convey different levels of assertiveness. Their results revealed that NPCs displaying low assertiveness levels led to increased engagement and heightened emotional responses. Normoyle et al. (2013) explored gaze patterns' influence on trust perception during verbal interactions with NPCs in RPGs, highlighting the importance of NPCs' eye contact in conveying trustworthiness. Furthermore, Chowanda et al. (2016) investigated how NPCs' improved social and emotional skills could enhance players' evaluation. They introduced the ERiSA Framework, emphasizing emotional expression, social interactions, and adaptive responsiveness. Participants engaged with NPCs crafted with this framework, with interactions recorded and questionnaires assessing personality, social relationships, emotional impact, and the overall experience. Results demonstrated that incorporating adaptive emotional and social attributes into NPCs enhances player experiences. Additionally, Dechant et al. (2022) unveiled the role of player personality traits in shaping emotional perceptions, verbal responses, and movement patterns during interactions with NPC. In conclusion, the above-mentioned studies collectively underscore the intricate player-NPC relationship, requiring a balance of visual cues from their appearance and behaviors to foster positive evaluations and enrich immersion. Understanding how players interpret cues and respond to behaviors empowers designers to craft compelling and relatable NPCs, enhancing the gaming experience.

1.3 Players' expectations about NPCs

When players encounter NPCs during their gaming experience, they have certain expectations that influence their evaluation and attitudes towards them. This section defines players' expectations and presents their implications in the gaming experience. Subsequently, the development of expectations in human interactions is then explored before addressing the violation of expectations in the following section.

1.3.1 Definition of players' expectations

In the gaming experience, one of the challenges players face is making effective decisions during their interactions with NPCs. Players' decisions to engage with NPCs heavily rely on their objectives in the game scenario combined with the detection of visual cues that convey NPCs' roles in the narrative (Linderoth, 2013). For instance, in Super Mario (Nintendo, 1985), the Goomba is

the first enemy encountered by Mario (see figure 11). Its slanting eyebrows convey its negative intentions, allowing players to engage appropriately.



Figure 11: Goomba is the first encountered character in *Super Mario bros.* (Nintendo, 1985)

As a result, the combination of visual cues from NPCs' appearances and players' video game knowledge shapes expectations regarding their behavior during interactions (Warpefelt, 2015). These expectations are conditional beliefs that players hold about the probabilities of NPCs' actions during interactions (Roese & Sherman, 2007; Hoorens & Ramachandran, 2012). Expectations are influenced by players' mental models related to video games – cognitive frameworks individuals construct to comprehend and predict the dynamics of their surroundings (Rabardel, 1995; Moray, 1996). These mental models are characterized by their goal-driven and simplified nature, aiming to provide players with a structured understanding of the game's challenges. For instance, when encountering enemies in video games, players primarily perceive them as challenges that align with their mental model of NPCs, simplifying their identity and purpose. This mental model doesn't delve into the nuanced backgrounds or motivations of each enemy but rather streamlines the perception of enemies as obstacles to overcome, aligning with the player's primary goal within the game. Consequently, the mental model supports efficient decision-making and action, allowing players to engage with enemies strategically and progress in their game experience. In other terms,

mental models often incorporate both the knowledge required to pursue a given goal and data extracted from the environment. This enables individuals to build mental representations that emphasize essential features while potentially overlooking irrelevant information (Ochanine, 1978). This intricate interplay between players' expectations and their mental models creates a dynamic feedback loop that game developers are keenly aware of and manipulate to enhance players' experiences. By skillfully combining NPCs' appearances with players' expectations, developers can streamline interactions, contributing to a more engaging and immersive gaming experience.

1.3.2 State of the art on players' expectations about NPCs

Warpefelt's research (2015) delves into players' identification of visual cues from NPCs' appearances and how these cues shape players' expectations of NPCs' roles and functions in the narrative. Through a two-part survey, the study gathered players' personal information that could influence their mental models, including demographics and media consumption habits, and assessed their expectations of NPCs based on appearance. Participants classified NPCs' images into conditional functions such as "Provide services" or "Guard places" and provided explanations for their choices (see figure 12).



Figure 12: Example of functions identified from NPCs appearance taken from *Skyrim* (2011). (Left) Provide services to the main character. (Right) Guard places in the virtual environment.

For instance, when presented with an image of an NPC located behind a shop counter surrounded by goods, participants commonly associated this visual cue with the function of selling items, and thus the role of a shopkeeper or merchant. The presence of the counter and goods, along with the physical arrangement, signaled to participants that the NPC is likely to engage in commerce-related activities. The study revealed that players' expectations are influenced by visual attributes and environmental context, highlighting the interplay between visual cues and mental models in guiding player interactions with NPCs. Thus, this experiment emphasizes the importance of coherent observation-based evaluation to comprehend users' expectations. Similarly, Rogers et al. (2018) conducted a study to unravel how players identify the narrative roles of NPCs in games and proposed a method for designing NPCs that visually align with expectations associated with specific roles like mentor, companion, enemy, and personality traits such as «wise,» «knowledgeable,» and «helpful». The researchers used an online survey to evaluate participants' expectations regarding the roles and traits of NPCs, whose gender was manipulated. The study examined how well-designed character images align with intended roles and whether gender plays a role in influencing perceived character roles. The study provides insights into how NPCs' roles can be conveyed through appearance alone, benefiting both game designers and researchers in enhancing player experience.

1.4 Violation of players' expectations

Video games offer narrative experiences, prompting game developers to incorporate surprising and unexpected events to heighten players' attention and engagement (Lavik, 2006). NPCs, being central to the game experience, often serve as such unexpected elements due to players' substantial expectations surrounding them. For instance, in the western action-adventure game *Red Dead Redemption 2* (Rockstar, 2018), the main character can encounter civilians along the roads who ask for help but often turn out to be trying to rob the main character if he stops to assist them. This unexpected event can occur when players identify NPCs based on their appearance and encounter a violation of their expectations, rooted in the mental model of needy characters commonly found in video games. The aim of this event is to inject dynamism into the game experience and create stress in players' minds as the main character must make rapid decisions to overcome unexpected enemies. The following subsection introduces the concept of violating expectations and explores its impact on those holding these expectations.

1.4.1 Definition of violation of players' expectations

When navigating in the virtual environment, players can be surprised by unexpected events similar to the one described above. These unexpected situations disrupt players' expectations, potentially leading to revisions of their mental models of the game world. This phenomenon has spurred extensive research into the violation of human expectations, resulting in the development of various theoretical models that shed light on their cognitive and behavioral impact. In the following, the different models are presented and adapted in the video game context, using the example of a military shooter game to explain how players adapt to unexpected events.

The Predictive Processing framework (Friston et al., 2012; Ransom et al., 2020) posits that the brain's core objective is to minimize surprise by continuously comparing incoming sensory data with its internal predictions. This process involves adjusting expectations through mechanisms like perceptual inference and active inference. Perceptual inference occurs when the brain revises its predictions to align with the external game environment, thereby reducing prediction errors and minimizing surprise. For instance, in a military shooter game's situation, players might anticipate encountering a small group of enemies in a specific location but find an unexpectedly large one instead. Players' perceptual inference mechanism would adjust their expectation about the forthcoming battle's intensity based on the new information. On the other hand, active inference comes into play when they adapt their strategy to align with their revised expectations. In the context of the game, this could involve changing tactics and seeking cover to handle the unexpected reinforcements efficiently. Together, these processes dynamically adjust players' expectations to maintain their immersion and their understanding of the gaming experience.

The Violation of Expectation Model (Gollwitzer et al., 2018; Doering et al., 2018), the Meaning Maintenance Model (Proulx et al., 2012; Proulx & Inzlicht, 2012), and the Expectation Disconfirmation Model (Oliver, 1980, 2010) collectively provide valuable insights into how humans respond to violated expectations. The ViolEx Model delves into how individuals navigate unexpected information that contradicts their established expectations, focusing on preserving and altering those expectations. For instance, in a military shooter game, players might anticipate an allied NPC's support on the battlefield, only to witness them unexpectedly retreat. This unexpected event could prompt players to redefine the NPC's role in the game's narrative, accommodating the new information while maintaining their overall positive perception of the character. Alternatively,

players might assimilate the unexpected event into their pre-existing mental model, viewing it as an isolated incident rather than a recurring behavior. Building on this, the Meaning Maintenance Model and Expectation Disconfirmation Model shed light on humans' cognitive responses to unexpected events such as inconsistencies. The Meaning Maintenance Model suggests that players use assimilation or accommodation to restore cognitive coherence when faced with unexpected occurrences. For example, if a powerful weapon in a military shooter game behaves unexpectedly due to a glitch, players might accommodate this anomaly by attributing it to a unique feature of the weapon rather than discarding their entire understanding of its mechanics. Moreover, the EDM addresses recalibrating expectations. Imagine players anticipating a challenging boss battle at a certain point in the game, only to find the encounter less intense than expected. In response, players might recalibrate their expectations, considering that the easier battle was intentionally designed to maintain pacing, rather than perceiving it as a letdown. Furthermore, the Expectancy Violations Theory (Burgoon, 2015) extends understanding to social interactions. This theory suggests that humans respond to unexpected behaviors during their social interactions by employing compensatory strategies, like compensation and reciprocity. For instance during an intense battle, if a typically cautious ally NPC unexpectedly charges into a dangerous situation, players might compensate by adjusting their own behaviors to provide additional support and ensure success. Similarly, players could exhibit reciprocity by mirroring the ally's unexpected behavior, thus maintaining a sense of consistency within the game's narrative dynamics. These compensatory strategies outlined by the EVT offer valuable insights into how players manage unexpected behaviors within the context of video game interactions, contributing to the overall coherence and engagement of the gaming experience. These models collectively illustrate how humans navigate the intricate interplay between their expectations and real-time experiences in their surroundings. By using cognitive mechanisms such as accommodation, assimilation, and recalibration, players strive to restore coherence and meaning within the gaming context. These insights highlight the dynamic interaction between players' mental models and their experiences, underscoring the importance of maintaining consistency and logical behaviors in video game design to ensure an immersive and engaging gaming experience.

2. Believability

This section delves into one of the key dimensions contributing to players' immersion in video games: believability. Given that video games offer narrative experiences that sometimes encompass fantasy storyline and involve interactions with AI-driven characters, it becomes essential for players to suspend their disbelief and fully immerse themselves in the gaming experience to derive maximum enjoyment. For example, this involves embracing the logic of a fictional world, even if it contradicts the rules of the real world, such as accepting the existence of broomsticks in a fantasy storyline like *Harry Potter* (Rowling, 1997). Believing in NPCs follows a similar principle, as players need to invest belief in the characters they encounter, which may include accepting their stereotyped behaviors or fantastical appearances. The following sections discuss the definition of believability and its various approaches in NPCs' design, followed by an overview of the state of the art in research investigating this topic.

2.1 Definition of NPCs' believability

The concept of believability in characters' conception has evolved across artistic mediums and time. Originating in the 19th century as «suspension of disbelief», coined by poet Samuel Taylor Coleridge (1985), it initially referred to accepting supernatural elements in poetry. This idea, as explained in Coleridge's *Biographia Literaria*, aimed to create a semblance of truth and inspire imaginative engagement. Over time, the concept broadened, encompassing fictional scenarios in various artistic forms beyond poetry. Notably, during the 1930s, Disney's animation principles, explored in *The Illusion of Life: Disney Animation* by Thomas and Johnston (1981), enabled characters to transcend drawings and appear vividly alive with emotions and motivations. This milestone cemented the tie between believability and the «illusion of life». These principles have resonated through computer science and AI, guiding the creation of believable characters, bridging the gap between traditional animation and interactive digital realms to improve users' immersion (Mateas, 2001).

In the realm of video games, believability plays a pivotal role in enhancing players' overall enjoyment of their gaming experience (Brown & Cairns, 2004). Approaches to achieving NPCs' believability in video games can be categorized into two distinct perspectives. One perspective revolves around the endeavor to simulate human-like behavior in NPCs, aiming to challenge players' perceptions of interacting with limited entities (Even et al. 2018). Thus, the believability

of NPCs is contingent upon their ability to create the illusion of being under a human player's control, rather than being governed solely by AI algorithms (Livingstone, 2006; Gomez-Gauchía & Peinado, 2006). This intricate approach delves into the depths of artificial intelligence, employing common techniques such as behavior and dialogue trees to intricately design characters that communicate traits of rationality, dynamism, and lifelikeness (Loyall, 1997). For instance, in shooter games that provide both solo and multiplayer experiences, players face off against NPCs in solo mode and human players in multiplayer mode. When NPCs exhibit human-like behaviors in solo play, it boosts both players' enjoyment and their sense of progress and skill in anticipation of their multiplayer experience (Soni & Hingston, 2008). Furthermore, the work of Biocca et al. (2001) posited that NPCs endowed with complex behaviors to convey humanness tend to trigger a sense of social presence within players.

On the other hand, an alternative approach maintains the foundational definition of believability and the artistic dimension of the game experience. Mateas (2002) defined NPCs as artistic abstractions with simplified and exaggerated behaviors, designed to engage the audience in the intended game experience. In this perspective, NPCs possess limited behaviors, but they are meticulously designed to align with players' expectations, thereby preserving players' immersion and minimizing disruptions within the gaming experience (Bailey & Katchabaw, 2008). Ermi and Mäyrä (2005) emphasize that for players to become fully immersed in the game experience, they must perceive the NPCs they encounter as believable and lifelike characters. Pursuing this path allows players to maintain a state of suspended disbelief, accepting NPCs' limited behaviors as part of the overall gaming experience. In the fantasy RPG *Skyrim* (Bethesda, 2011), NPCs have predefined schedules and behaviors that they follow throughout the day. While these behaviors may seem limited on an individual basis, they collectively contribute to creating an immersive and believable game world. For instance, a blacksmith NPC might spend the mornings hammering at its forge, the afternoons interacting with customers, and the evenings relaxing in a local tavern. Although the individual behaviors are relatively simple, the combination of these routines makes the NPCs feel like they are part of a social world. Therefore, players accept these limited behaviors as they confirm their expectations about NPCs' role in the narrative but also enable them to understand and predict NPCs' actions while preserving their immersion in the game experience. Gomes et al. (2013) listed key dimensions that contribute to the believability of characters within

interactive experiences, noting that asking an audience to rate their evaluation of a character's believability can be difficult and abstract. Therefore, these dimensions aim to encompass various aspects that contribute to the perspective of believability. To illustrate the dimensions linked to believability listed by Gomes et al. (2013), let us consider the NPC named Elena, a skilled elven ranger in a fictional fantasy RPG:

- *Behavior Coherence*: Throughout the game, Elena consistently demonstrates her characteristic preference for stealth and closed-ranged attacks, embodying behavior coherence.
- *Change with Experience*: As the main character's journey with Elena, her initial aloofness gradually evolves into camaraderie with her companions, reflecting her change with experience.
- *Awareness*: Elena adeptly detects hidden traps, points out concealed pathways, and communicates subtle shifts in the environment, showcasing her heightened awareness.
- *Behavior Understandability*: Players understand that Elena's health is low when she starts displaying cautious behavior.
- *Personality*: Elena's sharp wit, love for nature, and tendency to express herself through cryptic poems reflect her unique personality.
- *Emotional Expressiveness*: When faced with a critical choice, Elena's conflicted expressions and thoughtful dialogue responses mirror her internal emotional struggle.
- *Social Relationships*: Elena's bond with her fellow rangers is illustrated through shared stories, synchronized combat strategies, and mutual trust during quests.
- *Visual Impact*: Elena's distinct appearance, marked by intricate elven designs, ethereal glow, and her trusty bow, instantly captures players' attention.
- *Predictability*: When the main character and Elena initiate an ambush, her behaviors are not reckless and can be anticipated to result in a strategic approach.

Incorporating these dimensions into the design of a character ensures to convey believable characters but also provides game developers with a comprehensive framework to evaluate players' perceptions. Notably, Gomes et al. (2013) suggest that all the dimensions can be adjusted or

adapted to the game's setting, and certain dimensions might hold more importance than others, such as predictability. The absence of predictability can indeed harm behavior coherence; however, an excessive predictability could also impact believability, as NPCs might be perceived as overly robotic (Lester & Stone, 1997).

In conclusion, the quest for believable NPCs has become central to crafting immersive video game experiences. This pursuit takes two intertwined paths: one driven by technological innovation, where AI breathes life into characters, and the other rooted in artistic intention, where characters harmonize with players' expectations. These perspectives are not conflicting, but rather complementary, shaping a holistic approach to NPCs' believability. The goal is not mere realism, but to immerse players in dynamic narratives. As gaming evolves, the endeavor to enhance NPCs' believability remains a testament to the industry's dedication to unforgettable storytelling.

2.2 State of the art on believable NPCs

Understanding how players evaluate NPCs' believability is crucial, yet it presents some challenges. Several studies have delved into enhancing the evaluation of NPCs' believability for players. Various approaches have been used, ranging from analyzing NPCs' appearance and behaviors to implementing and manipulating NPCs' parameters. Lee and Heeter (2012) investigated the factors influencing NPCs' believability and players' enjoyment by comparing how players evaluated friendly and hostile NPCs in various RPG games. Participants watched videos depicting interactions between the main character and different NPCs and rated their enjoyment and NPCs' dimensions of believability. These dimensions involved participants' evaluation of NPCs' personality, appearance, behavior, social interaction, goals, and emotional expressions. Their findings revealed that NPCs' believability was significantly influenced by attributes such as appearance, behaviors, and the roles that the NPCs played within the game's narrative. NPCs' believability was primarily affected by their appearance, as it played a crucial role in participants' first impressions. Other attributes such as personality, emotion, goals, and behaviors contributed to NPCs' overall believability, while social interaction had a weaker correlation in the provided context. Additionally, friendly NPCs were evaluated as more believable and enjoyable than their hostile counterparts. Lee and Heeter (2012) suggested that hostile characters are often considered less convincing due to their restricted scripted behavior (i.e., only aggressive behaviors) and limited

interaction time. Furthermore, their results indicated a positive correlation between enjoyment and believability, emphasizing the crucial role of this dimension in players' enjoyment of the game experience (Togelius et al., 2013). Warpefelt and Strååt (2013) analyzed NPCs in various video games and how they exhibit behavior coherent with the game environment and human-like interactions. Warpefelt and Strååt (2013) observed video recordings of game sessions depicting NPCs in actions with the main character to focus on instances where NPCs' behaviors undermined player immersion by disrupting their believability. Their analysis led to the formulation of a set of heuristics, termed «anti-heuristics», which aimed to identify specific undesirable NPC behaviors. The study highlighted the frequent violation of certain behavior categories:

- Adaptation, which referred to NPCs' inability to adjust behavior according to contextual changes;
- Lack of awareness where NPCs exhibited a lack of understanding of their surroundings or actions;
- Models of others, signifying NPCs' failure to emulate realistic human interactions;
- and Models of self, pointing to instances where NPCs did not act consistently with their established behaviors or traits.

These findings underscored the importance of NPCs possessing attributes such as omniscience, situational awareness, and self-awareness to enhance their believability. Their analysis shed light on issues within NPCs' behaviors but also offered insights that can guide game developers in creating more immersive and coherent game experiences. Afonso and Prada (2009) created a conceptual model to enhance social believability and integrated social relationships among NPCs. Their model was designed with the specific goal of improving the social believability of NPCs, recognizing the importance of social interactions in gaming experiences. Therefore, their model encompasses a multifaceted approach:

- Personality Modeling: It simplifies NPC personalities, focusing on Agreeableness inspired by psychological theories like the Five Factor Model (Boele De Raad, 2000).
- Relationship Dynamics: NPCs form meaningful relationships influencing their interactions.

- Theory of Mind (ToM): NPCs could understand others' beliefs, adding depth to interactions.
- Interaction Decisions: NPCs autonomously choose whom to interact with based on relationships.
- Information Exchange: Relationships and personalities impact conversations, with liked characters receiving more information.

The authors evaluated their model in a RPG game where participants engaged with two game versions, differing in the inclusion of the model in NPCs' design. The findings revealed that players favored the version featuring integrated social relationships, underscoring the potential for enriching RPGs with more immersive and socially dynamic NPCs. Similarly, Tjokrosetio and Chowanda (2021) aimed to enhance player immersion in RPGs by improving NPCs' believability through the integration of Facial Expression Recognition technology, which recognizes and interprets players' real-time facial expressions to influence in-game character behaviors. They designed and implemented Facial Expression Recognition's model in the video game *Skyrim* (Bethesda, 2011). Participants took part in an experiment where they observed interaction simulations between characters in both the modified and original game versions. Their evaluation encompassed various metrics, including characters' believability using the scale suggested by Gomes et al. (2013). Their results demonstrated significant improvements in NPCs' believability when they had more reactive facial expressions during their interaction with the main character.

In essence, the described studies collectively underscore the complexity of evaluating NPCs' believability and highlight various approaches for enhancing it. From appearance and behavior to social relationships and real-time facial expressions, each dimension contributes to creating more immersive and believable gaming experiences. In this manuscript, believability is approached with the perspective that NPCs can have limited interactions until they are aligned with players' expectations.

3. Perceived Intelligence

Users perceive artificial agents as social entities and attribute human-like characteristics to them during their interactions (Nass & Reeve, 1996). Among the recurrent qualities that players frequently mention in their feedback on forums and playtests is their assessment of NPCs' intelligence. Notably, NPCs often seem to lack what could be termed as «intelligence». However, as mentioned in Section 1 of this chapter, NPCs are artificial agents constrained by their roles and functions in the game experience; thus, they cannot possess intelligence similar to humans. Hence, game designers, when crafting NPCs to enhance their believability, strive to create the illusion of intelligence. In that sense, players must perceive NPCs' intelligence conveyed by the parameters of NPCs' design. This section introduces and explores the concept of perceived intelligence within the context of users' interactions with NPCs. Subsequently, a comprehensive overview of the current state of research on perceived intelligence will be presented.

3.1 Definition of NPCs' perceived intelligence

Perceived intelligence refers to users' assessment of artificial agents' actions, including their evaluation of the apparent logic of the behaviors and the agents' performance in executing tasks (Bartneck et al., 2009). Moussawi et al. (2021) emphasized that a positive assessment of artificial agents' perceived intelligence is crucial for fostering users' engagement and adoption in new users. To aid designers understand how artificial agents are perceived as intelligent, Bartneck (2009) proposed a scale comprising five semantic differential items: Incompetent/Competent, Ignorant/Knowledgeable, Irresponsible/Responsible, Unintelligent/Intelligent, Foolish/Sensible. This scale offers distinct advantages as it comprehensively addresses both the evaluation of the logic of the behaviors (reflected in Responsible, Intelligent, and Sensible ratings) and performance (indicated by Competent and Knowledgeable ratings). By utilizing this scale, it becomes possible to effectively analyze users' perceptions of artificial agents' intelligence. This assessment considers not only users' evaluation of agents' performance but also agents' capacity to facilitate meaningful interactions.

Video games designers strive to convey the illusion of intelligence to enrich players' experience since the first interactions with NPCs. In this context, game designers' approach to NPCs' intelligence has long been associated with the performance of the characters during their interaction encompassing their decision making and reactivity to the main character's actions. For instance,

the video game *Space Invaders* (Taito), released in 1978, featured alien enemies displaying random, volatile changes in their trajectories, which players interpreted as intelligence by identifying patterns. However, this strategy to convey a sense of intelligence is effective only for short interactions and becomes ineffective as players realize the randomness of their behavior (Moussawi et al., 2021). As mentioned in Section 1, one of the aspects of NPCs is their limited abilities; therefore, the more players interact with them, the quicker they notice these limitations (Bartneck, 2009). Game designers recognized the significance of creating responsive and contextually relevant actions within the game world (Isla, 2005) and introduced behavior trees to address such limitations. A behavior tree is a metaphorical representation of decision-making processes, dictating how an NPC responds to various stimuli or the main character's actions. This innovation aims to contribute to the illusion of intelligence strived by game designers as NPCs would be more reactive to the virtual world. In Halo (Bungie, 2001), the behavior tree system empowered NPCs to exhibit a range of actions, from taking cover during combat to coordinating flanking maneuvers. By implementing behavior trees, developers improved players' experience through NPCs' reactive behaviors (Isla, 2005). Therefore, crafting NPCs' behaviors is crucial as it heavily influences how players will perceive their intelligence. Additionally, players' mental models of NPCs during an interaction influenced their evaluation, as these models establish expectations for the logic behind the agent's behaviors and their accurate performance (Ohrberg, 2019). As reported in Section 1, players might have different expectations about NPCs' role and function suggested by their appearance (Warpefelt, 2015). For instance, prior to the introduction of behavior trees, players' expectations of enemies' behaviors encompassed a limited range of actions, often consisting of random behaviors (Isla, 2005). However, with the incorporation of behavior trees into their gaming experience, enemies displaying a wide array of reactive behaviors began to defy these expectations. This, in turn, significantly enhanced players' perception of the enemies' intelligence as their unexpected complex behaviors induced more challenges in their game experience. In this situation, NPCs' illusion of intelligence was conveyed by their enhanced performance in the game.

In conclusion, while perceived intelligence reflects users' assessment of artificial agents' actions, encompassing logic and competence, the illusion of intelligence in NPCs is an intricate game design challenge that seeks to enrich players' experiences through contextually responsive and

performant behaviors. Therefore, the illusion of intelligence in NPCs could benefit from players' evaluation of NPCs' perceived intelligence, as it provides insights into how intelligence is conveyed during their in-game interactions.

3.2 State of the art on perceived intelligence

While crafting enemies for the video game Halo 2 (Bungie, 2004), Butcher and Griesemer (2004) conducted an experiment to explore players' perceptions of enemies' intelligence. Within the gaming experience, enemies were characterized by their high aggressiveness towards the main character and their resilience, as reflected in their health points. Their hypothesis centered around the belief that players expect intelligent enemies to present greater challenges in the game. To test this hypothesis, they created two versions of the game, each with distinct attributes determining enemies' strength: the damage they could inflict upon the main character and their allocated health points. As a result, the version with higher enemy's strength is more challenging than the other one. Researchers prompted participants to evaluate their perception of enemies' intelligence, leading to the observation that only 8% of participants gave positive evaluations in the version with lowered challenge, whereas 43% did so in the version with enhanced challenge. These findings indicate that manipulating indirect parameters of NPCs, such as their damage potency and resilience, can significantly influence how players perceive their intelligence. Additionally, Bartneck et al. (2009b) explored how the embodiment of artificial agents influences users' evaluation of their intelligence. By employing iCat and Robovie II as distinct robot embodiments, the researchers evaluated the impact of their appearances on users' intelligence assessment using the perceived intelligence scale suggested by Bartneck et al. (2009a). In their experiments, participants engaged in various interactions with the two robots and rated their perceptions of intelligence. Their results indicated that the two unique embodiments yielded differing measurements of perceived intelligence. iCat, with its animated facial expressions, was often perceived as more intelligent due to its engaging and interactive demeanor. In contrast, Robovie II, despite lacking animated facial features, conveyed a unique human-like presence through its mobility and articulated arms. The specific design elements and visual cues of each embodiment were significant in shaping users' evaluations of intelligence during their interactions. Furthermore, the impact of perceived intelligence on users' adoption of artificial agents has been studied by Moussawi et al. (2021). Their experiment examined the influence of perceived intelligence of

personal assistants like Siri (Apple) or Alexa (Amazon) on their adoption by users. In their experiment, new users interacted with personal assistants and were exposed to scenarios where the agents displayed varying degrees of perceived intelligence. Their result demonstrated a strong positive correlation between users' perceptions of higher intelligence in these personal assistants and their inclination to adopt them.

Collectively, these experiments demonstrate the intricate role of perceived intelligence during users' interactions with artificial agents. In the context of video games, crafting NPCs that accurately convey an illusion of intelligence would enhance the gaming experience and, consequently, players' engagement.

4. Research Questions and Hypotheses

The previous sections discussed how NPCs influence players' perceptions through their appearance and behaviors, impacting the overall game experience. NPCs assume specific roles in the game's narrative, shaped by developers who use appearance and behaviors as design elements to convey these roles to players. Players construct mental models of NPCs' roles based on their general knowledge and stereotypes, shaping their expectations for interactions. To enhance the game experience, developers align NPCs' appearances with stereotypical behaviors, ensuring coherence and meeting player expectations. Players assess NPCs, focusing on believability and perceived intelligence, which are crucial for creating an immersive and engaging gaming experience. While game developers can introduce unexpected events, like manipulating the alignment between NPCs' appearance and behaviors, in order to introduce narrative twists and captivate players' attention, the influence on NPC evaluation remains unclear. This thesis investigates how manipulating NPC design influences perceived intelligence and believability, with the hypothesis that coherence with player expectations enhances both dimensions, ultimately contributing to a superior gaming experience.

The following part of this manuscript presents the three studies conducted to answer this hypothesis. The general objective of these studies is to explore how manipulating NPCs' coherence influences players' assessments of perceived intelligence and believability. The initial study

focused on players' evaluation of NPCs' hostility in a military shooter game. In this study, we examined the impact of threatening appearances and aggressive behaviors on players' evaluations of hostility. To assess players' perceptions, we employed various scales from existing literature to gain a deeper understanding of how design parameters influence players' evaluations of NPCs. We hypothesized that players would assess NPCs' hostility based on both their appearance and behaviors.

The second study explored how manipulating NPCs' designs influenced players' gaming experiences, their expectations, and the impact of violated expectations on NPCs' perceived intelligence and believability. The parameters of NPCs' design were manipulated following the coherence and the consistency of the association between appearance and behaviors. During this study, players' behaviors and physiological activities were recorded as they interacted with NPCs. Perceived intelligence and believability were assessed after players' gaming experiences, involving both coherent and incoherent NPCs' designs. We hypothesized that players' gaming experiences would be influenced by the coherence between their expectations and NPCs' designs. For example, players interacting solely with coherent NPCs' designs would likely have a less challenging gaming experience compared to those engaging with both coherent and incoherent NPCs' designs. Additionally, we expected that players' evaluations of perceived intelligence and believability would be influenced by their gaming experiences. For instance, players who interacted exclusively with incoherent NPCs' designs might evaluate incoherent NPCs more positively than those who encountered only coherent NPCs.

The third study focused on an additional modality of interaction: textual interactions with NPCs. In this study, we analyzed the influence of NPCs' design parameters, including their role in the narrative and communication style, on players' gaming experience and their evaluation of perceived intelligence and believability. We hypothesized that coherence between NPCs' design parameters and players' expectations is crucial in textual interaction, similar to nonverbal interaction, for perceived intelligence and believability.

In essence, these studies collectively aim to deepen our understanding of how NPCs' coherence, shaped by design parameters, influences players' judgments of perceived intelligence and believability in various interaction modalities.

CHAPTER 2: Evaluation of virtual agents' hostility in video games

Poivet, R.^{1,2}, de Lagarde, A.¹, Pelachaud, C.¹, & Auvray, M.¹ (under review). Evaluation of virtual agents' hostility in video games. *IEEE Transaction on affective Computing*.

¹ Sorbonne Université, CNRS, Institut des Systèmes Intelligents et de Robotique, ISIR, Paris, France

² Ubisoft Paris studio, Montreuil, France

Abstract — Non-Playable Characters (NPCs) are a subtype of virtual agents that populate video games by endorsing social roles in the narrative. To infer NPCs' roles, players evaluate NPCs' appearance and behaviors, usually by ascribing human traits to NPCs, such as intelligence, likability, and morality. In particular, hostile NPCs in video games are essential to build the games' inherent challenges. The three experiments reported here investigated the extent to which the perception of hostility in a military shooter game (including both threat of appearance and aggressiveness in behaviors) is influenced by the appearance and the behaviors of NPCs thanks to perceived intelligence, likability and morality-related questionnaires. Our results first show that hostility is efficiently conveyed through NPCs' behaviors, but not significantly by their appearance. Second, our study allows identifying the main predictors of hostility perception, namely unfriendliness, knowledge, and harmfulness.

Index Terms — Affective issues in enhancing machine/robotic intelligence, Cognition, Entertainment, Games

Keywords — Video games, Hostility, Likability, Perceived intelligence, Non-playable characters

1. Introduction

In video games, the virtual agent embodied by the human player (referred to as the main character) interacts with virtual agents that populate the game: the non-playable characters (NPCs). These NPCs, which can be considered as a specific type of highly interactive virtual agents, are defined

by their role in the game's narrative. For instance, some NPCs can be designed as friendly artificial agents aiming to help the main character in pursuing the game (i.e., offering positive outputs for the player). On the other hand, some NPCs can be hostile toward the main character, and thus have negative interactions with them. The role of an NPC shapes the type of interaction with the main character which implies the manipulation of design factors to convey the intended gaming experience. In most video games, hostility is a crucial aspect as enemies endorse the challenge of the gaming experience. Hostility in games is conveyed through an antagonistic stance which implies threatening appearances and aggressive actions towards the main character. Game designers manipulate the appearance and behaviors of hostile NPCs to create diverse enemy types, enhancing the immersive gaming experience.

To communicate hostile NPCs' intentions, game designers rely on visual cues associated with hostile stereotypes and aggressive behaviors. Thus, creating NPCs involves manipulating their design factors during the production of the game; but how players experience and evaluate the hostility of the enemies they encounter remains unclear. In particular, the respective impact of visual and behavioral designs on players' evaluation of NPCs' hostility is yet to be unveiled.

The factors manipulated to convey NPCs' roles and hostility have the potential to elicit diverse expectations of personality traits. However, the precise impact of such interactions on personality traits remains unclear. For instance, in the game's narrative, morality holds substantial importance, and NPCs designed according to their roles, such as friendly versus hostile NPCs, should be perceived differently in terms of their moral character. Similarly, traits such as likability and perceived intelligence can vary depending on their appearance or behaviors, contributing to an immersive experience, and fostering a positive appreciation of interactions with NPCs. Consequently, the creation of engaging enemies requires a careful consideration of factors that not only convey hostility but also consider their potential influence on other traits that actively shape players' overall gaming experience such as their likability, morality, and perceived intelligence. To investigate the factors influencing players' perception of NPCs, and of their hostility, we conducted three experiments in the context of a military shooter game. Outcomes of such studies can inform game designers about the specific factors that contribute to the perception of hostility and subsequently help them modulate NPCs' parameters to enhance the overall gaming experience.

In the next section we give definitions of hostility and list its characteristics related to appearance and behavior. Section 3 introduces the three personality traits our study focuses on, namely morality, likability, and perceived intelligence. From Section 4 onward, we present the three conducted studies. Section 5 describes the methodology. The next three sections detail each of the experiments and section 9 discusses their results. Finally, the paper ends by presenting some limitations and future works.

2. Background

Antagonists are designed to play a unique role inside a video game’s narrative, and they need to look and act like the “bad guys”. The worst of them, the “final boss”, is reached after interactions with less hurtful antagonists. The hierarchy in challenges during the narrative allows keeping the game entertaining. To better understand how antagonists are conceived by game designers, the definition of hostility for virtual agents needs to be clarified. To do so, the parameters which allow virtual agents, including NPCs, to convey hostility will be detailed below.

2.1 Definition of hostility and its characteristics.

Most studies on human-agent interaction tries to develop artificial agents that adopt a positive stance towards humans (e.g., chatbots designed for elderly care, see [1]). Nonetheless, there could be some potential benefits to creating virtual “bad guys” and in particular when designing hostile NPCs. We detail the case of the perception of hostile virtual agents and NPCs.

Hostility has been the focus of studies in clinical psychology where it is usually defined as an antagonistic attitude towards people including cognitive, affective, and behavioral components [2], [3]. Cognitive aspects of hostility refer to cynical beliefs and mistrust of others (e.g., believing that an agent possesses negative intentions). The affective dimension covers the negative emotions induced by the antagonistic attitude (e.g., anger, disgust). The behavioral component involves verbal or physical aggressiveness which can be defined as impulsive or goal-oriented behaviors towards others motivated by a desire to harm them.

Hostility involves two core characteristics: threat and aggressiveness. Threat refers to the perceived potential for harm or danger posed by an agent or a situation [4]. It is often associated with the anticipation of negative outcomes or the activation of self-protective mechanisms. On the other hand, aggressiveness primarily focuses on verbal or physical actions directed towards others with

harmful intentions [5].

In social psychology research, hostility perception in others is characterized by the detection of physical and verbal aggressiveness [6], [3]. However, it is important to note that aggressiveness is a narrower construct than hostility, as it primarily encompasses observable behaviors. In contrast, hostility encompasses a broader range of negative attitudes, emotions, and antagonistic actions directed towards others. When interacting with a hostile person, individuals often infer the inner states of this person based on visual cues that communicate potential aggressiveness, which is perceived as a threat (e.g., facial expressions associated with negative emotions). These cues allow individuals to adapt their attitudes and behaviors accordingly, and the person's subsequent behaviors associated with aggressiveness serve to confirm the initial inference. Perceiving hostility of virtual agents relies on the same mechanism of decoding a combination of visual cues related to threatening appearances and aggressive behaviors. [7] created aggressive virtual agents to measure their effects on human perception, and to assess their potential benefits for clinical applications. They compared the level of stress induced in participants facing an aggressive human versus an aggressive virtual agent. In this study, both the virtual agent and the human adopted an aggressive stance towards the participant (e.g., screaming at them). Both interactions generated physiological stress in the participant, although the induced stress was lower when interacting with a virtual agent than when interacting with a human. [8] accounted for this result by the fact that virtual "bad guys" are non-consequential, whereas human-to-human interactions can have consequences on the human state. In the field of video games, designing hostile NPCs is highly relevant to make the game more challenging and engaging for the players. Hostile NPCs can inflict damage to the main character, which will be evidenced to the player by a loss in their so-called health points. Interacting with a hostile NPC could have multiple potential consequences, such as the need to find a friendly NPC to heal the main character's wounds or even its death. For instance, in the shooter game Ready or not the main character is heading a SWAT in multiple crisis scenarios (e.g., hostage rescue or bomb defusing) and they need to rapidly decide if the NPCs they encounter are hostile or not.

2.2 Threatful appearance

In video games, the rapid decision to interact with NPCs is often influenced by visual cues from their appearance design. The decision process is crucial when facing potential threats. In Ready or not, the main character may enter a room with hostile NPCs ready to attack them, which can lead

to the end of the game session if the player is not rapid enough. Detecting NPCs' social roles has been studied by [9] across different game types, revealing that players rely on stereotypical visual cues to understand NPCs' roles. For example, NPCs having the appearance of mentors may be associated with a beard, which conveys older age and higher competence [10]. These visual cues aim to activate stereotypes in the player's mind [11]. In Super Mario Bros., the aggressive and angry nature of the Goombas is conveyed through their slanting eyebrows. Furthermore, [12] examined the effect of manipulating facial features on humans' attribution of personality traits to virtual agents and found that wider faces influence the perception of trustworthiness, aggressiveness, and dominance. The narrative context also plays a role in determining how an appearance is perceived as a threat. In games like Call of Duty 4: Modern Warfare, a military shooter setting, human soldiers are considered as enemies, whereas in science fiction shooter games like Halo, they can be perceived as allies. Apart from stereotypes, the mere presence of a weapon in an NPC's appearance is often sufficient to communicate a threat in players' minds. Detecting the presence of a weapon provides information about potential behavioral patterns, attack range, frequency, and any special abilities the NPC possesses [13]. Consequently, when the main character encounters NPCs for the first time, players can make predictions about their abilities, intentions, and emotional states, which significantly influence their subsequent interactions.

2.3 Aggressive behaviors

For virtual agents, behaviors are mainly referring to the nonverbal abilities that turn their static appearance into a dynamic stance. Nonverbal behaviors include facial expression [14], gaze direction [15] and different types of gestures [16]. As nonverbal communication adds supplementary information to the verbal communication, it enhances the believability during the interaction with the agent [17]. The implementation of such nonverbal cues on virtual agents influences human perception, as shown in [18] study. In their experiment, users evaluated their interactions with virtual agents more positively when they perceived them as more reactive and socially present. The interaction with virtual agents gains relevance in the eyes of the users when they are designed with behaviors such as facial expressions or gestures.

NPCs can be seen as interactive virtual agents, which can be designed with complex behaviors. In that sense, NPCs' behaviors refer to the sets of predefined sequences of actions leading to interactions with their surroundings. In shooter games, hostile NPCs are characterized by their

aggressiveness toward the main character. Game designers manipulate the aggressiveness of enemies by utilizing a classification system relying both on their behaviors and their weapons [13]. For example, enemies armed with shotguns follow a sequence of actions that involve shortening the distance due to the weapon's limited range, while enemies equipped with sniper rifles behave differently by maintaining their distance from the main character. By considering these parameters, game designers allow players to anticipate the behaviors of enemies from the identification of their weapons.

In addition, game designers have the flexibility to manipulate other parameters, such as enemies' shooting precision (i.e., their accuracy in hitting targets), the damage they can inflict, and even their decision-making processes as they approach the main character. Decision-making particularly affects how enemies manage cover sequences, as game environments often include cover walls that create safe zones for both the main character and NPCs. For instance, enemies that utilize cover tactics generate less pressure on the main character compared to those who recklessly charge without seeking cover. In conclusion, game designers shape the behaviors of hostile NPCs to convey the game's level of challenge by carefully manipulating a combination of parameters and action sequences. These design choices directly influence the gaming difficulty, as a more aggressive NPC presents a greater threat to the main character and increases the risk of defeat.

3. Personality traits

First impressions play a crucial role in evaluating others, as they activate our knowledge and concepts to predict interactions and influence decision-making [19]. This extends to artificial agents, as humans tend to perceive computers and robots as social actors [20]. NPCs in video games can also make first impressions on players, who infer both basic (good or bad) and complex (personality traits) information about them. When the main character encounters an NPC for the first time, players' preconceived notions of artificial agents come into play, shaping their expectations for future interactions, and influencing their decision to engage. For example, in the comparison between games like Call of Duty 4: Modern Warfare and Halo, players might be more inclined to approach visually similar soldiers in Halo due to their knowledge of the game's context, where enemies are non-humanoid. On the other hand, encountering a humanoid soldier in Call of duty would be a threat in players' minds. Thus, visual cues from NPCs' appearance allow players

to ascribe personality traits, such as morality, and predict their role in the game [21]. Visual cues and the game context can lead players to infer that a character is friendly and increase the likelihood of their interaction, while NPCs perceived as unfriendly may be approached with caution. These assumptions rely on stereotypes associated with likability and morality, which are evaluated based on NPCs' appearance [22]. During gameplay, players compare their expectations with the behaviors they observe in NPCs, updating their mental representation. The ascription of personality traits to NPCs, including morality, likability, and perceived intelligence, is a crucial aspect of the gaming experience, making the interaction more engaging [23]. Examining players' perception of these traits in relation to hostility can provide valuable insights for game designers, as it allows a deeper understanding of how hostility influences players' perception during interactions with NPCs. To investigate this further, our study analyzes scales' items that focus on three dimensions relevant to social perception: morality, likability and perceived intelligence.

3.1 Morality

As most video games are narrative-oriented, the perception of NPCs' roles is important for players' mental representations. The role of a character in fictional media can be described thanks to their morality traits within the narrative. According to [24], morality can be defined as “the differentiation of intentions, decisions and actions between those that are distinguished as proper and those that are improper”. In the field of video games, morality has been investigated through the antagonistic characters' physical appearances [21]. In fact, the concept of antagonist is bonded to the notion of morality ascription as they are introduced as “the bad guy” of the story ([25], [26], [27]). The morality of antagonists is based on the evaluation of their actions and the nature of their intentions. To assess fictional characters' perceived morality, [28] introduced a questionnaire based on the Moral foundation theory [29], relying on five-dimension items: Harm/Care, Fairness/Reciprocity, Ingroup/Loyalty, Authority/Respect, Purity/Sanctity. Using these scales, [21] outlined visual cues from the appearance that triggers mental representations associated with morality. They noted that players were able to ascribe moral intention to characters they had never seen before using only their physical appearance. Their results show that certain stereotypical items extracted from NPCs' appearance were associated with a perception of immorality (i.e., covered faces, dark colors clothes, skin problems, shape dysmorphia). Therefore, as part of the ascribed human traits, morality perception participates in players' mental representations of NPCs. Players'

evaluation of an encountered character's morality empowers their decision to interact with them and the way they do so.

3.2 Likability

One dimension which has been shown to influence human's perception of agents is their likability. In that sense, [30] developed a likability evaluation scale made of five differential items: Like/Dislike, Kind/Unkind, Pleasant/Unpleasant, Friendly/Unfriendly, Nice/Awful. This scale has been used to assess the first impression that humans have of robots' likability [31] and it can be used for NPCs in the context of video games. In the case of the Goombas in Super Mario Bros. introduced earlier, the choice of certain visual cues (i.e., the slanting eyebrows) aims to be automatically perceived as unlikeable by players. Having this negative first impression solely based on the Goombas' appearance allows players to rapidly detect the Goombas' intentions and to expect subsequent unlikable behaviors. Therefore, very much like humans meeting other humans, NPCs such as Goombas can be perceived as more or less likable. This dimension can be manipulated by game designers to convey NPCs' role within the game.

3.3 Perceived intelligence

Humans attribute intelligence to artificial agents based on their appearance and behavior. People's evaluation of intelligence relies on two main dimensions: understandability and performance. On the one hand, [32] underlined the difficulty to understand the underlying intention of an artificial agent from its behaviors. For instance, in video games, NPCs have limited behaviors (i.e., sets of available interactions) while players are unpredictable. The combination of limited abilities and unpredictability can lead NPCs to have inaccurate behaviors, which are seen as mistakes by players. When players cannot explain NPCs' behaviors, this negatively influences their perceived intelligence and thus their overall evaluation. On the other hand, [33] explained that virtual agents' performance to complete their task is crucial to users' evaluation. For instance, hostile NPCs are designed to create difficulty by endorsing the role of opponents. Players have expectations about what opponents should be able to do in each context and if hostile NPCs are not performing well in this role (e.g., they cannot defeat the main character), they are perceived as lacking intelligence. In 2002, the video game company Bungie conducted an experiment while creating their game Halo [34]. They manipulated game sessions' difficulty by creating different hostile characters who can be defeated more or less easily and who can inflict different levels of damage. Then, they asked

players to judge the presence or absence of intelligence in NPCs on a binary scale and found that NPCs that are more challenging for players are perceived as more intelligent. [32] proposed a scale to assess users' evaluation of artificial agents' perceived intelligence that consist of five semantic differential items: Incompetent/Competent, Ignorant/Knowledgeable, Irresponsible/Responsible, Unintelligent/ Intelligent, Foolish/Sensible. Using the scores of each of the items, the scale has the advantage to cover the evaluation of both understandability (Responsible, Intelligent and Sensible scores) and performance (Competent and Knowledgeable scores).

4. Experiments on hostile NPCs

To summarize, players rely on their knowledge and stereotypes to infer NPCs' role inside the game by attributing human traits to them and infer potential hostile intentions based on their appearance and behavior. To convey hostility, NPCs' appearance must be perceived as a threat; this perception aims to help the player make the best decision, namely, to decide whether to interact or not. Then, during interactions, players update their judgment about NPCs. For instance, interacting with hostile NPCs might result in negative consequences for the main character, such as a loss of health points or a defeat, which induces the player to categorize their behaviors as aggressive. The aim of the present research is to investigate how hostility is conveyed by non-playable characters. We hypothesize that hostility evaluation in virtual agents is a combination of their appearance activating threatful concepts and their display of aggressive behaviors. In the present study, three experiments were conducted to investigate these factors of appearance and behavior involved in hostility evaluation as well as their respective contributions. Hostility evaluation was investigated through the analysis of scales' items of the three dimensions: likability, perceived intelligence, and morality. More precisely, in the context of a Ubisoft military shooter game, the first experiment focuses on the evaluation of the threat conveyed by NPCs' appearances. The second experiment assessed the evaluation of the aggressiveness in NPCs' behaviors. The last experiment used the results from the previous two experiments; the appearance and behavior of NPCs were manipulated to evaluate their impact on hostility evaluation. In addition, we measured the scales' items predicting the perception of hostility.

5. General Methods

Participants

For each study, participants were volunteer players registered at the Ubisoft User Research Laboratory platform (i.e., a mailing platform where players can deliberately enter their information to participate in Ubisoft's research). All the participants were contacted by email and were provided with information about the content of the research. The email informed participants of the free nature of their participation and that they could stop at any time they wanted. Each experiment took approximately 15 minutes to complete and was conducted in accordance with the Declaration of Helsinki.

Procedure and materials

The three experiments were conducted online. The participants were first welcomed and introduced to the topic of the research. Then, they were presented with the stimuli and asked to answer questionnaires.

To create the stimuli, NPCs from the military video game Ghost Recon Breakpoint were used. This is a realistic military game in which the main character fights against hostile soldiers and interacts positively or in a neutral way with civilians. In the game, there are more soldiers than civilians, as in military shooter games soldiers are at the core of the action. In our study, this distribution was respected, and hence, six different archetypes of NPCs were selected: five soldiers and one civilian. Archetypes are represented through an appearance that involves specific visual cues and behaviors that aim to express NPCs' role inside the game (see Figure 1 for an illustration and Table 1 for the details). For each archetype, three variations were chosen to display different samples of appearance (i.e., 3 different civilians that differ by the color of their outfit).

TABLE 1
DESCRIPTION OF THE APPEARANCES OF NPCs SELECTED FOR EXPERIMENT 1

Archetype of the NPC	Role in the game	Head	Body	Accessories
Civilian	Not opponent	Uncovered face	Unmuscular shape. Recreational light-colored clothes (jacket or t-shirt with graphics, classic pants, and baskets shoes)	None
Drone Carrier	Opponent	Covered face (Balaclava)	Athletic shape. Black t-shirt, blue jeans, black boots, and gloves	Ammunition harness (dark colors), backpack, empty holster, and small screen on the right wrist
Rifleman	Opponent	Military cap (dark colors)	Athletic shape. Dark t-shirt, blue jeans, dark boots, and gloves	Light harness and empty holster. Headset and sunglasses
Rocket Launcher	Opponent	Partly covered face (dark colors). Red military cap	Athletic shape. Grey t-shirt, blue jeans, and black boots	Backpack and empty holster. Headset
Rusher	Opponent	Partly covered face (dark scarf on the lower face)	Athletic shape. Black t-shirt, blue jeans, black boots	Ammunition harness (light colors), empty holster and ski goggles
Sniper	Opponent	Uncovered face	Athletic shape. Dark t-shirt, blue jeans, dark boot.	Ammunition harness (dark colors), a tight protection and empty holster



Fig 1. NPCs' design from left to right: Civilian (one of the three displayed civilians, friendly role, will get on their knees if the main character targets them with their weapon), Drone carrier (enemy role, spawn drones to attack as their main behavioral pattern but can use a rifle at close ranges), Rifleman (enemy role, use a rifle), Rocket launcher (enemy role, use a rocket as their main weapon but can use a rifle at close ranges), Rusher (enemy role, use a shotgun and will run toward a main character located at a close range), and Sniper (enemy role, use a sniper but can use a rifle at close ranges).

6. Experiment 1: Threat of NPCs' appearances

Experiment 1 aims at evaluating players' perception of the threat conveyed by NPCs based on their appearance (i.e., NPCs' shapes and the threatening cues from their outfit). As the aim of our study is to investigate players' evaluation of NPCs from their appearance, all soldiers were displayed without their weapon to limit participants' inference of their abilities and focus on NPCs' shape and outfit. Following their role in the game, our hypothesis is that the appearance of civilians should be perceived as less threatening than soldiers' ones. This difference would reflect the type of interaction expected in the game, as civilians are neutral or friendly characters while soldiers are enemies. In addition, soldiers should be perceived at a distinct level of threat based on the visual cues (e.g., military gear) from their appearance.

Participants, stimuli, and procedure

Thirty-nine French participants completed Experiment 1 (thirty-six men, one woman and two others, mean age = 31.4 years old, SD = 9.2). The gender repartition can be explained by the significant part of men registered on the Ubisoft User Research mailing list and that the survey was randomly sent to players and was about a military game (i.e., potentially interested gamers were mostly men). In Experiment 1, the materials consisted of static images of the six standing NPCs (see Figure 1 and Table 1). Each NPC is shown with a neutral posture, no visible weapon (i.e., soldiers wore an empty gun holster and different ammunition belts), no visible background, and controlled skin color (see Figure 2 for a detailed description of NPCs' appearance). NPCs' appearances were randomly presented three times each. As said earlier, there were more stimuli of soldiers than of civilians to respect the repartition in the original game. For each presentation, participants rated the amount of threat they perceived on a continuous scale (from '0 - not threatening at all' to '100 - totally threatening') and responded to the Likability and Perceived Intelligence scales (Godspeed, [32]) and to the morality questionnaire (CMFQ-S, [28]).

Results and discussion

Scores were obtained from participants' ratings of NPCs' threat in appearance, and the scales' items of Likability, Perceived intelligence, and Morality. In order to maintain the representativeness of the dataset, no explicit outlier treatment was applied, hence all the collected data was included in

TABLE 2
DESCRIPTIVE STATISTICS FOR THREAT SCORES

	Civilian	Drone Carrier	Rifleman	Rocket Launcher	Rusher	Sniper
N	78	78	78	78	78	78
Mean	1.417	51.2	29.75	30.883	42.017	29.317
SD	2.872	20.757	18.636	19.297	18.813	19.827

the statistical analysis. The threat's scores did not follow a normal distribution. Therefore, a one-way ANOVA on ranks (Kruskal-Wallis H test) was conducted in order to investigate the significant

difference of threat between NPCs' appearances ($df = 5, p < .001, \eta^2 = 0.429$). Civilian's appearance was perceived as significantly less threatening than soldiers' ones. Moreover, Drone carriers' appearance was perceived as the most threatening among all the soldiers (see Figure 2). A significant correlation was found between threat and the items from the Likability scale "Friendly" ($r = -0.364$), "Kind" ($r = -0.235$), "Pleasant" ($r = -0.17$), "Nice" ($r = -0.351$), "Intelligent" ($r = 0.249$), "Competent" ($r = 0.43$), as well as with the Morality scales regarding "Ingroup/Loyalty" ($r = 0.261$), "Harm/Care" ($r = 0.64$), "Authority/Respect" ($r = 0.67$) and "Fairness/Reciprocity" ($r = 0.515$). A multiple linear regression was conducted to identify the predictors of the score of threat perception. All the items explained 56.2% of the variability of the threat score ($R^2 = 0.562$). However, the items "Competent" ($\beta = 0.143, p = .034$), "Harm/Care" ($\beta = 0.255, p < .001$) and "Authority/Respect" ($\beta = 0.357, p < .001$) were significant predictors of threat perception. A multiple linear regression with those items showed that they predict 53.5% of the variability of the threatening score ($R^2 = 0.535$).

Experiment 1's results show a significant difference in threat perception between the NPCs (see Figure 2). The civilian's appearance was perceived as less threatening compared to all soldiers' appearances. This difference is explained by very distinct characteristics in appearance, such as casual versus military clothes and differences in body shape. The overall shape of soldiers is more squared with a prominence of muscles. These results are in line with [35], according to whom square shapes in fictional characters refer to strength and stability while round shapes are linked to safety and friendliness. In addition, soldiers' appearances include military accessories which are

associated with concepts of war and violence. The linear regression showed a granularity of the threat attribution within soldiers' appearances. NPCs with the most threatening appearances were perceived as more competent, dangerous, and antipathetic. However, none of the items from the Likability scale significantly predicted the threat score, suggesting the low importance of this dimension to evaluate enemies' threat. Furthermore, one of the differences between soldiers' appearances was the visibility of their military gear. Soldiers perceived as the most threatening had also their face hidden (see the general discussion in Section 9, for further discussion of the results).

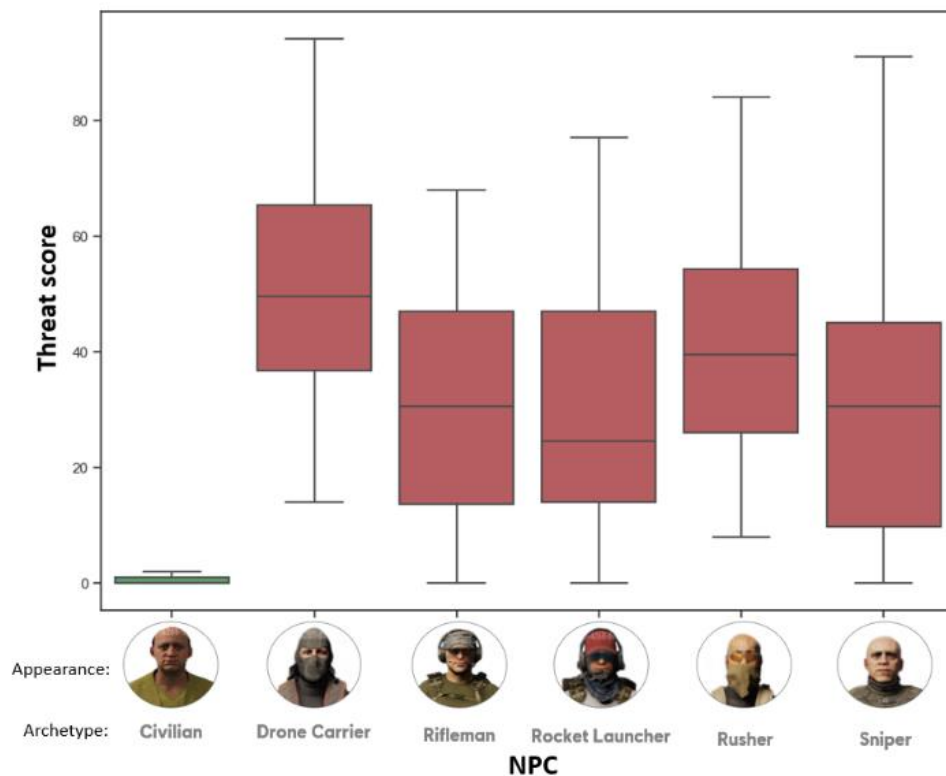


Fig. 2. Boxplots of NPCs' threat scores: boxplots' colors represent NPCs' role inside the narrative (Civilian in green and Soldiers in red). The boxplots show the distribution of the threat score by NPCs' appearances, their median and interquartile range. Black error bars represent the standard error of the mean.

7. Experiment 2: Aggressiveness of NPCs' behaviors

Experiment 2 investigates how players evaluate the aggressiveness of several NPCs' behaviors. Based on their roles and the intended challenge within the narrative, NPCs behave differently to create different types of interactions with the main character. Civilians are friendly NPCs in military games, they should not be perceived as aggressive. Soldiers are the opponents in the narrative and their goal is to try and kill the main character. Therefore, unlike civilians, soldiers' behaviors should be perceived as aggressive. In addition, as hostile NPCs have different behaviors to express their stance toward the main character, participants should rate their aggressiveness differently.



Fig. 3-a. (left). In-game screen capture of NPCs' appearance used for the video. Note that all NPCs had the same appearance (same face, body shape, clothes, and weapon). Fig. 3-b. (right). Illustration of the task: the main character moves toward the NPC inside a corridor. Here the main character is in the background while the NPC is in the foreground. Note that for each image, watermarks are automatically generated and required by Ubisoft.

Participants, stimuli, and procedure

Forty-three French participants completed Experiment 2 (thirty-eight men, four women and one other, mean age = 31.9, SD = 8.2). In Experiment 2, we controlled the appearance of the NPCs; the

six NPCs' appearances were identical (see Figure 3-a), which means that all NPCs' appearances consisted of the same face, body shape, clothes, and weapon (a rifle, which is the most common weapon in military shooter games). The only difference between them was their behaviors, that is, their sequence of actions (see Table 3). For instance, civilians in military games get on their knees while soldiers attack the main character. For soldiers, the behaviors differ as a function of two parameters: shooting (number of shooting sequences, precision, and distance) and cover sequences (hiding behind a wall or rushing toward the main character). As a result, the manipulated parameters lead to different consequences for the main character's state (i.e., dead or alive). NPCs' behaviors were shown to the participants via videos of pre-recorded interactions between the main character (controlled by the experimenter) and each of the NPCs. In each sequence, the main character moves toward the NPC inside a corridor. The video camera was placed in a specific angle to help participants focus on NPCs' behaviors (see Figure 3-b). For each video, the participants completed the same scales as those used in Experiment 1, except the threat score which was replaced by aggressiveness (from '0 - not aggressive at all' to '100 - totally aggressive').

TABLE 3
DESCRIPTION AND CONSEQUENCES OF THE BEHAVIOR OF NPCs SELECTED FOR EXPERIMENT 2

Archetype of the NPC	Role	Shooting distance	Length of shooting sequences (seconds)	Cover sequences	Main character's state at the end of the interaction
Civilian	Not opponent	Not applicable	0	Get on their knees	Alive
Drone Carrier	Opponent	Long distance	6	Hide behind a wall	Alive
Rifleman	Opponent	Long distance	3	Hide behind a wall	Dead
Rocket Launcher	Opponent	Long distance	10	Hide behind a wall	Alive
Rusher	Opponent	Short distance	3	Rush toward the main character	Dead
Sniper	Opponent	Long distance	5	Hide behind a wall	Dead

Results and discussion

Scores were obtained from participants' ratings of aggressiveness of NPCs' behaviors and their ratings in the scales' items of Likability, Perceived intelligence and Morality. In order to maintain the representativeness of the dataset, no explicit outlier treatment was applied, hence all the

collected data was included in the statistical analysis. The perceived aggressiveness's scores did not follow a normal distribution. Therefore, a one-way ANOVA was conducted on ranks (Kruskal-Wallis H test) to assess the significant difference of aggressiveness between NPCs ($df=5$, $p < .001$, $\eta^2 = 0.555$). The less aggressive behavior was of the civilian, which is characterized by getting on his knees when the main character comes too close. The most aggressive behavior was the Rusher's, which is characterized by a reduction of the distance between him and the main character. A significant correlation was found with aggressiveness for the items: "Friendly" ($r = -0.657$), "Kind" ($r = -0.415$), "Nice" ($r = -0.492$), "Responsible" ($r = 0.311$), "Sensible" ($r = 0.384$), Knowledgeable ($r = 0.503$), "Intelligent" ($r = 0.494$), "Competent" ($r = 0.558$), "Harm/Care" ($r = 0.739$), "Purity/Sanctity" ($r = 0.505$), "Authority/Respect" ($r = 0.688$) and "Fairness/Reciprocity" ($r = 0.632$). A multiple linear regression was conducted to identify the predictors of the score of aggressiveness. All the items explained 71.1% of the variability of the aggressiveness score ($R^2 = 0.711$). However, among them, only the items "Friendly" ($\beta = -0.279$, $p < .001$), "Knowledgeable" ($\beta = 0.120$, $p = .043$), "Harm/Care" ($\beta = 0.214$, $p < .001$), "Authority/Respect" ($\beta = 0.167$, $p = .005$), and "Fairness/Reciprocity" ($\beta = 0.148$, $p = .004$) were significant predictors of aggressiveness perception. Indeed, a multiple linear regression with those items showed a prediction of the variability of 69.5% of the aggressiveness score ($R^2 = 0.695$).

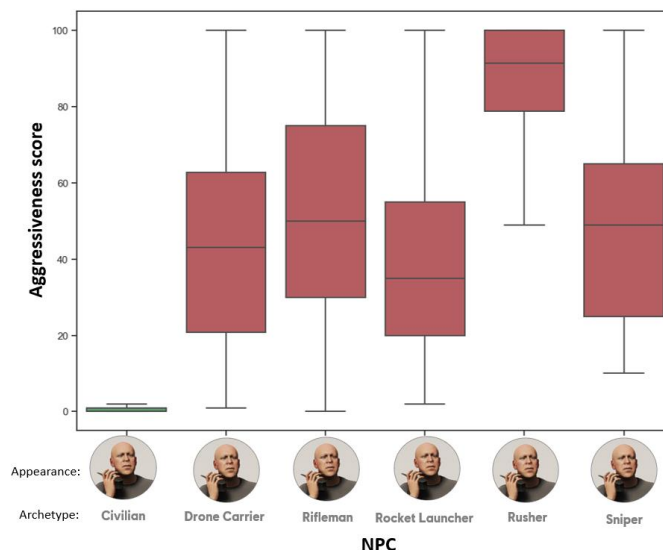


Fig. 4. Boxplots of NPCs' aggressiveness scores: boxplots' colors represent NPCs' role inside the narrative (the green color for Civilian and red for Soldiers). The boxplots show the distribution of

the aggressiveness score by NPC behavior, their median and interquartile range. Black error bars denote the standard error of the mean.

The results from Experiment 2 indicate a significant difference of aggressiveness evaluation between all the NPCs. The difference of aggressiveness between civilian and soldiers' behaviors is explained by the lack of damage resulting from the civilian's activity. Indeed, the combination of the items "Friendly", "Knowledgeable", "Harm/Care", "Authority/Respect", and "Fairness/Reciprocity" refers to the perceived efficiency to harm the main character's health. The behavior perceived as the most aggressive one has the particularity of distance reduction between the two agents (i.e., rushers' behavior). This behavior induces in participants' the perception of an intrusion into their main character's safety zone. Other soldiers' behaviors were defined as less mobile (i.e., hiding behind a wall), meaning that the NPC is less intrusive into the main character's safety zone.

8. Experiment 3: Contribution of appearance and behavior to the perception of hostility

Experiment 3 investigates the influence of NPCs' appearance and behaviors on hostility perception. More precisely, the aim is to investigate within the predictors of threat and aggressiveness, those influencing players' evaluation of hostility. During this experiment, the participants completed the individual items that were found to be predictors of threat ("Competent", "Harm/Care" and "Authority/Respect") and aggressiveness ("Friendly", "Knowledgeable", "Harm/Care", "Authority/Respect" and "Fairness/Reciprocity") in Experiments 1 and 2.

Participants, stimuli, and procedure

Fifty-one French participants completed Experiment 3 (forty-six men, four women and one other, mean age = 30.03, SD = 6.4). Four NPCs were created based on the results from the two previous experiments. The appearances and behaviors with the highest and lowest threat and aggressiveness scores from Experiments 1 and 2 were used. From these, two congruent NPCs (most and least threatening appearance and aggressive behavior combined) and two incongruent NPCs (most threatening appearance and least aggressive behavior, and vice versa) were created.



Fig. 5-a. (left) Illustration of the NPCs' appearance (most threatful appearance on the left: Drone Carrier and least on the right: Civilian), made of a screenshot of the NPCs encountered in the video. Fig. 5-b. (right) Screenshots from the videos illustrating the interactions between the main character and the NPCs. The two pictures on the top depict an NPC with a civilian appearance behaving non-aggressively (left) and aggressively (right) toward the main character. At the bottom, an NPC with a soldier appearance follows the same dichotomy. Same as E2, watermarks are automatically generated and required by Ubisoft when screens are captured in the game's editor.

For each NPC, an interaction with the main character was recorded. The position of the camera was placed behind the main character, as it is the case in most video games. The videos stop after the interaction (i.e., either the main character is killed by the NPC or the main character reaches the crouched NPC). Before watching each video, the participants first had to rate the threat of the appearance (from '0 - not threatful at all' to '100 - totally threatful'), based on an in-game picture of the NPC (see Figure 5-a). Then, the participants watched the videos and evaluated NPCs' aggressiveness in behavior (from '0 - not aggressive at all' to '100 - totally aggressive') and hostility in NPCs (from '0 - not hostile at all' to '100 - totally hostile'). After that, the participants were asked to complete the following individual items: "Friendly", "Competent", "Knowledgeable", "Harm/Care", "Authority/Respect" and "Fairness/Reciprocity".

Results and discussion

No explicit outlier treatment was applied; hence all the collected data was included in the statistical analysis. The hostility scores did not follow a normal distribution. Therefore, a two-way ANOVA on ranks (Kruskal-Wallis H test) was conducted to assess the significant difference in hostility perception between NPCs ($df = 3, p < .001$). No significant difference between appearances was found, while there was a significant difference between behaviors ($p < .001$) (see Figure 8). There was no significant interaction between appearance and behavior. A multiple linear regression was conducted to identify the predictors of the score of hostility. All the items taken together explained 47.9% of the variability of the hostility score ($R^2 = 0.479$). However, only the items “Friendly” ($\beta = -0.253, p = .004$) “Knowledgeable” ($\beta = 0.214, p = .051$) and “Harm/Care” ($\beta = 0.39, p < .001$) were significant predictors of hostility perception. A multiple linear regression with these predictors showed a prediction of 47% of the variability of the hostility score ($R^2 = 0.47$). A two-way ANOVA was conducted to assess the significant difference between appearance and behavior for the items “Friendly”, “Knowledgeable” and “Harm/Care”. No significant effect of appearance nor interaction between appearance and behavior were observed for the item Friendly, while a significant effect was found for the behavior ($p < .001$). No significant effect of appearance and behavior was found for the item “Knowledgeable”. Significant effects of appearance ($p = .015$), behavior ($p < .001$) and an interaction between Appearance and Behavior ($p = .02$) were found for the item “Harm/Care”. Finally, a linear regression with the scores of threat and aggressiveness was conducted to assess their impact on the hostility score. The two scores explained 48.7% of the variability of the final score ($R^2 = 0.487$). Only the score of aggressiveness was a significant predictor ($\beta = 0.690, p < .001$).

The results from Experiment 3 show that hostility perception is different for the four NPCs. This

TABLE 5
DESCRIPTIVE STATISTICS FOR HOSTILITY SCORES

	Not aggressive soldier	Aggressive soldier	Not aggressive civilian	Aggressive civilian
N	51	51	51	51
Mean	44.740	87.960	38.200	87.700
SD	33.245	15.585	36.006	25.463

difference is due to the variability in the aggressiveness of the behaviors. Moreover, friendliness, knowledge and harm’s perception were the main predictors of hostility attribution. Those results

follow findings from Experiment 2 and confirm the importance of behaviors for hostility attribution.

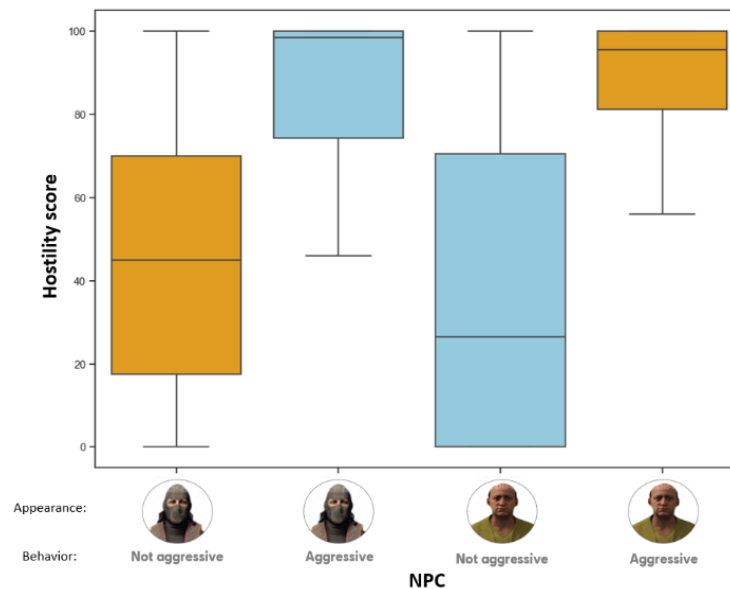


Fig. 6. Boxplots of NPCs’ hostility scores: boxplots’ colors represent NPCs’ congruence during the experiment (the blue color for congruent and orange for incongruent). The boxplots show the distribution of the hostility score by NPCs’ manipulated parameters, their median and interquartile range. Black error bars denote the standard error of the mean.

9. General Discussion

9.1 Summary

The results from these three experiments provide information about hostility evaluation in video game players. As the vast majority of participants were men, the results only generalize to this population. The approach adopted in this research was to study NPCs’ hostility by means of players’ evaluation of the threat of their appearance and the aggressiveness of their behaviors. It is all the more important in the case of a military game, as evaluating the hostility conveyed by NPCs is essential to establish a hierarchy of danger and to adopt the best strategy to progress in the game (e.g., starting to interact with less threatening NPCs as it might be easier to defeat them). Moreover, NPCs’ aggressive behaviors participate in the challenge of the game as they are at the core of the interactions with the main character. Game designers gradually increase difficulty regarding NPCs’

aggressiveness to create diversity and an impression of moving forward into the narrative. Each of the three experiments outlined the elements in the NPCs' design that enables players to assess their hostility, thus providing precise hostility predictors and an interpretation of players' perception. In the following, the three experiments' results are discussed separately, as each one made specific predictors of hostility emerge, before addressing limitations of the study and future research.

9.2 Threat of the appearance

The first experiment focused on the perception of threat conveyed by NPCs' appearance. Three main results were found: (i) there was a significant difference of threat evaluation between NPCs looking like civilians versus soldiers, (ii) there were significant differences of threat perception among soldiers, and (iii) the level of threat conveyed by the different NPCs rely on the morality and competence ascribed to them.

First, participants attributed different levels of threat to civilian versus soldier NPCs. This result might rely on participants' detection of stereotypical items from the NPCs' appearance (e.g., NPCs' shapes, clothes, or gears). For instance, the plain and neutral looks of civilian NPCs must have been implicitly associated with the harmless environment of everyday life, while the gears worn by the soldiers must have triggered stereotypes related to the military world, which is generally associated with danger and violence, thus with threat.

Second, a significant difference in threat evaluation was observed depending on soldiers' archetype. In particular, soldiers with covered faces and dark clothes were perceived as more threatening than soldiers with uncovered faces and light clothes. So, beyond the global appearance of an NPC allowing participants to implicitly categorize them as civilians or soldiers, specific visual cues appear to influence threat perception as well. This result is in line with [21] study, which outlined the visual cues influencing the perception of immorality within antagonists in video games, and among them covered faces and darker colors. Finally, three items used to evaluate the threat in NPCs' appearances turned out to be significant predictors of such ratings: "Competent", "Harm/Care" and "Authority/Respect". NPCs that were perceived as the most threatening were also perceived as the most competent, the most harmful and the most inclined to cause chaos and disorder, thus as the most immoral. The first predictor of threat perception, the item "Competent", comes from [32] perceived intelligence questionnaire. In Experiment 1, NPCs' appearances that were perceived as more threatening were perceived as more competent, which naturally follows

the context of the narrative. Civilians are expected to be less competent in a military game, thus less threatening than soldiers who evolve in familiar territory. Other predictors of threat perception were the items “Harm/Care” and “Authority/Respect” taken from the morality CMFQ-S scale, which reflect participants’ assessment of the perceived NPCs’ morality. These moral dimensions refer to the tendency of the character to cause harm, chaos, and disorder.

Therefore, Experiment 1 outlined that participants rely on the threat of NPCs’ appearance to recognize their role inside the narrative (i.e., civilians are less threatening than soldiers in a military game). This finding follows a multifactorial approach which suggests that hostile NPCs are not simply perceived on a single axis from good to bad. From the detection of specific visual cues, participants rated soldiers’ appearance on different levels of threat. Finally, the combination of ascribed competence and morality based on NPCs’ appearance seems to predict the participants’ anticipation of the NPCs’ abilities and intentions towards the main character.

9.3 Aggressiveness of the behavior

The second experiment studied players’ perception of aggressiveness through the observation of a pre-recorded interaction between the main character and an NPC. NPCs’ behaviors were manipulated to display the different sequences of actions found in the gaming experience (i.e., civilians’ and soldiers’ behaviors). Three main results were found: (i) there was a significant difference of aggressiveness evaluation between civilians and soldiers’ behaviors, (ii) among soldiers’ behaviors and (iii) aggressiveness evaluation relied on friendliness, knowledge, and morality ascriptions.

First, participants perceived different levels of aggressiveness in NPCs who looked the same but displayed either a typical civilian’s or soldier’s behavior. More precisely, a civilian’s behavior endorses a submissive stance (crouching) when confronted with an interaction, while soldiers try to defeat the main character (shooting). Subsequently, soldiers’ behaviors through their shooting activity usually entail negative consequences (i.e., the harm or death of the main character). Thus, participants evaluated the civilian behavior as less aggressive than the soldier’s one, while all NPCs displayed the same appearance.

Participants ascribed different levels of aggressiveness among soldiers’ behaviors. Soldiers’ behaviors were perceived as increasingly aggressive depending on their efficiency in defeating the main character and on their mobility. Among soldiers’ behaviors, some of them failed in defeating

the main character. For instance, some NPCs were not accurate with their weapon, their shooting resulting in a small loss in the main character's health points. NPCs that were able to kill the main character were perceived as more aggressive. Moreover, the behavior evaluated as the most aggressive was characterized by a reduction of the distance between the main character and the NPC's initial location (i.e., NPC rushing toward the main character). As described in proxemic theory, social spaces can be defined by the distance between the agents of a scene [36]. With this approach, NPCs and the main character would possess their own personal space, that is, the distance in which even a virtual agent looks like it feels comfortable. In the case of the most aggressive behavior, the main character's personal space can be perceived as violated. This invasion of the main character's personal space could thus explain the high amount of perceived aggressiveness. With this information, designers could make enhanced decisions when they plan the encountered enemies during a mission to convey the amount of intended aggressiveness. For instance, at the start of the game, the main character may encounter soldiers who are less accurate shooters and tend to stay in cover. However, as the main character progresses towards the end of their mission, they may come across soldiers with greater shooting accuracy and enhanced mobility, demonstrated by a reduction in the distance between them.

Lastly, significant predictors of aggressiveness evaluation were the items "Harm/Care", "Authority/Respect", "Fairness/Reciprocity", "Knowledgeable" and "Friendly". The first three are moral dimensions which refer to the ascription of interpersonal traits related to violence and authority (i.e. "This character would cause chaos and disorder", "This character would deny others their rights", "This character would physically hurt another person"). Again, this association between soldiers' behavior and morality ascription could be explained by players' mental representations of the military world and their video games' knowledge. In many military video games, soldiers are opponents of the main character, as they endorse an antagonistic stance through the improper nature of their acts (e.g., killing civilians) which triggers the main character's moral compass (e.g., soldiers oppressing civilians). Thus, NPCs' behaviors perceived as harmful, authoritarian, and unfair were also perceived as the most aggressive. Another predictor of aggressiveness was the item "Knowledgeable", taken from the perceived intelligence scale. This item refers to the perceived efficiency of NPCs to achieve their task. In the military world, knowledge would refer to different abilities and skills (e.g., weapon usage, personality, and emotional control) that are required for the military. In Experiment 2, overly aggressive NPCs were

perceived as more knowledgeable than less aggressive ones. The last predictor of aggressiveness perception was friendliness: the more aggressive NPCs were evaluated, the less friendliness they were ascribed. The results from this second experiment outlined a difference of aggressiveness evaluation within NPCs' behaviors. This difference was mainly explained by NPCs' role inside the narrative of the game (i.e., civilian versus soldier). Moreover, among soldiers' behaviors, NPCs who were efficient in defeating the main character and who violated their personal space were perceived as the most aggressive. Finally, predictors of aggressiveness evaluation suggest that aggressive behaviors are perceived as immoral, knowledgeable, and unfriendly.

9.4 Hostility within NPCs

Experiment 3 combined the evaluation of threat of appearances and aggressiveness of behaviors to assess hostility more holistically. Participants observed four interactions between the main character and an NPC. The NPCs could be either congruent or incongruent. For instance, an incongruent NPC displays a non-threatening appearance (i.e., a civilian appearance) and an aggressive behavior (i.e., the most aggressively perceived soldiers' behavior, the rusher), or the other way round. Directly after their observation, participants evaluated the hostility of the NPCs and rated their perception with the predictors of the threat and aggressiveness obtained through the previous experiments of this study. Two main results were found: (i) hostility was conveyed through aggressive behaviors rather than threatening appearances and (ii) Some predictors are similar for hostility and aggressiveness.

First, the results show that aggressive behaviors had a significant impact on participants' evaluation of hostility whereas there was no effect of threatening appearance. Creating incongruent NPCs can have interesting outcomes for the conception of video games as it can create surprise in players' minds. For instance, designers could manipulate the appearance and behaviors of NPCs inside a mission to create ambiguity and challenge without affecting players' evaluation of hostility.

In addition to hostility evaluation, the items "Friendly", "Harm/Care" and "Knowledgeable" were significant predictors of the participants' evaluation of hostility. In the context of military games, NPCs that were perceived as the most hostile were perceived as the most unfriendly, harmful, and knowledgeable. These three predictors were part of the ones found for aggressiveness in Experiment 2. Therefore, not only did participants perceive hostility when they also perceived aggressive behaviors, but both hostility and aggressiveness perception share some of the same

predictors.

10. Limitations and future research

The main limitations of this research, which open the road to future studies, touch on three areas: (i) participants, (ii) material and (iii) measures. First, the participants completing the studies belonged to Ubisoft user research database and they are known to be highly engaged players. To investigate whether the predictors of hostility differ between highly trained players and more naive ones, it would be interesting in future research to conduct the experiment in groups of players with different levels of gaming experience. Moreover, as most of the participants were men, it would be important for a full generalization of the results to conduct a comparison between genders on the evaluation of hostility in video games. Understanding the difference between genders on hostility evaluation could lead to a better approach when designing enemies in video games (i.e., conceiving NPCs' hostility levels based on the players' gender, which will allow refining the gaming experience). Second, our three experiments were conducted in the context of military video games. It would be interesting to investigate the extent to which the obtained results generalize to other contexts. For instance, enemies in a more cartoonish context might trigger different hostility perceptions, such as in Super Mario Odyssey, where enemies are non-humanoid agents. Third, in our studies, participants were watching videos of interactions with NPCs. However, it is possible that experiencing hostility differs between watching videos and interacting with hostile NPCs. Future research should investigate whether players observing versus directly engaging with NPCs perceive hostility differently. Finally, this study used explicit scales of measure which requires participants to consciously rationalize their often-unconscious perception. It would be interesting in future experiments to add indirect measures such as electro-physiological ones (e.g., heart rate, skin-conductance response) to investigate implicit evaluations in video games. This would allow comparison of the explicitly perceived (through questionnaires) and implicitly felt (through electro-physiological recordings) hostility, a direction which is currently being explored in our team.

11. Conclusion

Creating hostile characters is a crucial part of designing shooter video games. Indeed, enemies are part of the inherent challenges of the game and might represent most of the interactions in these contexts. Understanding how the factors affect the perception of the three personality traits of likability, perceived intelligence and morality can help designers in creating more engaging and immersive enemy characters in military shooter games. This knowledge can lead to the development of more captivating gameplay, where players are immersed in a challenging and realistic virtual environment that heightens their sense of excitement and enjoyment. By considering the interplay between visual design, behavioral design, and players' evaluation, game designers can create enemies that evoke the desired emotional responses and enhance the overall gameplay experience. To study hostility perception, three experiments were conducted. The first experiment gave a fresh understanding of players' evaluation of the threat conveyed by NPCs' appearance. Participants were able to ascribe competence and morality to characters based on cues from their appearance only (e.g., civilian- versus soldier-looking, gears or covered faces). The second experiment measured the perception of aggressiveness based on NPCs' behaviors and outlined the impact of proxemic behaviors on aggressiveness attribution. Perception of aggressiveness was predicted by several traits (i.e., friendliness, knowledgeable, immorality) and NPCs that were perceived as the most aggressive were those intruding the main character's personal space. The last experiment aimed to evaluate players' perception of hostility by manipulating both NPCs' appearance and behavior. While the threat of the appearance conveyed information about the role and intention of characters, only behaviors were significant to convey hostility. Moreover, hostility perception was predicted by three items which reflect the likability (friendly), the perceived intelligence (knowledgeable) and the morality (harm/care) of the character, in line with this study's hypothesis. This research thus gives a new understanding of how hostility can be conveyed in military games and suggests that behaviors are the main parameters to influence players' perception of NPCs.

ACKNOWLEDGEMENTS

Remi Poivet is funded by the French ANRT and the company Ubisoft.

References

- [1] W.-D. Liu, K.-Y. Chuang, and K.-Y. Chen, “The Design and Implementation of a Chatbot’s Character for Elderly Care,” in 2018 Int. Conf. System Science. and Engineering (ICSSE), New Taipei, Jun. 2018, pp. 1–5, doi: 10.1109/ICSSE.2018.8520008.
- [2] J. C. Barefoot, “Keeping Conflicting Findings in Perspective: The Case of Hostility and Health,” in *Mayo Clinic Proc.*, vol. 68, no. 2, pp. 192–193, Feb. 1993, doi: 10.1016/S0025-6196(12)60170-0.
- [3] J. C. Barefoot and I. M. Lipkus, “The assessment of anger and hostility,” in A. W. Siegman & T. W. Smith (Eds.), *Anger, hostility, and the heart*, Lawrence Erlbaum Associates, Inc, 1994, pp. 43–66.
- [4] W. G. Stephan and C. W. Stephan, “Intergroup Anxiety,” *J. of Social Issues*, vol 43, no. 3, pp. 157-175, 1985.
- [5] B. J. Bushman and C. A. Anderson, “Is it time to pull the plug on the hostile versus instrumental aggression dichotomy?,” *Psychol. Rev.*, vol 108, no. 1, pp. 273–279, 2001.
- [6] J. C. Barefoot, “Developments in the measurement of hostility,” in *Hostility, coping, & health.*, H. S. Friedman, Ed. Washington: American Psychological Association, 1992, pp. 13–31, doi: 10.1037/10105-001.
- [7] R. Blankendaal, T. Bosse, C. Gerritsen, T. de Jong, and J. de Man, “Are Aggressive Agents as Scary as Aggressive Humans?,” in *Proc. 2015 Int. Conf. Autonomous Agents and Multiagent Systems*, 2015, pp. 553–561.
- [8] T. Bosse, T. Hartmann, R. A. M. Blankendaal, N. Dokter, M. Otte, and L. Goedschalk, « Virtually Bad: A Study on Virtual Agents that Physically Threaten Human Beings », in *Proc. 17th Int. Conf. Autonomous Agents and MultiAgent Syst.*, Richland, SC, jul. 2018, p. 1258 1266.
- [9] K. Rogers, M. Aufheimer, M. Weber, and L. Nacke, “Towards the Visual Design of Non-Player Characters for Narrative Roles,” *Graphics Interface 2018*, 2018 pp. 154–161, doi: 10.20380/GI2018.21.
- [10] J. A. Reed and E. M. Blunk, “The influence of facial hair on impression formation,” *Social Behav. and Personality: an Int. J.*, vol. 18, no. 1, pp. 169–175, Jan. 1990, doi: 10.2224/sbp.1990.18.1.169.
- [11] M. Haake and A. Gulz, “Visual Stereotypes and Virtual Pedagogical Agents,” *Educational Technology and Society*, vol. 11, no. 4, pp. 1–15, 2008.

- [12] Y. Ferstl, M. McKay, and R. McDonnell, “Facial Feature Manipulation for Trait Portrayal in Realistic and Cartoon-Rendered Characters,” *ACM Trans. Appl. Percept.*, vol. 18, no. 4, pp. 1–8, Oct. 2021, doi: 10.1145/3486579.
- [13] R. Giusti, K. Hullett, and J. Whitehead, “Weapon design patterns in shooter games,” in *Proc. of the 1st Workshop on Des. Patterns in Games*, 2012, pp. 1-7.
- [14] C. Becker, S. Kopp, and I. Wachsmuth, “Simulating the Emotion Dynamics of a Multimodal Conversational Agent,” in *Tutorial and Research Workshop on Affective Dialogue Systems*, in *Affective Dialogue Systems*, vol. 3068, E. André, L. Dybkjær, W. Minker, and P. Heisterkamp, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 154–165, doi: 10.1007/978-3-540-24842-2_15.
- [15] K. Ruhland, C. E. Peters, S. Andrist, J. B. Badler, N. Badler, M. L. Gleicher, B. Mutlu, and R. McDonnell, “A Review of Eye Gaze in Virtual Agents, Social Robotics and HCI: Behaviour Generation, User Interaction and Perception: A Review of Eye Gaze,” *Computer Graphics Forum*, vol. 34, no. 6, pp. 299–326, Sep. 2015, doi: 10.1111/cgf.12603.
- [16] N. Sadoughi and C. Busso, “Speech-driven animation with meaningful behaviors,” *Speech Communication*, vol. 110, pp. 90–100, Jul. 2019, doi: 10.1016/j.specom.2019.04.005.
- [17] A. B. Loyall, “Believable Agents: Building Interactive Personalities,” Ph.D. dissertation, Carnegie Mellon University, 1997.
- [18] K. Kim, L. Boelling, S. Haesler, J. Bailenson, G. Bruder, and G. F. Welch, “Does a Digital Assistant Need a Body? The Influence of Visual Embodiment and Social Behavior on the Perception of Intelligent Virtual Agents in AR,” in *2018 IEEE Int. Symp. Mixed and Augmented Reality (ISMAR)*, Munich, Germany, Oct. 2018, pp. 105–114, doi: 10.1109/ISMAR.2018.00039.
- [19] T. L. Robbins and A. S. DeNisi, “A closer look at interpersonal affect as a distinct influence on cognitive processing in performance evaluations,” *J. of Applied Psychol.*, vol. 79, no. 3, pp. 341–353, Jun. 1994, doi: 10.1037/0021-9010.79.3.341.
- [20] B. Reeves and C. Nass, *The media equation: how people treat computers, television, and new media like real people and places* Choice Reviews. Cambridge, UK, 1996.
- [21] R. Pradantyo, M. V. Birk, and S. Bateman, “How the Visual Design of Video Game Antagonists Affects Perception of Morality,” *Frontiers in Computer Science*, vol. 3, 2021.
- [22] H. Warpefelt. “Cues and insinuations: Indicating affordances of non-player character using

- visual indicators,” in Proc. DiGRA 2015: Diversity of play: Games - Cultures – Identities, May. 2015. [Online]. Available: <http://www.digra.org/digital-library/publications/cues-and-insinuations-indicating-affordances-of-non-player-character-using-visual-indicators/>
- [23] M. Lee and C. Heeter, “What do you mean by believable characters?: The effect of character rating and hostility on the perception of character believability,” *J. of Gaming and Virtual Worlds*, vol 4, no. 1, pp. 81-97, 2012.
- [24] A. A. Long and N. Sedley, *The Hellenistic Philosophers: Volume 2, Greek and Latin Texts with Notes and Bibliography*. Cambridge University Press, 1987.
- [25] A. Eden, M. B. Oliver, R. Tamborini, A. Limperos, and J. Woolley, “Perceptions of Moral Violations and Personality Traits Among Heroes and Villains,” *Mass Communication and Soc.*, vol. 18, no. 2, pp. 186–208, Mar. 2015, doi: 10.1080/15205436.2014.923462.
- [26] T. Manninen and T. Kujanpää, “The value of virtual assets: The role of game characters in MMOGs,” *Int. J. of Bus. Sci & Appl. Manage.*, vol. 2, no. 1, pp. 21–33, 2007.
- [27] P. Vorderer, T. Hartmann, and C. Klimmt, “Explaining the enjoyment of playing video games: the role of competition,” in Proc. 2nd Int. Conf. Entertainment computing, 2003, pp. 1–9.
- [28] M. Grizzard, K. Fitzgerald, C. J. Francemone, C. Ahn, J. Huang, J. Walton, C. McAllister, and A. Eden, “Validating the extended character morality questionnaire,” *Media Psychology*, vol. 23, no. 1, pp. 107–130, Jan. 2020, doi: 10.1080/15213269.2019.1572523.
- [29] J. Haidt and C. Joseph, “The Moral Mind: How Five Sets of Innate Intuitions Guide the Development of Many Culture-Specific Virtues, and Perhaps Even Modules” in *Evolution and Cognition. The Innate Mind, Volume 3, 1st ed.*, P. Carruthers and S. Laurence, Eds. Oxford University Press New York, 2008, pp. 367–392, doi: 10.1093/acprof:oso/9780195332834.003.0019.
- [30] J. L. Monahan, “I Don’t Know It But I Like You The Influence of Non-conscious Affect on Person Perception,” *Human Communication Res.*, vol. 24, no. 4, pp. 480–500, Jun. 1998, doi: 10.1111/j.1468-2958.1998.tb00428.x.
- [31] C. Bartneck, T. Kanda, H. Ishiguro, and N. Hagita, “My robotic doppelgänger - a critical look at the Uncanny Valley,” in *RO-MAN 2009 – The 18th IEEE Int. Symp. Robot and Human Commun.*, 2009, p. 269–276, doi: 10.1109/ROMAN.2009.5326351.
- [32] C. Bartneck, D. Kulić, E. Croft, and S. Zoghbi, “Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of

- Robots,” *Int. J. of Social Robot.*, vol. 1, no. 1, pp. 71–81, 2009, doi: 10.1007/s12369-008-0001-3.
- [33] T. Koda and P. Maes, “Agents with faces: the effect of personification,” in *RO-MAN’96 – The 5th IEEE Int. Workshop Robot and Human Commun.*, 1996, pp. 189–194, doi: 10.1109/ROMAN.1996.568812.
- [34] C. Butcher and J. Griesemer, «The illusion of intelligence: The integration of AI and level design in Halo,» presented at the *Game Developers Conference*, San Jose, CA, March 21-23, 2002.
- [35] T. Bancroft, *Creating characters with personality*. New York: Watson-Guptill, 2006.
- [36] R. Gifford, “The experience of personal space: Perception of interpersonal distance,” *J. Nonverbal Behav.*, vol. 7, no. 3, pp. 170–178, Mar. 1983, doi: 10.1007/BF00986947.

CHAPTER 3: Breaking expectations in video games: the role of coherence and consistency in players' experience

Poivet, R.^{1,2}, Pelachaud, C.¹, & Auvray, M.¹ (in preparation). Breaking expectations: the role of coherence and consistency in players' experience.

¹ Sorbonne Université, CNRS, Institut des Systèmes Intelligents et de Robotique, ISIR, Paris, France

² Ubisoft Paris studio, Montreuil, France

Abstract — In video games, non-player characters (NPCs) have an essential role in shaping players' experiences, hence their design, comprising appearance and behaviors, influences players' expectations. Coherence and consistency between NPCs' appearance and behaviors during players' interaction contribute to maintaining their expectations, while violating these expectations aims to create surprise and engagement in the game experience. Additionally, players rely on their expectation to evaluate NPCs' perceived intelligence and believability - two dimensions associated with adoption and engagement toward artificial agents. Therefore, the impact of violating players' expectations on their evaluation of NPCs remains to be unveiled. Two experiments were conducted in a military shooter game. The first experiment investigated the impact of breaking players' initial expectations on the evaluation of NPCs' perceived intelligence and believability. The second experiment focused on the influence of coherence and consistency manipulation of NPCs' design on players' expectations and their evaluation of NPCs' perceived intelligence and believability. To investigate players' expectations a combination of perceptual, behavioral and physiological measures were used. Our results show that breaking players' expectations influenced their evaluation of NPCs, with coherent and consistent design reinforcing expectations and incoherent design challenging them. Participants' reactions to unexpected behaviors and their adaptability to different designs varied, leading to differences in perceived intelligence and believability ratings. A better understanding of how players form expectations and react to coherence and consistency manipulation can help tailor NPCs' design to enhance player engagement and immersion in the narrative.

Index Terms — Affective issues in enhancing machine/robotic intelligence, Cognition, Entertainment, Games

Keywords — Video games, Non-playable character, violation of expectations, perceived intelligence, believability

1. Introduction

In video games, human players control a main character who interacts with non-player characters (NPCs) that populate the game. NPCs are characterized by their design, which is a combination of features chosen by game designers to communicate NPCs' role in the narrative. NPCs can assume various roles, such as allies or enemies. Game designers often use two features to convey NPCs' roles in the narrative: appearance and behaviors.

Appearance refers to NPCs' virtual embodiment, consisting of visual cues aiming at activating stereotypes in the players' minds, allowing them to anticipate their behaviors. NPCs' behaviors consist in sequences of predefined actions specific to their roles in the narrative [1]. Players form expectations when they encounter NPCs based on the coherence between their stereotypes triggered by NPCs' appearance and their observation of NPCs' actions in the narrative [2]. Players' expectations can be reinforced by the consistency of NPCs' design in their following interactions with similar NPCs in the narrative [3]. For instance, in a military shooter game, encountering an aggressive soldier at the start of the game experience can lead to hostile expectations when another soldier appears with a similar appearance. If the newly encountered soldier confirms players' expectations, then meeting a third soldier might be easier to predict hostility with higher certainty.

During their interactions, players evaluate NPCs' perceived intelligence and believability as it contributes to their engagement and appreciation of the overall game experience [4, 5]. Perceived intelligence is defined by the player's evaluation of NPCs' abilities to achieve their purpose in the narrative, encompassing two dimensions: behaviors' understandability and performance [6]. Believability, on the other hand, is defined as the gap between players' expectations and their observations in the game, impacting their engagement in the narrative and immersion [5]. These two aspects are crucial in delivering engaging narrative experiences. Therefore, considering players' evaluation when designing NPCs is important. However, to enhance the narrative

experience designers introduce twists in the game [7]. Using NPCs as a twist, game designers manipulate the coherence and the consistency between their appearance and behaviors to create surprise in the players' minds. For example, in the western action-adventure game *Red Dead Redemption 2* [8], the main character can encounter NPCs who have a non-threatening appearance, asking for help along the road but suddenly turn out to be aggressive toward the main character if players decide to stop and help them. Breaking human expectations can have both positive and negative aspects [9]. In the example of *Red Dead Redemption 2*, it aims to convey a positive break of expectations as it introduces challenges in the game experience and conveys a sense of hostility that players could expect from the social environment in a game depicting a western setting. However, the impact of violating players' expectations on their evaluation of NPCs' perceived intelligence and believability is yet to be unveiled. As part of their evaluation, players rely on their expectations and knowledge of NPCs to ascribe them intelligence and believability. We hypothesize that breaking players' initial expectations will decrease their evaluation of NPCs' perceived intelligence and believability. Our second hypothesis posits that after multiple interactions with NPCs that break players' initial expectations, new expectations may form, leading to improved evaluations of perceived intelligence and believability.

In this study, two experiments in a military shooter game were conducted. The first experiment investigated the violation of players' expectations on their evaluation of NPCs' perceived intelligence and believability. The second experiment focused on the impact of consistency and coherence manipulation of NPCs' design on players' expectations using questionnaires, behavioral and physiological measures, and assessing the influence of expectation on players' evaluation of NPCs' perceived intelligence and believability.

2. Background

2.1 Designing NPCs: the role of appearance and behaviors

Creating NPCs involves setting parameters that are connected and consistent throughout the game experience [1]. For example, in shooter games, enemies are subdivided into archetypes defined by their appearances, the weapons they use, and their moving abilities. Consequently, the design of an NPC refers to the choice made by game designers to convey their intended game experience.

Appearance

NPCs' appearance refers to their virtual embodiment as perceived by players in the game. The visual cues chosen to create NPCs' appearance aim to activate stereotypes in the players' mind [10]. Based on these stereotypes, players can understand the social roles of NPCs in the game [11]. For instance, an NPC dressed in military gear in a military shooter game might be evaluated as more threatening compared to another NPC wearing casual clothing. In this scenario, the military gear triggers stereotypes linked to violence and implies potential aggressiveness toward the main character. Thus, as part of their identification of NPCs' roles, players anticipate NPCs' behaviors and abilities. In military shooter games, the type of weapon carried by an enemy provides valuable clues about their abilities. For example, an NPC wielding a shotgun would likely attempt close-range attacks, while one armed with a sniper rifle would engage from a distance. Consequently, players can use these cues to make informed decisions during the game, such as whether to interact with an NPC or not. Additionally, the context of the game plays a crucial role. In a zombie shooter game, allies may possess military gear in their appearance, but players do not evaluate them as enemies due to the specific context of the game. Other factors, such as the shape or facial features of the NPCs, can also be crucial for players to anticipate their interactions [12]. In summary, appearances play a crucial role in players' video game experience as they activate stereotypes in players' minds and players expect NPCs' behaviors to be aligned with their anticipation during the interaction.

Behaviors

NPCs' behaviors consist of predefined sequences of actions that shape how they interact with others. These behaviors can vary, ranging from simulating casual activities to attacking the main character. However, the nature of their actions is constrained by their roles in the narrative, meaning that friendly NPCs cannot attack the main character, as it is not their intended purpose. For enemies, the type of weapon they carry shapes their behaviors, as illustrated in the earlier comparison between shotguns and sniper rifles. Moreover, in shooter games, the aggressiveness of enemies can be manipulated by other parameters, such as their shooting precision or the damage they can inflict to the main character. Therefore, during the creation of NPCs, their behaviors are interconnected with their appearance, following a consistent and coherent design along the narrative. Consistency refers to the alignment over time between appearance and behaviors, ensuring that NPCs' actions

match their visual cues and role in the game [1]. On the other hand, coherence pertains to the logic of the association based on players' stereotypes, ensuring that NPCs' behaviors make sense within the larger context of the game world [13].

The objective of maintaining both consistency and coherence between these two features is to enable players to form accurate expectations of enemies or allies and, as a result, enhance the overall game experience [14]. When NPCs' actions align with their appearance and fit logically within the game world, players can better immerse themselves in the narrative and make informed decisions throughout the gameplay.

2.2 Breaking players' expectations

During their game experience, players can learn associations between NPCs' appearance and behaviors to predict new encounters. To add complexity to the narrative and to create surprise in players, game designers can manipulate the design of NPCs so that their appearances and behaviors are unexpected to the participants.

How do we form and break expectations?

Players form lasting beliefs about NPCs' behaviors, known as expectations, through a cognitive process that involves gathering information from previous experiences, knowledge, and contextual cues within the game [15]. Drawing from their past video game experiences, players are familiar with the limitations and predefined behaviors of NPCs. Consequently, they expect NPCs to exhibit stereotyped sequences of actions that confirm their roles [10]. When encountering enemies, players expect them to engage in attacking the main character until they are defeated. In the game experience, the main character may come across multiple NPCs with identical designs in appearance and behaviors. Players hold explicit memories of their interactions with similar NPCs which allows them to adapt to various situations and strengthen the certainty of their decision-making along the narrative. This mechanism relies on the human tendency to reinforce beliefs when events happen as expected, validating their expectations, even when the events occur randomly [3]. Thus, when events do not conform to their beliefs, it violates players' expectations.

Violation of players' expectations refers to the introduction of unanticipated actions or events in the game experience [3]. To break players' expectations, game designers can manipulate both coherence and consistency of NPCs' design. On the one hand, incoherent NPCs' design would refer

to the intended mismatch between NPCs' appearance and behaviors. For instance, an incoherent association between appearances and behaviors convey false information for players' anticipation. Violation of expectations elicit compensatory behaviors reflected as exploratory behaviors seeking for explanation [17]. For instance in *Ready or not* [18], aggressive NPCs with a civilian appearance can be explained by the context, and the challenge of the game (i.e., identifying threats between NPCs is the inner challenge of the game). On the other hand, game designers can manipulate the consistency of NPCs' design, which means altering the occurrence of a specific association between an appearance and behaviors in a game experience (e.g., an NPC with given appearance and behaviors display the opposite type of behaviors). Altering consistency can affect players' expectations similarly to manipulating NPCs' coherence. [19] measured participants' physiological responses toward inconsistent breaks of their expectations in a computer game. The experiment used a cyberball game where participants' characters had to throw a ball at NPCs. Researchers manipulated participants' ability to throw the ball by occasionally overruling their actions. The aim was to assess the impact of consistency manipulation on the expectation of motor control. The results indicated that inconsistent expectations triggered physiological responses similar to aversive events. In a game experience, such manipulations can have a positive impact as it increases the inner challenge of the game. In *Ready or not*, as civilians can be either aggressive or not, players have to raise their attention to make quick and accurate decisions to shoot at NPCs.

The impact of breaking expectations

Breaking expectations influences individuals' cognitive and emotional responses, as suggested by the ViolEx model [20]. When unexpected events violate humans' expectations, it triggers a cognitive process of reassessment and adjustment. The ViolEx model proposes three coping processes in response to expectation violations: accommodation, assimilation, and immunization, with the aim of maintaining individuals' understanding of their surroundings. Accommodation involves adjusting beliefs after observing the outcome of unexpected events, while assimilation entails avoiding events that challenge initial expectations, and immunization involves reducing the relevance of the violation or reframing the meaning of the initial expectation. The selection of the coping strategy is influenced by individual factors such as personality traits, cognitive abilities, or past experiences [21, 22].

The impact of breaking expectations can lead to cognitive flexibility and adaptability, as individuals adjust their beliefs based on new information [23], but it can also trigger emotional responses such as frustration or surprise [24]. Violation of expectations induces an appraisal process that may be moderated by the reward of the violation [25]. Positive and negative attributions to the violations of expectations have been suggested by Burgoon to distinguish more or less favorable outcomes. Thus, the nature of the violation affects how humans evaluate unexpected events [25]. Game designers may manipulate the design of NPCs to surprise players and convey positive violations of expectations, aiming to create twists and interest in the narrative [26]. However, the impact of such manipulation on players' evaluation of NPCs remains unclear.

3. How to measure perceived intelligence and believability of NPCs?

In video games, players' positive evaluation of NPCs' perceived intelligence and believability is associated with the narrative's enjoyment, motivation to interact and immersion [5, 27]. Perceived intelligence refers to players' evaluation of NPCs' abilities to interact with their surroundings [6]. To evaluate NPCs' abilities to interact, two dimensions are important: their understandability and performance [28]. Both dimensions are based on players' understanding of the NPCs' purposes and efficiency in achieving their goals in the game experience. As a result, players' evaluation of NPCs' perceived intelligence is influenced by their expectations regarding the NPCs' roles in the narrative. For instance, allies are designed to assist the main character, while enemies are meant to pose challenges in the game experience. Therefore, the perceived intelligence of enemies would rely on their abilities to defeat the main character in an understandable manner. In this case, factors like enemies' strength or amount of health points can affect players' evaluation of NPCs' perceived intelligence, as they contribute to increased challenges in the game [4]. To investigate users' evaluation of artificial agents' perceived intelligence, [6] proposed an evaluation scale based on five semantic items: Incompetent/Competent, Ignorant/Knowledgeable, Irresponsible/Responsible, Unintelligent/Intelligent, Foolish/Sensible. The scores on each item are used to assess the two dimensions of understandability and performance. Game designers can benefit from using such a measure, as the scale conveys a deeper understanding of how NPCs' intelligence is perceived by players, allowing them to make the necessary adjustments to align with their intended game experience.

Aside from intelligence evaluation, players must believe in their interactions with NPCs to feel immersed and engaged in the narrative [5]. As NPCs possess predefined behaviors to simulate their interactions with the main character, it is essential to understand how their design conveys believability from players' perspective. Since players evaluate NPCs based on their expectations of how they should behave, believable agents should closely correspond to players' expectations during interactions (Loyall, 1997). Consequently, players' identification of NPCs' roles can influence their evaluation, as they will expect these agents to serve different purposes (e.g., interactions will differ when occurring with allies or enemies). [29] proposed a scale composed of various dimensions that play crucial roles in players' evaluation of believable characters in narrative experiences. Using the items listed below, players can indicate their evaluation of the believability of any NPCs (Where <X> is replaced by the evaluated NPC).

- Awareness: < X > perceives the world around him/her.
- Behavior understandability: It is easy to understand what < X > is thinking about.
- Personality: < X > has a personality.
- Visual impact: < X >'s behavior draws my attention.
- Predictability: < X >'s behavior is predictable.
- Behavior coherence: < X >'s behavior is coherent.
- Change with experience: < X >'s behavior changes according to experience.
- Social: < X > interacts socially with other characters.

This understanding of how players perceive NPCs in terms of believability and intelligence provides game designers with valuable insights into the strengths and weaknesses of their characters across various dimensions that contribute to engaging interactions. By using these scales, game designers can systematically iterate on the parameters they manipulate to precisely convey the desired game experience. This iterative approach empowers designers to fine-tune NPC

behaviors and appearances, aligning them more effectively with player expectations and ultimately leading to the creation of NPCs that offer a richer and more immersive gaming experience

4. Experiments

As seen above, players form expectations of NPCs before they interact with them. These expectations are based on knowledge, past experiences, and stereotypes triggered by NPCs' appearance. During interactions, players use NPCs' behaviors as feedback to validate their expectations. Game designers can manipulate the coherence and consistency of NPCs' design to influence players' expectations, allowing for the creation of unexpected events to deepen the narrative. However, the impact of such manipulation on players' evaluation of NPCs remains to be revealed, which is our aim here.

As players immerse themselves in the narrative, they evaluate NPCs' intelligence and believability based on their expectations. We hypothesize that the coherence manipulation of NPCs' design initially decreases players' evaluation as it violates their expectations. Additionally, manipulating the consistency of NPCs' design affects players' experience and evaluation of NPCs. To investigate these violations in players' experience, two experiments were conducted in the context of a Ubisoft military shooter game. The first experiment investigated the influence of coherence between NPCs' appearance and behaviors on players' initial evaluation, while the second experiment investigated the impact of consistency and coherence on players across multiple interactions in game sessions. The scales of perceived intelligence and believability were employed to measure players' evaluations of NPCs. Players' game experience was assessed through a variety of dimensions, encompassing their in-game behaviors, physiological activity, and expectations regarding NPCs' appearances across multiple game sessions.

5. General Methods

Materials and Procedure

The two experiments were set in a military shooter game environment. The NPCs used in the experiments were selected from Ubisoft's video game *Ghost Recon Breakpoint* [30]. The first

experiment was conducted online and involved participants watching four videos depicting interactions between a main character and NPCs with manipulated designs. The second experiment took place at Ubisoft Paris and included video game sessions preceding the viewing of the same video used in Experiment 1. During the game sessions, questionnaires about their expectations of NPCs, behavioral and physiological measures were used to investigate participants' interactions with NPCs and the evolution of their expectations. In both experiments, participants assessed their perception of NPCs featured in the four videos using the Perceived Intelligence questionnaire [6] and the Believability questionnaire [29].

Creation of NPCs' design

To create different NPCs' design for the experiments, appearances and behaviors of soldiers and civilians were used and manipulated to convey either coherent or incoherent design of NPCs. In military shooter games, NPCs with a civilian appearance are typically associated with innocent characters, while soldiers are expected to be enemies and display aggressiveness toward the main character. These behaviors were categorized as either submissive or aggressive. Submissive behaviors were defined as NPCs getting on their knees when the main character approached them. Conversely, aggressive behaviors were characterized by NPCs shooting at the main character while rushing toward them. The manipulation of NPCs' coherence was achieved by adjusting these two parameters; for example, an aggressive civilian would be considered incoherent.



Fig. 1. Appearance of NPCs and their typical behaviors in military shooter games: On the left, civilian appearances are defined by their casual clothes and non-muscular shape. Their behavior is

to crouch and surrender when the main character reaches them (as shown graphically by the blue silhouette). On the right, soldiers' appearance is characterized by their muscular shape and military gear, with their faces covered and wearing dark clothes. Their behavior is to rush toward the main character and shoot at them (displayed by the red silhouette).

Videos for participants' evaluation

Four videos displaying coherent and incoherent NPCs design were created. Participants evaluated NPCs' perceived intelligence and believability observed in each video. In the videos, the camera follows the main character moving inside a corridor, and an NPC is positioned at the end of it. The video pauses after the interaction, with outcomes varying between the main character being killed and the NPC surrendering. For example, a coherent video would depict an NPC with a soldier's appearance and aggressive behavior while an incoherent one would depict a submissive soldier (see Fig. 2).



Fig 2. Screenshot of the videos illustrating the interaction between the main character and the four NPCs, the characters furthest away in the images. The two pictures on the top depict an NPC with a civilian appearance behaving non-aggressively (left) and aggressively (right) toward the main character. At the bottom, an NPC with a soldier appearance follows the same dichotomy. Note that for each image, watermarks are automatically generated and required by Ubisoft.

6. Experiment 1

Participants

One-hundred and two participants completed the experiment (ninety-one men, ten women and one other, mean age = 32.156 years, SD = 8.570). Participants were volunteer players registered at the Ubisoft User Research Laboratory platform (i.e., a mailing platform where players can deliberately enter their information to participate in Ubisoft's research). The gender distribution may be attributed to the self-selection of participants already familiar with the military shooter game setting, potentially introducing a gender bias in their motivation for this specific type of game.

Procedure and materials

All participants were contacted by email and were provided with information about the content of the research. The email informed participants of the free nature of their participation and that they can stop at any time they want. The experiment took approximately 15 minutes to complete and was conducted in accordance with the 1992 Declaration of Helsinki. Participants were instructed about the context of the game (i.e., a military shooter game) prior to watching the videos. Participants watched the four videos in a random order and evaluated the NPCs using the items of the scales for perceived intelligence and believability.

Results and discussion

Participants indicated their perception of the four NPCs using the items of perceived intelligence and believability's scales. In order to maintain the representativeness of the dataset, no explicit outlier treatment was applied, hence all the collected data was included in the statistical analysis. A two-way ANOVA using the factors "Appearance" (soldier or civilian) and "Coherence" (coherent or incoherent design) was conducted on the participants' ratings of the scales' items. There was a main effect of the factor "Appearance" on the items "Visual impact" ($df = 404$, $p = .001$, $\eta^2 = 0.023$) and "Behavior coherence" ($df = 404$, $p = .003$, $\eta^2 = 0.016$) of the believability scale. Participants rated the behaviors of soldiers as drawing more attention compared to civilians' ones. However, civilians' behaviors were perceived as more coherent than soldiers' ones. There was a main effect of the factor "Coherence" on the items "Competent" ($df = 404$, $p = .003$, $\eta^2 = 0.019$), "Intelligent" ($df = 404$, $p < .001$, $\eta^2 = 0.030$), "Knowledgeable" ($df = 404$, $p = .003$, $\eta^2 = 0.019$), "Sensible" ($df = 404$, $p < .001$, $\eta^2 = 0.111$), "Responsible" ($df = 404$, $p < .001$, $\eta^2 = 0.056$) from the perceived intelligence scale. Incoherent NPCs were perceived as less intelligent than coherent ones. There

was a main effect of the factor “Coherence” on the items “Awareness” ($df = 404, p = 0.024, \eta^2 = 0.012$), “Behavior understandability” ($df = 404, p < .001, \eta^2 = 0.093$), “Predictability” ($df = 404, p < .001, \eta^2 = 0.239$) and “Behavior coherence” ($df = 404, p < .001, \eta^2 = 0.222$) from the believability scale. The manipulation affected participants’ evaluation of the readiness of NPCs’ behaviors, as coherent NPCs received higher ratings on items associated with the NPCs’ understandability. There was a significant interaction of the two factors on the items “Competent” ($df = 404, p < .001, \eta^2 = 0.128$), “Knowledgeable»” ($df = 404, p < .001, \eta^2 = 0.124$), “Sensible” ($df = 404, p < .001, \eta^2 = 0.033$), “Responsible” ($df = 404, p = .002, \eta^2 = 0.022$) from the perceived intelligence scale and the items “Awareness” ($df = 404, p < .001, \eta^2 = 0.044$), “Personality” ($df = 404, p = .011, \eta^2 = 0.016$), “Visual impact” ($df = 404, p < .001, \eta^2 = 0.085$), “Predictability” ($df = 404, p = .036, \eta^2 = 0.008$) and “Behavior coherence” ($df = 404, p = .048, \eta^2 = 0.007$) from the believability scale.

Experiment 1 revealed a significant effect of the factors Appearance and Coherence on participants’ ratings of perceived intelligence and believability (see Fig. 3). The results suggest that participants formed expectations of NPCs’ behaviors based on the evaluation of their appearance. In a shooter game with a coherent design, soldiers are enemies, while civilians are neutral or friendly NPCs. As a result, participants evaluated the NPCs based on their knowledge and expectations of a coherent NPCs’ design, which led them to expect aggressive behaviors from soldiers and neutral or non-aggressive behaviors from civilians.

During the experiment, participants evaluated incoherent NPCs as less intelligent and believable compared to their coherent counterparts. This finding indicates that participants had a more positive evaluation of coherent NPCs, as their behaviors matched their expectations. These results align with the definition of believability introduced by [5] and the dimension of understandability of perceived intelligence presented by [6]. Participants observing incoherent NPCs were surprised and unable to accurately explain their behaviors, leading to a decrease in their evaluation of NPCs’ believability and perceived intelligence compared to the coherent ones.

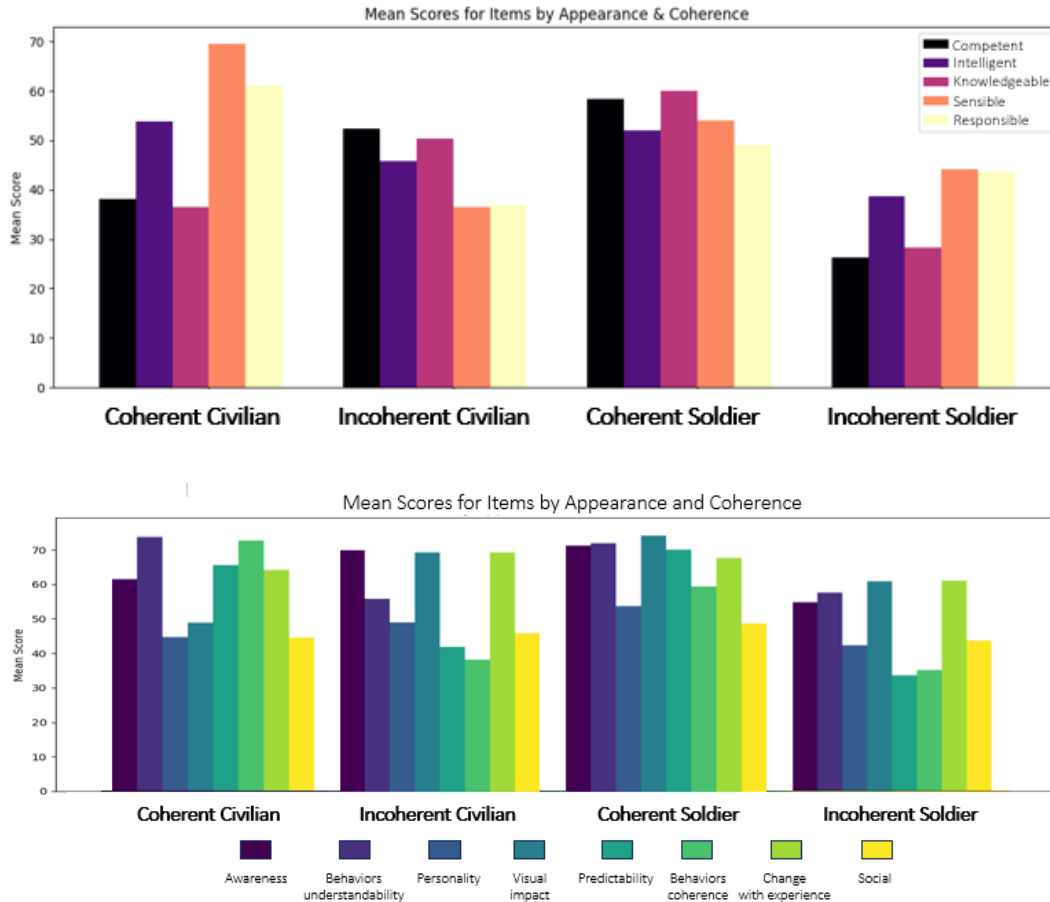


Fig. 3. Bar Charts of participants' ratings on items from the perceived intelligence (top) and believability (bottom) scales, based on the interaction between coherence and NPCs' appearance.

7. Experiment 2

Participants

Ninety-one French employees of Ubisoft participated in the experiment, comprising fifty-six men and thirty-five women with an average age of 27.5 years (SD = 6.197). Participants rated their frequency in playing video games and military shooter games using scales ranging from “0 - I never play” to “100 - I play every day”. There was a gender difference in military shooter game habits, with women reporting a mean of 15.37 (SD = 14.81; barely play to play sometimes) and men reporting a mean of 45.57 (SD = 31.08; play often to play very often), while no difference was found in video game habits (mean = 76.57, SD = 24.05, play almost every day). All participants

were contacted via email and provided with information regarding the general purpose of the research. The email emphasized the voluntary nature of participation, the different measures recorded and the option to withdraw at any time. The experiment lasted approximately 50 minutes and was conducted following the principles outlined in the Declaration of Helsinki.

Procedure and materials

The experiment consisted of completing two game sessions and watching four videos. Participants evaluated NPCs' perceived intelligence and believability after each video similarly to Experiment 1. To explore participants' expectations, questionnaires, behavioral and physiological measures were used during their game sessions (see Fig. 4). Information about game sessions and the measures recorded are detailed below.



Fig. 4. Illustration of the coherent and consistent NPCs' design procedure: Participants complete two military missions before watching videos, with questionnaires, behavioral, and physiological measures recorded during these sessions.

Game sessions

The experiment consisted of two game sessions. Participants were engaged in a military mission and were instructed to shoot at NPCs they identify as enemies with high certainty. The main objective for the main character controlled by the participants was to reach a specific location in the virtual environment. Along their way, ten NPCs were strategically placed to depict either innocent or enemies. The two game sessions differed in the order of interaction with NPCs in the

virtual environment. For instance, the first session might start with an enemy while in the second session, it is an innocent one in this location. Three variations of NPCs’ design were used and participants were randomly assigned to one of them (see Fig. 4). Group 1 had a consistent and coherent NPCs design, defined as encountering only aggressive soldiers (enemies) and submissive civilians (innocent). Group 2 had a consistent and incoherent NPCs design, soldiers being submissive and civilians being aggressive. In Group 3, NPCs were inconsistent, alternating between coherent and incoherent designs, allowing soldiers and civilians to be either aggressive or submissive.



Fig. 5. The three designs created for the game sessions. The red and blue figures illustrate the behaviors of the NPCs, either surrender (blue figure) or attack (red figure). On the left, Group 1: coherent design, civilians surrender when the main character approaches while soldiers try to defeat them. In the center, Group 2: incoherent design, the strict opposite of the coherent design. On the right, Group 3, inconsistent design, both soldiers and civilians exhibit the same behaviors, occasionally rushing toward the main character or surrendering when approached.

Questionnaires

Participants indicated their intention to shoot at NPCs (“I would shoot at this NPC in the context of a military shooter game: Yes or No”) based on their appearance (i.e., one appearance of a soldier and one of a civilian). Additionally, they rated the certainty of their intention to shoot ranging from “0 - I’m not certain at all” to “100- I’m totally certain”. This questionnaire was presented before the first session, between and after the two game sessions.

Behavioral measures

During the game sessions, participants facing enemies have to defeat them to progress in the virtual environment. To reflect their ability to identify NPCs as enemies, the distance between the main character and defeated enemies in the virtual environment was measured.

Physiological measures

During participants' interactions with NPCs, autonomous activities were recorded to assess their reaction in response to NPCs' behaviors. Skin conductance responses (SCR) were measured, focusing on amplitude and rise time during each interaction. These SCR measurements were obtained from sensors placed on participants' left hand, specifically on the ring and little fingers. To ensure consistency, participants were instructed to use only their thumb, index, and middle fingers while engaging with an Xbox controller. Additionally, their heart rates were recorded using Photoplethysmography (PPG). PPG data, capturing heart rate information, was collected by placing a sensor on participants' left earlobe. To ensure accurate readings, participants were instructed to remove any earrings, and precautions were taken to secure loose hair that might interfere with the sensor's contact with the skin. These measures aimed to minimize potential sources of interference and enhance the reliability of the heart rate measurements throughout the study.

Results and discussion

In this experiment, participants completed three questionnaires regarding their decision to shoot NPCs based on their appearance, engaged in two game sessions, and watched four videos. This section will first investigate the results of the three questionnaires. Then, the behavioral analysis of participants during their game sessions, as well as their SCR and heart rate, will be analyzed. Lastly, participants' ratings of the four videos will be examined.

Questionnaires' analysis

Regarding participants' responses to the questionnaires, chi-squared tests were conducted to analyze their decisions to shoot based on NPCs' Appearances and their respective game design groups. Before their game sessions, there was no significant difference between groups for both soldier ($\chi^2 = 1.042$, $p = .594$) and civilian ($\chi^2 = 0.337$, $p = .845$) appearances. Participants were expecting to shoot at soldiers and avoid civilians. After the first game session, significant

differences were observed between groups for both soldier ($\chi^2 = 17.838$, $p < .001$) and civilian ($\chi^2 = 8.128$, $p = .017$) appearances. These differences were also observed after the two game sessions for both soldier ($\chi^2 = 25.082$, $p < .001$) and civilian ($\chi^2 = 14.155$, $p < .001$) appearances. Following their game sessions and their interaction with the manipulated NPCs' design, participants' decision to shoot at NPCs was influenced, suggesting an evolution of their expectation about enemies. A repeated-measure ANOVA was performed to compare the effect of the factor Group of game sessions on participants' certainty to shoot at soldiers and civilians. For participants' certainty to shoot at soldiers, a Mauchly's test of sphericity on the repeated-measure ANOVA indicated that the assumption of sphericity had been violated ($\chi^2(2) = 8.327$, $p < 0.001$). Thus, a Greenhouse-Geisser correction was applied, revealing a significant effect of their group of game sessions on participants' certainty to shoot at soldiers ($df = 3.665$, $p < 0.001$, $\eta^2 = 0.063$). Post-hoc analysis using the Holm-Bonferroni method revealed significant differences between Groups 1 and 3 ($p < 0.001$) and Groups 2 and 3 ($p = 0.009$) after the two game sessions. Participants from Group 1, with the consistent and coherent NPCs' design were more certain of their decision to shoot at soldiers compared to participants from Group 3, with an inconsistent NPCs' design. For participants' certainty to shoot at civilians, a Mauchly's test of sphericity on the repeated-measure ANOVA indicated that the assumption of sphericity had not been violated ($\chi^2(2) = 2.551$, $p = 0.279$). The analysis revealed a significant effect of their group of game sessions on participants' certainty to shoot at civilians ($df = 4$, $p = 0.017$, $\eta^2 = 0.033$). Post-hoc analysis using the Holm-Bonferroni method showed significant differences between Groups 1 and 3 ($p < 0.001$) and Groups 2 and 3 ($p = 0.004$) after the two game sessions. Participants in Group 3 rated their certainty toward shooting or not at civilians lower compared to the other two groups after their game sessions.

Behavioral analysis

For the behavioral analysis, the distance (in meters inside the virtual environment) between the participants' characters and the defeated enemies was analyzed. A two-way ANOVA was performed to examine the effects of the factors Group of game sessions and Session's order on the distance to defeat enemies (see Fig. 6). There was a main effect of the Group of game sessions on the measured distance ($df = 2$, $p < 0.001$, $\eta^2 = 0.014$). Post-hoc analysis indicated that the distance between the participants' characters and enemies was larger in Group 1 compared to participants in Group 2 ($p = 0.003$) and Group 3 ($p = 0.003$). However, there was no significant difference between Groups 2

and 3 ($p = .999$). Furthermore, there was a main effect of Session's order on the distance to defeat enemies ($df = 1, p < .001, \eta^2 = 0.016$). The distance between participants' characters and defeated enemies was significantly larger in the second session. There was no significant interaction between the factors Group of game sessions and Session's order ($df = 2, p = .411, \eta^2 = 0.002$).



Fig. 6. Evolution of the distance to defeat enemies in participants' game sessions between the first and second sessions, based on group of game sessions. Group 1 is represented in blue, Group 2 in red, and Group 3 in purple.

Physiological analysis

A two-way ANOVA on participants' heart rate (beats per minute) did not reveal any significant effect neither of Group of game sessions ($p = .345$) nor of Session's order ($p = .991$) on participants' mean heart rate. For participants' SCR, two data points were analyzed: amplitude and rise time of the SCR during each interaction with NPCs. A two-way ANOVA was performed to examine the effects of the factors Group of game sessions and Session's order on SCR's amplitude and rise time (see Fig. 7). There was a main effect of the factor Group of game sessions on the signal's amplitude ($df = 2, p < .001, \eta^2 = 0.018$). Post hoc analysis using the Holm-Bonferroni method indicated a significant difference between Groups 2 and 3 ($p < .001$), but no significant difference between Groups 1 and 2 ($p = .053$) or between Group 1 and 3 ($p = .074$).

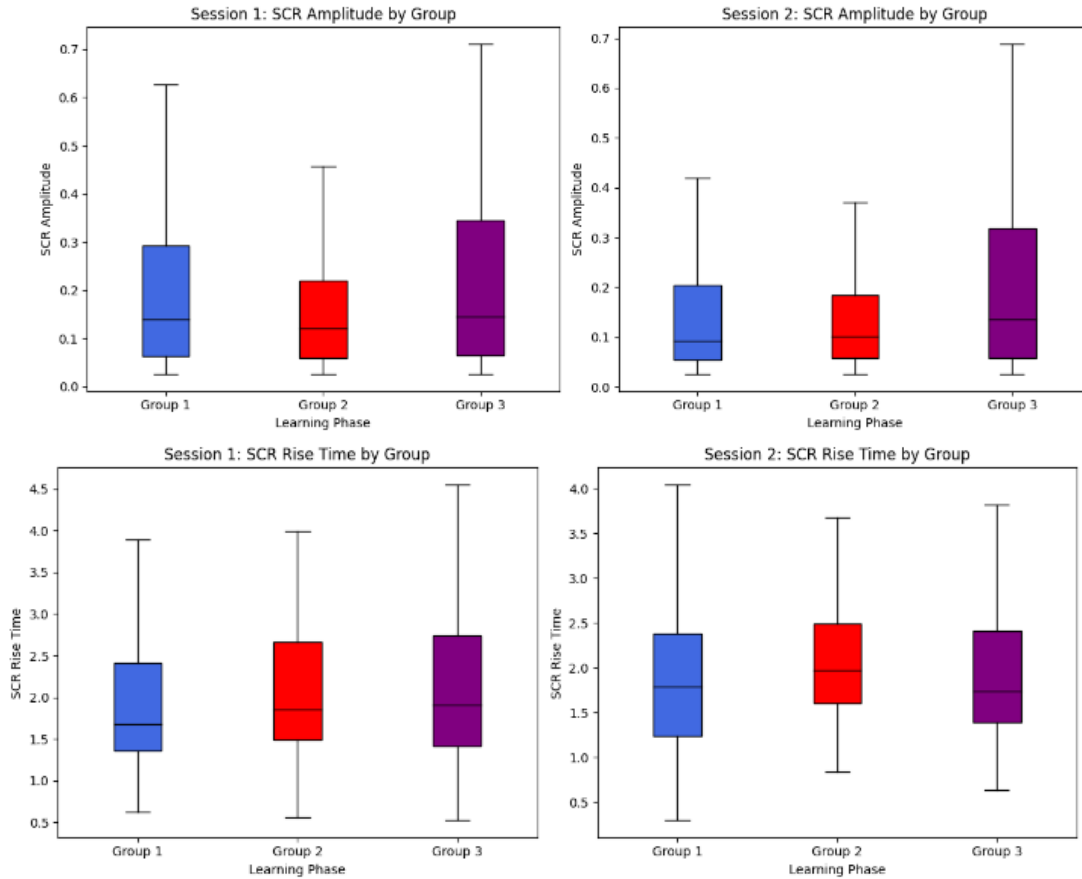


Fig. 7. Boxplots of participants' skin conductance response during their interaction with NPCs: (Top) Amplitude (top) and Rise time (bottom) of their response by session order.

Participants in Group 3 had significantly higher responses during their interaction with NPCs compared to the other two groups. Additionally, there was a main effect of Session's order on the SCR's amplitude ($df = 1$, $p = .015$, $\eta^2 = 0.006$). Post hoc analysis using the Holm-Bonferroni method showed a significant difference between the first and second sessions ($p = .015$), with all participants' responses being lowered during the second session. However, there was no significant interaction between the factors Group of game sessions and the Session's order on participants' amplitude of SCR ($df = 2$, $p = .578$, $\eta^2 = 0.001$). Regarding SCR's rise time, there was a main effect of Group of game sessions on SCR rise time ($df = 2$, $p = .005$, $\eta^2 = 0.010$). Post hoc analysis using the Holm-Bonferroni method revealed a significant difference between Groups 1 and 2 ($p = .004$) and Groups 2 and 3 ($p = .044$). However, there was no significant difference between Groups 1 and

3 ($p = .620$). Participants in the second group had significantly longer rise times compared to the other two groups. There was no significant effect of Session's order ($p = .799$) nor any interaction between the two factors ($p = .095$) on participants' rise time of their SCR.

Videos' evaluation

After the game sessions, participants watched four videos and rated NPCs using the items of the perceived intelligence and believability scales. A three-way ANOVA was conducted, considering the factors Group of game sessions, Appearance, and Coherence to analyze participants' ratings on the items. There was no main effect of the factor Group of game sessions on participants' ratings. There was a main effect of the factor Appearance on participants' rating. Soldiers were rated as more knowledgeable ($df = 1, p = .001, \eta^2 = 0.024$) and their behaviors were perceived as more attention-grabbing ($df = 1, p = .002, \eta^2 = 0.023$) than civilians, although civilians' behaviors were evaluated as more coherent ($df = 1, p = 0.04, \eta^2 = 0.009$). The factor Coherence had a main effect on several items: Coherent NPCs were perceived as more sensible ($df = 1, p < .001, \eta^2 = 0.112$), responsible ($df = 1, p = .003, \eta^2 = 0.02$), predictable ($df = 1, p < .001, \eta^2 = 0.341$), and understandable ($df = 1, p < .001, \eta^2 = 0.114$), while incoherent NPCs were perceived as having stronger personalities ($df = 1, p = 0.034, \eta^2 = 0.012$), more reactivity ($df = 1, p = .038, \eta^2 = 0.011$), and higher visual impact ($df = 1, p < .001, \eta^2 = 0.082$). There was no significant effect of the interaction between the factors Group of game sessions and Appearance. However, the interaction between Group of game sessions and Coherence had a significant effect on the items «Personality» ($df = 2, p = .002, \eta^2 = 0.035$), «Predictability» ($df = 2, p < .001, \eta^2 = 0.060$), «Behavior coherence» ($df = 2, p = .002, \eta^2 = 0.026$), and «Change with experience» ($df = 2, p = .047, \eta^2 = 0.016$) (see Fig. 8). Furthermore, the interaction between Appearance and Coherence had a significant effect on the items «Competence» ($df = 1, p < .001, \eta^2 = 0.188$), «Knowledgeable» ($df = 1, p < .001, \eta^2 = 0.158$), «Sensible» ($df = 1, p < .001, \eta^2 = 0.114$), «Responsible» ($df = 1, p < .001, \eta^2 = 0.192$), «Behavior understandability» ($df = 1, p < .001, \eta^2 = 0.038$), «Visual impact» ($df = 1, p < .001, \eta^2 = 0.052$), «Change with experience» ($df = 1, p < .001, \eta^2 = 0.03$), and «Social» ($df = 1, p = .031, \eta^2 = 0.013$). Lastly, there was no significant effect of the interaction between the three factors on participants' ratings.

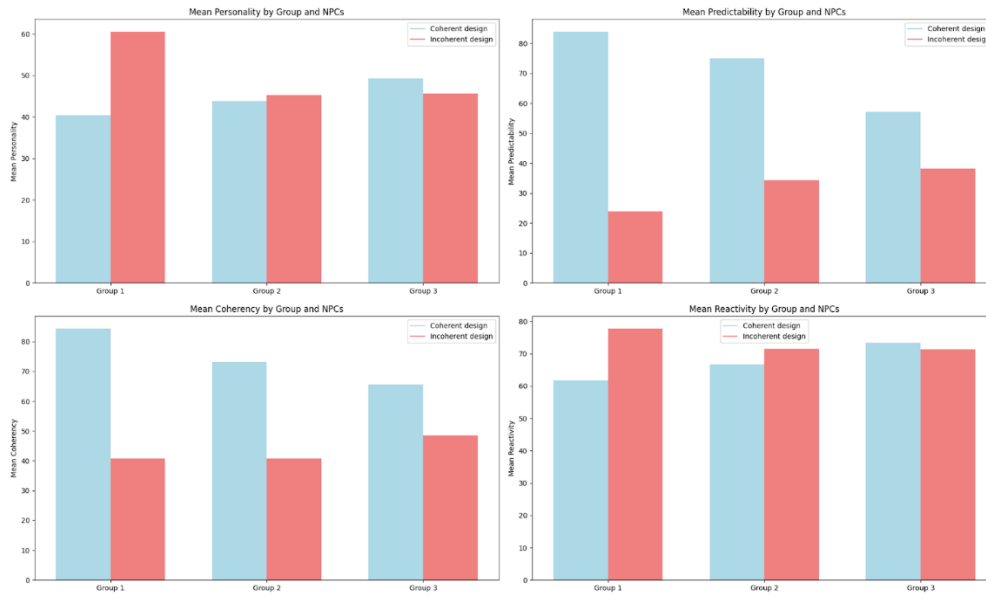


Fig. 8. Bar charts displaying the four items with significant differences influenced by the Group of game sessions and Coherence. The items ‘Behavior coherence’ and ‘Change with experience’ from the believability scale have been renamed as ‘Coherency’ and ‘Reactivity,’ respectively, for readability.

The results of Experiment 2 revealed that NPCs’ design during the game sessions significantly influenced participants’ expectations. Prior to the game sessions, there were no significant differences between groups in participants’ decision to shoot at NPCs based on their appearances. However, after the game sessions, participants in Group 1 were reinforced in their initial decision to shoot at soldiers, while participants in Group 2 had to change their decision and shoot at NPCs with a civilian appearance. Group 3, characterized by inconsistent NPCs’ design, exhibited lower rating of certainty in shooting at both appearances in the questionnaires.

Behavioral and physiological analyses during participants’ game sessions showed significant differences between groups. Participants in Group 1 maintained a larger distance to defeat enemies, likely due to the coherence and consistency of NPCs’ design, as their behaviors confirmed participants’ expectations. The other two groups exhibited more uncertainty during their decision to shoot as revealed by the questionnaires and their behavioral analysis, with Group 2 accommodating their expectations, and Group 3 struggling to form accurate expectations. The

significant change in all groups between the two sessions suggests that participants formed distinct expectations about NPCs following their interactions from the first session. Specifically, Group 2 associated aggressiveness with civilian appearances, while Group 3 learned about the unpredictability of NPCs. Regarding physiological measures, the absence of a significant difference between groups in heart rate can be attributed to participants' limited engagement in the game experience, which might not have been strong enough to elicit changes in their heart rates. For example, the absence of a compelling scenario or the simplicity of the virtual environment could have restrained participants' willingness to emotionally invest during their interactions. However, regarding SCRs, the significant difference in amplitude for the Group 3 suggests a higher sensitivity toward NPCs during their game sessions. Participants in this condition were unable to rely on accurate expectations, making them more prone to stress caused by uncertainty. In summary, the result concerning players' experience of their game sessions suggests that consistency during interactions in a military shooter game is an important parameter as it enables players to form accurate and strong expectations about NPCs' roles. Hence, the manipulation of consistency confirms the ViolEx Model [20]: participants in Group 2 accommodated their expectations to the outcome of their interactions. Therefore, participants in Groups 1 and 2 were able to make accurate decisions to shoot at NPCs based on their identification of NPCs' appearance, while participants in Group 3 had to enhance their reactivity toward aggressiveness, resulting in more stress during their game sessions.

After the game sessions, participants who encountered only coherent NPCs were the most affected by the unexpected behaviors of incoherent NPCs in their ratings of perceived intelligence and believability. Conversely, participants who encountered only incoherent NPCs rated their understanding of incoherent NPCs higher compared to the other groups, as they were expecting aggressive civilians and submissive soldiers. Their ratings for coherent NPCs were similar to participants from Group 1. Furthermore, participants who encountered NPCs with inconsistent design had ambiguous ratings for coherent and incoherent design, reflecting flexibility in their evaluation of incoherent NPCs, but a decrease in their evaluation for coherent NPCs compared to the other groups. Overall, Experiment 2 reveals that interacting with different NPCs' designs significantly influence participants' expectations, resulting in different expectations about the roles of NPCs associated with their appearance. Additionally, participants who encountered only

coherent NPCs were more affected by unexpected behaviors, while those who encountered only incoherent NPCs had higher understanding of incoherent NPCs. Participants interacting with inconsistent NPCs' design exhibited flexibility in their evaluation, but a decrease in their evaluation for coherent NPCs compared to the other groups.

8. General Discussion

8.1 Summary

The results of the two experiments provided information about the impact of breaking players' expectations on their evaluation of NPCs' perceived intelligence and believability. To break players' expectation, the coherence and consistency of NPCs' design (i.e., their appearance and behaviors) in a military shooter game were manipulated. In the first experiment, the focus was on breaking players' expectations by disrupting the coherence between the two parameters, leading to inaccurate expectations in players' minds. As a consequence, unexpected NPCs' behaviors resulted in misunderstanding, leading to a decrease in their perceived intelligence and believability. The second experiment explored the effect of consistency manipulation between the two parameters on the game experience and players' expectations. Three designs were created to depict either coherent and consistent NPCs, incoherent and consistent NPCs or inconsistent NPCs in the game sessions. Facing consistent NPCs, players were able to form new expectations with incoherent NPCs following their interactions with them. This highlights the importance of consistency in enabling players to form accurate expectations when incoherent NPCs' design has initially broken their expectations. Hence, players' evaluation of NPCs' perceived intelligence and believability was significantly influenced by their expectations. Understanding how coherence and consistency impact players' expectations is crucial in game design, as it can shape players' overall experience and immersion in the narrative. In the following, the results of the two experiments are discussed separately before addressing the limits of the study.

8.2 Coherence manipulation

The first experiment aimed at investigating the impact of breaking players' expectations on their evaluation of NPCs' perceived intelligence and believability. To do so, coherent and incoherent NPCs were created by manipulating their appearance and behaviors to convey inaccurate

expectations in a military setting. During the experiment, participants rated aggressive soldiers and submissive civilians higher in perceived intelligence and believability compared to submissive soldiers and aggressive civilians. This significant difference can be explained by previous research on believability and perceived intelligence [5, 6], as players heavily rely on their expectations to evaluate NPCs. Players rated unexpected behaviors based on their knowledge of video games, the specific type of game (i.e., a military shooter), and the stereotypes triggered by NPCs' appearance [10]. On the one hand, the higher evaluation of aggressive soldiers as more intelligent and believable may be attributed to the confirmation of players' expectations of soldiers' purpose in the game. In military shooter games, the core experience and challenge involve interacting with aggressive soldiers. Encountering submissive soldiers, which deviate from this norm, might lead to misunderstanding and disappointment in such a game, and thus decrease players' evaluation of their perceived intelligence and believability. On the other hand, aggressive civilians are not a common occurrence in military shooter games. Players usually associate them with non-threatening roles, making it counterintuitive to engage with them as enemies. The decision to shoot at a civilian might be distasteful for players as it goes against the moral compass they adopt while playing a military game [31]. Therefore, submissive civilians are more positively evaluated as they confirm players' expectations and help maintain their immersion in the game world. In summary, players tend to evaluate NPCs more positively when their behaviors confirm their expectations, as it makes them easier to understand and aligns with their overall expectation of the in-game setting.

8.3 Consistency manipulation

Game experience when manipulating coherence and consistency

In Experiment 2, participants engaged in a fictional military shooter game featuring manipulated NPCs' designs. Participants were divided into three groups based on NPCs' design. Group 1 encountered coherent and consistent NPCs, Group 2 faced incoherent but consistent ones, and Group 3 confronted an inconsistent design with random possibilities, encompassing both coherent and incoherent NPCs. Our hypothesis revolved around NPCs' design and its impact on players' post-game session expectations. The results indicate that participants adjusted their expectations of NPCs based on their interactions.

In Group 1, participants' expectations were reinforced, aligning with typical beliefs about enemies in military shooter games. Their high certainty ratings for shooting NPCs based on appearances were supported by the greater distance required to defeat enemies during their game sessions.

In Group 2, NPCs' design broke participants' initial expectations, akin to the findings in Experiment 1. Participants anticipated aggressive soldiers but encountered the opposite, leading them to make decisive choices to shoot at civilians to avoid being defeated. As mentioned in the ViolEx Model [20], repeated interactions with incoherent NPCs prompted participants to adapt their expectations to maintain their understanding of the fictional game. This adaptation process was reflected in their SCR's rise time, which notably exceeded that of the other two groups, suggesting increased participants' attention [32, 33]. After two sessions, participants adjusted their expectations, evident in their certainty scores when shooting NPCs based on appearance and the increased distance required to defeat enemies. Group 3 participants encountered NPCs with an inconsistent design, hindering their ability to adjust expectations from the absence of consistency across interactions. This inconsistency entirely disrupted their expectations, illustrated by their decreased certainty to shoot NPCs and the lack of evolution of their distance to defeat enemies across sessions. Participants in this group had to react to aggressiveness instead of anticipating hostility based on appearances, resulting in uncertainty, as reflected in their high SCR's amplitude. Precisely, significant differences emerged between groups and sessions. Participants in Groups 1 and 2 demonstrated a marked decrease in their SCR's amplitude between sessions, indicating their capability to accurately anticipate NPC interactions, leading to reduced physiological responses. However, participants in Group 3 did not exhibit the same learning pattern, remaining uncertain during and between sessions. Following [34] description of high amplitude in SCR, newly encountered NPCs in this group conveyed higher stress to players' minds, as they held the potential for fatal consequences regardless of their appearance.

Overall, manipulating NPCs' design in terms of coherence and consistency of their parameters significantly impacts the game experience. Incoherent NPCs provoke unexpected events in players' minds; nevertheless, consistency emerges as a pivotal factor for managing player expectations. Players demonstrate their ability to adjust their expectations when faced with consistent yet incoherent designs, which aligns with our initial hypotheses. Inconsistency, however, compels players to compensate for the decreased accuracy of their expectations by heightening their

reactivity toward enemies, resulting in elevated stress levels during game sessions. Consequently, game designers can leverage this knowledge to manipulate coherence and consistency to achieve their desired game experience.

Players' expectation of NPCs' intelligence and believability

Participants formed different expectations about the appearances of civilians and soldiers during their game sessions. Their evaluation of intelligence and believability of NPCs in the videos was influenced both by appearance and behaviors. Participants rated soldiers as more attention-grabbing and competent based on the context of a military shooter game. However, their specific expectations from the game sessions influenced some of their ratings.

Group 1, who encountered only coherent NPCs in their game sessions, were more affected by unexpected behaviors of incoherent NPCs in the videos. They evaluated incoherent NPCs as more attention-grabbing and ascribed them with more personality (positive violations of expectations), but less intelligence and believability than their coherent counterparts (negative violations of expectations). These results suggest that unexpected behaviors lead to both positive and negative violations of expectations, as NPCs challenge players' understanding while simultaneously piquing their interest in them. Following [5] definition of believability, players' attribution of personality to NPCs can be interesting for game designers to create immersive narrative. The attribution of personality reflects the player's acknowledgment of the NPCs' uniqueness, serving as a powerful motivator for player engagement. Those who encountered only incoherent NPCs in their game sessions rated coherent NPCs in the videos as more believable and intelligent than incoherent ones, but still rated believability of incoherent NPCs higher compared to the other groups. This difference can be explained by [35] perspective, defining believability as an evaluation based on players' expectations. Furthermore, players evaluate perceived intelligence of NPCs in the videos from their understanding based on their initial expectations which led to higher ratings for coherent ones in all groups. However, participants who encountered only incoherent NPCs had to accommodate their expectations to maintain their understanding during their game experience, leading to a higher evaluation of incoherent NPCs in comparison to the other groups. The gain in the evaluation of incoherent NPCs was not observed for participants who encountered inconsistent NPCs' design in their game sessions. The uncertainty during the game experience induced flexibility, leading them to lower their expectations and not particularly seek to explain and understand incoherent

behaviors. In summary, participants who encountered only coherent NPCs were more sensitive to unexpected behaviors due to strong expectations of the game. It decreased their evaluation of perceived intelligence and believability but raised their attention and evaluation of NPCs' personality compared to the other groups, which can be valuable for narrative experiences. Overall, the results suggest that players have the ability to form new expectations in the game when they encounter consistent NPCs' design, even if it goes against their initial expectations. Therefore, consistency of the association between appearance and behaviors can be used as a modulator to enhance players' investment of belief in NPCs. It is crucial for designers to be mindful of the potential conflicts inconsistency can have on players' immersion in the game. However, they can still leverage this information to manipulate NPCs' design and create alternative game experiences. For instance, they can strategically place unexpected behaviors after multiple expected interactions to surprise players' minds and enhance gameplay.

9. Limits and future research

The main limitation of this research pertains to two aspects: the selection of participants and the type of game used in the experiments. In Experiment 1, the participants were volunteer players registered in Ubisoft's user research database, known to be highly engaged players. Their high engagement may not represent all players' attitudes toward unexpected events in a game experience. In addition, there was a significant gender bias, with a predominance of male participants. This gender bias may have influenced the effect of breaking players' expectations on their evaluation of intelligence and believability. In Experiment 2, the participants were volunteer players who were Ubisoft employees. While the game material used in the experiment did not reflect Ubisoft games, we cannot rule out the possibility that participants' affiliation with the company may have influenced their evaluation of NPCs. Additionally, the game context used in the experiments was a military shooter game, which may not accurately represent other types of games. For example, inconsistent or incoherent NPCs' design in a science fiction game could have a different impact on players' experience. Therefore, the results obtained in this study are specific to military shooter games and may not be applicable to other genres. Future research should explore other types of games and consider individual factors, such as players' gender, game preferences, or cultural affiliation, when examining the impact of breaking expectations on the evaluation of NPCs.

This would provide a more comprehensive understanding of the factors influencing players' experiences and their expectations in different gaming contexts.

10. Conclusion

This study delved into the crucial aspects of players' interactions with NPCs in video games, specifically focusing on the impact of NPCs' design on players' expectations and their evaluation of perceived intelligence and believability. Through two experiments conducted in a military shooter game setting, the impact of breaking players' expectation on player experience was explored by manipulating the coherence and consistency of NPCs' design. Experiment 1 revealed that violating players' expectations regarding NPCs' appearance and behaviors led to a decrease in their evaluation of NPCs' perceived intelligence and believability. Players relied on their knowledge and prior experiences to form expectations, and when these expectations were disrupted, it negatively affected their assessment of NPCs' abilities to achieve their purpose in the narrative. Experiment 2 focused on the impact of coherence and consistency manipulation of NPCs' design on game experience and players' expectations. Game experience was analyzed using a combination of perceptual, behavioral, and physiological data, providing a deeper understanding of the impact of NPCs' design on players' experience. Therefore, the coherent and consistent design of NPCs reinforced players' expectations, leading to enhanced decision-making and evaluation of NPCs' perceived intelligence and believability. On the other hand, incoherent and inconsistent NPCs challenged players' expectations, but participants accommodated and adapted their expectations over time, resulting in more accurate evaluations. However, inconsistent NPCs' design created uncertainty and hindered participants from forming accurate expectations, leading to increased stress measured by the amplitude of their skin conductance responses during their interactions. Overall, the study shed light on the intricate relationship between NPCs' design, players' expectations, and their evaluation of intelligence and believability. Game designers can leverage this knowledge to create alternative game experiences by strategically breaking players' expectations or maintaining coherence and consistency in NPCs' design. Hence, understanding and considering the impact of these manipulations on players' expectations can lead to more engaging and immersive players' experiences.

ACKNOWLEDGEMENTS

Remi Poivet is funded by the French ANRT and the Company Ubisoft.

References

- [1] R. Giusti, K. Hullett, and J. Whitehead, “Weapon design patterns in shooter games,” in Proc. of the 1st Workshop on Des. Patterns in Games, 2012, pp. 1-7.
- [2] S. Ohrberg, «Recreating Believability In NPCs: The Effects Of Visual And Logical Behaviour,» 2019.
- [3] D. Besnard, D. Greathead, and G. Baxter, «When mental models go wrong: Co-occurrences in dynamic, critical systems,» in Int. J. of Human-Computer Studies, vol. 60, no. 1, pp. 117-128, 2004.
- [4] C. Butcher and J. Griesemer, «The illusion of intelligence: The integration of AI and level design in Halo,» presented at the Game Developers Conference, San Jose, CA, March 21-23, 2002.
- [5] A. B. Loyall, “Believable Agents: Building Interactive Personalities,” Ph.D. dissertation, Carnegie Mellon University, 1997.
- [6] C. Bartneck, D. Kulić, E. Croft, and S. Zoghbi, “Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots,” Int. J. of Social Robot., vol. 1, no. 1, pp. 71–81, 2009, doi: 10.1007/s12369-008-0001-3.
- [7] E. Lavik, «Narrative structure in The Sixth Sense: a new twist in ‘twist movies’?» The Velvet Light Trap, vol. 58, no. 1, pp. 55-64, 2006.
- [8] Rockstar, «Red Dead Redemption 2,» 2018..
- [9] C. Panitz, D. Endres, M. Buchholz, Z. Khosrowtaj, M. F. Sperl, E. M. Mueller, ..., M. Piquart, «A revised framework for the investigation of expectation update versus maintenance in the context of expectation violations: The ViolEx 2.0 model,» Frontiers in psychol., vol. 12, p. 726432, 2021
- [10] H. Warpefelt. “Cues and insinuations: Indicating affordances of non-player character using visual indicators,” in Proc. DiGRA 2015: Diversity of play: Games - Cultures – Identities, May.

2015. [Online]. Available: <http://www.digra.org/digital-library/publications/cues-and-insinuations-indicating-affordances-of-non-player-character-using-visual-indicators/>
- [11] K. Rogers, M. Aufheimer, M. Weber, and L. Nacke, “Towards the Visual Design of Non-Player Characters for Narrative Roles,” *Graphics Interface 2018*, 2018 pp. 154–161, doi: 10.20380/GI2018.21.
- [12] R. Pradantyo, M. V. Birk, and S. Bateman, “How the Visual Design of Video Game Antagonists Affects Perception of Morality,” *Frontiers in Computer Science*, vol. 3, 2021.
- [13] A. Ortony, “On making believable emotional agents believable. Emotions in humans and artifacts,” 2003.
- [14] K. Cheng and P. A. Cairns, «Behaviour, realism, and immersion in games,» in *CHI’05 ext. abstr. on Human factors in comput. syst.*, April 2005, pp. 1272-1275.
- [15] J. K. Burgoon and J. B. Walther, «Nonverbal expectancies and the evaluative consequences of violations,» *Human Commun. Res.*, vol. 17, no. 2, pp. 232-265, 1990.
- [16] J. K. Burgoon, «Interpersonal expectations, expectancy violations, and emotional communication,» *J. of language and soc. psychol.*, vol. 12, no. 1-2, pp. 30-48, 1993.
- [17] J. Perez and L. Feigenson, «Violations of expectation trigger infants to search for explanations,» *Cognition*, vol. 218, p. 104942, 2022.
- [18] Void Interactive, “Ready or not”, 2021.
- [19] M. Niedeggen, R. Kerschreiter, and K. Schuck, «Loss of control as a violation of expectations: Testing the predictions of a common inconsistency compensation approach in an inclusionary cyberball game,» *PLoS One*, vol. 14, no. 9, 2019.
- [20] M. Gollwitzer, A. Thorwart, and K. Meissner, «Psychological responses to violations of expectations,» *Frontiers in psychol.*, vol. 8, p. 2357, 2018.
- [21] M. Gollwitzer, T. Rothmund, and P. Süssenhach, «The Sensitivity to Mean Intentions (SeMI) model: Basic assumptions, recent findings, and potential avenues for future research,» *Soc. Personal. Psychol. Compass*, vol. 7, pp. 415–426, 2013.
- [22] P. Süssenhach, M. Gollwitzer, L. Mieth, A. Buchner, and R. Bell, «Trustworthy tricksters: Violating a negative social expectation affects source memory and person perception when fear of exploitation is high,» *Frontiers in psychol.*, vol. 7, p. 2037, 2016.

- [23] J. W. Sherman, A. Y. Lee, G. R. Bessenoff, and L. A. Frost, «Stereotype efficiency reconsidered: Encoding flexibility under cognitive load,» *J. Person. Soc. Psychol.*, vol. 75, pp. 589–606, 1998. doi: 10.1037/0022-3514.75.3.589.
- [24] R. Casati and E. Pasquinelli, «How can you be surprised? The case for volatile expectations,» *Phenomenology and the Cog. Sci.*, vol. 6, no. 1-2, pp. 171-183, 2007.
- [25] J. K. Burgoon, «Expectancy violations theory,» *The Int. Encyclopedia of Interpersonal Commun.*, pp. 1-9, 2015.
- [26] M. Beigi, J. Callahan, and C. Michaelson, «A critical plot twist: Changing characters and foreshadowing the future of organizational storytelling,» *Int. J. Manag. Rev.*, vol. 21, no. 4, pp. 447-465, 2019.
- [27] S. Moussawi, M. Koufaris, and R. Benbunan-Fich, «How perceptions of intelligence and anthropomorphism affect adoption of personal intelligent agents,» *Electron. Markets*, vol. 31, pp. 343-364, 2021.
- [28] T. Koda and P. Maes, «Agents with faces: the effect of personification,» in *RO-MAN'96 – The 5th IEEE Int. Workshop Robot and Human Commun.*, 1996, pp. 189–194, doi: 10.1109/ROMAN.1996.568812.
- [29] P. Gomes, A. Paiva, C. Martinho, and A. Jhala, «Metrics for character believability in interactive narrative,» in *Interactive Storytelling: 6th Int. Conf., ICIDS 2013, Istanbul, Turkey, November 6-9, 2013, Proceedings 6*, pp. 223-228, Springer International Publishing.
- [30] Ubisoft, «Ghost Recon Breakpoint,» 2019.
- [31] B. DeVane and K. D. Squire, «The meaning of race and violence in Grand Theft Auto: San Andreas,» *Games and Culture*, vol. 3, no. 3-4, pp. 264-285, 2008.
- [32] P. H. Venables, S. A. Gartshore, and P. W. O'riordan, «The function of skin conductance response recovery and rise time,» *Biol. psychol.*, vol. 10, no. 1, pp. 1-6, 1980.
- [33] M. Shibagaki and T. Yamanaka, «Attention of preschool children: temporal variation of electrodermal activity during auditory stimulation,» *Perceptual and Motor Skills*, vol. 72, no. 3_suppl, pp. 1115-1124, 1991.
- [34] M. E. Dawson, A. M. Schell, and D. L. Filion, «The electrodermal system,» *Handbook of psychophysiology*, vol. 2, pp. 200-223, 2007.
- [35] Magerko, B. (2007, November). Measuring Dramatic Believability. In *AAAI Fall Symposium: Intell. Narrative Technol.* (pp. 79-82).

CHAPTER 4: The influence of conversational agents' role and communication style on user experience

Poivet, R.^{1,2}, Lopez Malet, M.², Pelachaud, C.¹, & Auvray, M.¹ (under review). The influence of conversational agents' role and communication style on user experience. *Frontiers in Psychology Cognitive Science*.

¹ Sorbonne Université, CNRS, Institut des Systèmes Intelligents et de Robotique, ISIR, Paris, France

² Ubisoft Paris studio, Montreuil, France

Keywords: Conversational Agents, verbal interactions, human behaviors, human-machine interaction, believability

Abstract

In narrative experience, Conversational Agents (CAs) are characterized by their roles in the narrative and the communication style they adopt during conversations. Within computer games, users' evaluation of the narrative is influenced by their evaluation of CAs' intelligence and believability. However, the impact of CAs' roles and communication styles on users' experience remains unclear. This research investigates the influence of CAs' roles and communication styles on users' experience through a crime-solving textual game. Distinct CAs were developed, each were assigned either with the role of a witness or a suspect and would exhibit either an aggressive or cooperative communication style. A Wizard of Oz method was employed to simulate the communication styles. Users' experience was evaluated using perceptual measures (perceived intelligence and believability scales) and behavioral measures (including analysis of users' input length, input delay, and conversation length). Additionally, users were required to identify the culprit, assess the certainty of their judgments, and rank the CAs based on their conversational preferences. The findings reveal that both the role and the communication style of CAs have an influence on users' perception of CAs' intelligence and believability, as well as on users' behaviors. Users' evaluation was primarily influenced by the CAs' roles, while their conversational behaviors were influenced by the CAs' communication styles. The communication style significantly impacted the choice of the culprit and users' preferred conversations.

1. Introduction

During interactive experiences, conversational agents (CAs) convey the narrative through their interactions with users. To develop these agents, designers manipulate parameters that aim to influence users' experience of the narrative. For instance, users can be informed about the explicit roles of CAs in the narrative, allowing them to adjust their conversational strategy to better interact with the agent. Moreover, CAs' communication style is crucial in the interaction as it effectively communicates the narrative by generating content for users. Additionally, research on human-agent interactions demonstrated the importance of taking into account users' perception of intelligence and believability attribution to gain insights and create more engaging agents (Bartneck et al., 2009; Loyall, 1997). This study aims to provide insights about the influence of CAs' role and communication style on users' experience. A textual computer game was created to investigate their effect. In the game's narrative, users endorse the role of a detective who investigates a case by engaging in conversations with various CAs. Precisely, users were instructed to name a culprit between four distinct CAs whose role and communication style were manipulated. Users' evaluation of CAs' was measured through items of perceived intelligence and believability scales and users' behaviors were analyzed through the measure of their input length (number of characters in their message), delay (number of seconds to send their message), and conversations length (number of turns in their conversations).

In the remainder of the manuscript, we first highlight the importance and impacts of role and communication style factors in the creation of CAs within narrative environments. Then, we detail our experimental approach, outlining the narrative setting scenario and the manipulation of CAs' parameters. Finally, we present and analyze the findings, concluding the paper with our final remarks.

2. Background

CAs, also known as chatbots, are a type of artificial agents that aims to simulate human conversation through natural language processing and generation (Klüwer et al., 2011). In computer games, the number of textual narrative games is growing and so is the importance of CAs in their design. These games offer an interactive experience that relies on textual communication

between the user and CAs who take on different roles in the scenario. Effective communication is crucial for users' immersion and engagement in the narrative (Isbister and Nass, 2000), as it directly impacts their enjoyment of the game (Schaffer and Fang, 2019). Therefore, designing CAs requires careful consideration of users' perception and expectations.

CAs involved in narrative experiences are characterized by their role and their communication style toward the user. These factors aim to influence users' expectations, their evaluation of the agent, and their conversations (Mou and Peng, 2009; Nag and Yalçın, 2020). Users' expectations are influenced by their knowledge of CAs (Komatsu and Yamada, 2011) and by implicit stereotypes, which are associated with positive evaluations or negative ones (Brahnam and De Angeli, 2012).

Regarding roles, in narrative experiences, CAs can assume, for example, either friendly roles or opponent ones toward the user, which can be described in terms of their moral compass (Gomes et al., 2013). This role refers to the agents' function in the narrative. Information about the role can be explicitly communicated (e.g., the agent is introduced by the narrator as an opponent or as an ally) or inferred by users during their interaction (e.g., through the agent's communication style). Explicit roles aim to shape users' interactions by triggering pre-existing positive or negative stereotypes before the actual conversation occurs. For instance, the explicit role of an opponent affects users' expectations, which involves an anticipation of their interaction influenced by the stereotypes associated with hostility. This expectation would shape their decision to interact with the CAs (e.g., adopting an appropriate communication strategy or even avoiding the interaction).

Regarding communication style, CAs are characterized by the delivery form of verbal output during interactions with users. Communication style includes content generation, conversational strategy, and linguistic cues to convey CAs' intentions and personality traits (van Pinxteren et al., 2023). In narrative experiences, before an interaction with CAs, users make expectations of their communication style based on the stereotypes associated with CAs' explicit roles. For instance, users could expect an opponent to adopt an aggressive communication style, since this trait is associated with hostility (Infante, 1995). Therefore, designing communication style is crucial for users' interactions, which involves the choice of CAs' conversational strategy and linguistic cues to affect users' perception (Mairesse and Walker, 2009; Resendez, 2020). For instance, designers of CAs can use linguistic cues associated with the personality traits of the Big Five model to convey distinct personalities, such as extraversion, conscientiousness, agreeableness, openness to

experience, and neuroticism (Mairesse and Walker, 2007). For example, a CA using a formal lexicon conveys more conscientiousness to users but less extraversion than a CA with an informal one (Heylighen and Dewaele, 2002). Personality traits can also affect users' behaviors and engagement during their conversations. Ruane et al.'s (2021) study examined the impact of chatbots' perceived personality on users' engagement and preference. The results showed that users tend to adopt the communication style of chatbots. The authors hypothesized that users would engage in longer conversations with their preferred CA. They created two versions of chatbots using linguistic cues associated with the Big Five personality model: one with high extraversion and agreeableness, and the other with the opposite traits. The study measured participants' input length (number of words in their messages) and conversation length (number of minutes and turns), and found that participants tended to mimic the linguistic cues of the chatbot they were interacting with. In this context, the formal lexicon associated with the low extraversion of the second chatbot led to longer conversations. However, participants preferred their conversations with the first chatbot as it was perceived as more agreeable. In conclusion, the selection of appropriate linguistic cues and adjustment of the content's communication is crucial to create CAs with different personalities and engaging qualities (Følstad and Skjuve, 2019).

Creating opponent and friendly CAs involves different requirements. The communication style of opponent CAs has to be perceived as aggressive to accurately convey their intention. Their conversational strategy relies on verbal aggressiveness, which reflects an intention to attack the interlocutor. The desired outcome involves emotionally affecting the interlocutor, for example by inducing humiliation and negative feelings (Infante and Wigley, 1986). For instance, to operationalize the strategy, the negative content polarization, the lexicon formality, and the use of swear words are associated with aggressive ascriptions (Mehl et al., 2006; Pennebaker and King, 1999). Overall, an aggressive communication style conveys a disagreeable personality trait (Mairesse and Walker, 2009) which reflects the distinction between expected personalities and outcomes with opponent or with friendly roles in a narrative experience. Users expect opponents in the narrative to induce negative consequences based on their stereotypes, and thus agents are expected to be disagreeable and their communication style to be aggressive.

On the other hand, the communication style of friendly CAs involves linguistic cues associated with positive personality traits such as agreeableness, openness to experience, or extraversion

(Völkel et al., 2020) and a friendly conversational strategy (Simpson et al., 2020). Friendly CAs would be more inclined to positive content polarization and less inclined to negative topics (Mehl et al., 2006). For instance, a friendly CA in a narrative could be expected to diffuse tensions in their interaction with users using consilience markers (e.g., generate apologies during misunderstandings with users (de Sá Siqueira et al., 2023)). Moreover, friendly CAs are associated with higher extraversion and thus are more inclined to larger verbosity and informal lexicon (Mairesse and Walker, 2009). As a result, the combination of linguistic cues and conversational strategy used in a friendly communication style aims to make the friendly CAs more engaging and cooperative during their interactions with users.

3. Scales of perceived intelligence and believability

During interactive experiences, CAs' perceived intelligence and believability attributions are associated with users' enjoyment and motivation to interact (Loyall, 1997; Moussawi et al., 2021). Therefore, understanding how the role and communication style of CAs affect these attributions could greatly enhance their design, and thus users' experience. Two scales are particularly relevant to do so: Perceived intelligence and Believability.

Perceived intelligence can be used to probe users' evaluation of CAs' intelligence (Bartneck et al., 2008). Such evaluation relies on two dimensions: understandability and performance (Koda and Maes, 1996), both based on users' understanding of the agents' purposes and efficiency in reaching their goals. For instance, CAs' perceived intelligence would rely on their capacity to accurately simulate natural human communication with users. However, in narrative experiences, CAs' role can affect users' attitudes and thus, influence their expectations of the agents' communication style and purpose. In that sense, the perception of intelligence of a friendly CA or an aggressive one would be different, as their purposes in the narrative are not the same. Namely, friendly agents are conceived to help users, while aggressive ones make users increase the challenge of the game. To improve artificial agents' design by considering users' perception, Warner and Sugarman (1986) proposed an intelligence evaluation scale that relies on five semantic items: Incompetent/ Competent, Ignorant/ Knowledgeable, Irresponsible/ Responsible, Unintelligent/ Intelligent,

Foolish/ Sensible. This scale assesses users' judgment of the two dimensions of understandability and performance.

Besides intelligence perception, when users believe in their interaction with CAs their level of engagement increases (Nag and Yalçın, 2020). As these agents aim to simulate human communication, it is necessary to understand how their design conveys believability from the user's point of view. As users evaluate agents from their expectations of how they should behave, believable agents ought to have a close correspondence between users' expectations and their interaction (Loyall, 1997). Therefore, agents' roles can influence users' evaluation as they will expect these agents to have different purposes (e.g., interactions will then differ when occurring with friendly or aggressive agents). Gomes et al. (2013) developed a scale comprising multiple dimensions which play crucial roles in determining how interactive agents are perceived as believable in narrative experiences. These dimensions listed below allow one to quantify agents' believability through users' ratings (Where < X > is replaced by the evaluated agent).

- Awareness: < X > perceives the world around him/her.
- Behavior understandability: It is easy to understand what < X > is thinking about.
- Personality: < X > has a personality.
- Visual impact: < X >'s behavior draws my attention.
- Predictability: < X >'s behavior is predictable.
- Behavior coherence: < X >'s behavior is coherent.
- Change with experience: < X >'s behavior changes according to experience.
- Social: < X > interacts socially with other characters.

Considering users' ratings of these dimensions during CAs' design process could make them more believable. Thus, understanding the effects of the agents' design parameters on believability and on perceived intelligence could help designers conceive more engaging narrative agents.

To summarize, CAs' design is crucial for positive users' experiences, as these agents convey the core of the narrative. CAs' role can influence users' expectations and attitudes toward them, while

the communication style confirms or not their predictions and impacts their behaviors and their perception of the agents' personality traits. Previous research has shown that studying users' perceived intelligence and believability during their interaction with artificial agents can provide crucial information to improve designers' choices. However, it remains unclear how these design factors affect users' experience. The goal of this research is to understand how the roles and communication styles of CAs affect how users ascribe them intelligence and believability in narrative experiences. To this aim an experiment was conducted in which participants interact with different CAs set in a game scenario.

4. General methods and procedure

The experiment investigates how users' experience is influenced by the role and communication style of CAs. A French computer textual game was developed. The game's narrative was designed as a detective game where users assume the role of a detective and engage in various conversations with CAs to solve a case. The detective aims to name a culprit among four CAs. The CAs were defined by their identity, personality traits, and knowledge about the case. Before each discussion, the explicit role of the CA was given to the detective. CAs' answers were either aggressive or cooperative based on the manipulation of linguistic cues and conversational strategies. To generate the CAs' answers, a Wizard of Oz (WoZ) selected outputs from a predefined list of sentences. The participants engaged with the four CAs, with their roles and communication style pseudo-randomized. This ensured that each participant experienced all the experimental conditions. The duration of each discussion was set to ten minutes. The investigation of users' experience involved participants rating items on perceived intelligence (Bartneck et al., 2008) and believability (Gomes et al., 2013) scales next to each discussion. The lexicon of the believability scales was modified to suit the interaction context, replacing the term "behavior" in the various items with "discourse". Additionally, participants rated their perception of warmth (from "0 - very cold" to "100 - very warm") and cooperation/aggressivity (from "0 - cooperant" to "100 - aggressive"). Participants' behaviors were analyzed through the measure of their inputs' length (number of characters in their sentences) and delay (number of seconds to send their input), and the number of turns during their conversation. After the four interactions, participants were required to indicate a culprit and hierarchize the CAs based on their preferred conversation. They ranked the conversations, with the

first place indicating their most preferred conversation and the fourth place representing the least preferred one.

Participants

Thirty-two French employees of Ubisoft participated in the experiment, comprising 19 men and 13 women, with an average age of 29 years (SD = 7.3). The participants rated their frequency of playing video games and role-playing games using scales ranging from «0 - I never play» to «100 - I play every day.» There was a gender difference in role-playing game habits, with women reporting a mean of 31.4 (SD = 29.8) and men reporting a mean of 54 (SD = 32.9), while no difference was found in video game habits (mean = 73.9, SD = 25.1). All participants were contacted via email and provided with information regarding the general purpose of the research. The email emphasized the voluntary nature of participation and the option to withdraw at any time. The experiment lasted approximately 50 minutes and was conducted following the principles outlined in the Declaration of Helsinki.

5. Materials

The scenario of the game presents a police investigation where participants assume the role of a detective and must engage in conversations with the CAs to solve the crime case. Before their interaction, participants are provided with contextual information regarding the police investigation and the various CAs' identities. This information sets the narrative of the game and aims to guide participants' inquiries during their discussions with the CAs.

The four agents are defined by their identity, personality traits, and knowledge about the crime. Their identities refer to the CA's personal information (i.e., agent's name, age, and profession) and backstories that aim to add in-depth details during their conversation. For instance, one of the CA's identities was Christian, who is described as a 41-year-old man, working as a security agent in the casino. Each identity was ascribed personality traits from the Big Five model and was communicated in their content generation through linguistic markers. The purpose of describing each identity with different personality traits is to create diverse and engaging CAs by giving them distinct but consistent personalities. For instance, Christian is high in extraversion but low in conscientiousness as he will be more likely to use informal lexicon, slang, and in-group markers

such as ‘my pal’ when speaking with the detective (see Table 1 for the details of all the personality traits and the associated linguistic cues). Additionally, the CAs’ gender and age were controlled to avoid confounding variables (e.g., stereotypes on gender-aggressivity association). On the other hand, knowledge about the crime refers to the content they will communicate to participants. Moreover, the relationships between the CAs were priorly defined and controlled in their content, as social interactions between witnesses and suspects are important in a crime-solving situation (e.g., to solve a mystery, the detective might rely on the relationship between the suspects to name a culprit).

“Police inspector, you are contacted to take over the investigation of a colleague who is on leave. You are asked to find out more about a case of fraud in the casino "Le Prestige" in Le Havre. The director, Patrice Serres, has summoned the previous inspector to investigate a difference in the account for an amount of 400,000 euros. He suspects that one or more of his employees are embezzling from the casino.

You will have to question four of them:

- Mathieu, 39 years old, a seasoned croupier and expert in handling cards.*
- Enzo, 32, a tormented bartender who loads up the drinks to disinhibit the customers.*
- Anthony, 35, co-manager who replaced the old staff with his employees.*
- Christian, 41, a security guard, a disbarred ex-cop addicted to gambling.*

Your predecessor has already interviewed all the suspects and written notes on their potential involvement in this case. So, the people you are going to talk to have already been identified as witnesses or suspects.”

Figure 1: The scenario of the game communicated to participants.

Finally, CAs are defined by their own identity, which refers to their personal information and the linguistic markers associated with their personality that shape their content’s communication form (see Table 1). Each identity is described below.

- Mathieu Fournier, 39 years old, croupier. Mathieu is the oldest employee of the casino. He has been in the job for 15 years and is well-liked by the customers. He is described as very skilled by his coworkers. Mathieu is high in extraversion but low in agreeableness as he tends to use an informal lexicon and an impolite form of address. Thus, he has a quick wit and a dry sense of

humor. He sees Anthony as a very ambitious person and since his arrival as co-manager, he now has Christian at his table who participates in the games as a fake player. He doesn't like his presence as he feels watched but remains professional. He spends his free time at Enzo's counter without necessarily talking to him. He has a positive opinion of Enzo's competence in his profession.

- Enzo Lamy, 32 years old, barman. Enzo is a mixology enthusiast and is confronted with unpleasant behavior from drunken customers. In addition, he respects the alcohol dosages instructions given by the management which indicates a high conscientiousness. However, he is high in neuroticism, which involves anxious reactivity during his interactions with the detective. He has been hired by Anthony; they have a relationship of trust. He does not talk much to Mathieu although he finds him competent. He appreciates Christian's unconventional personality.

- Anthony Frey, 35 years old, co-manager. Anthony is a highly educated man; thus, he uses a formal lexicon and adopts a polite form of address toward the detective. He was the one that discovers the discrepancy in the accounts and notified the police. He is high in conscientiousness as he changed procedures and replaced staff before notifying the police. He recruited Christian, whom he finds useful to the casino despite their very different personalities. He recruited Enzo, whom he finds very competent in his work and contributes to a good customer experience. Mathieu is the only employee who has not been replaced by Anthony, so he does not know him well but has nothing against him.

- Christian Vigneron, 41 years old, security agent. Christian is a former police officer who was disbarred for alcohol and gambling problems. Christian is high in extraversion but low in conscientiousness as he tends to use familiar lexicon and in-group markers when he interacts with the detective. He has been recruited by Anthony, who he likes although he considers to be his opposite personality. He spends time at Mathieu's table as a fake gambler to keep an eye on the customers, thanks to an envelope given to him by the management. He thinks Mathieu is a good croupier and has nothing against him. He spends the rest of his time sitting at Enzo's bar. Mathieu appreciates that Enzo is a good listener.

Table 1 – Linguistic markers associated with each identity’s content generation.

Agent’s Identity	Linguistic markers
Mathieu Fournier (39 years old, Croupier)	Medium verbosity
	Contracted negation (e.g., ‘I can’t tell you...’)
	Informal lexicon (slight vulgarity, e.g., ‘it pisses me off when...’) Impolite form of address (sarcastic tendency, e.g., ‘Yeah, it’s the colonel mustard who did it.’)
Enzo Lamy (32 years old, Barman)	Strong verbosity
	Contracted negation (e.g., ‘I can’t tell you...’)
	Formal lexicon (no vulgarity unless under accusation in aggressive verbal behaviors)
	Polite form of address Emotional reaction (e.g., ‘oh no...’, ‘I’m so anxious about...’)
Anthony Frey (35 years old, Co-manager)	Strong verbosity
	Uncontracted negation (e.g., ‘I do not...’, ‘I cannot...’)
	Formal lexicon (no vulgarity and rich vocabulary, e.g., ‘This accusation is outrageous’) Polite form of address
Christian Vigneron (41 years old, Security guard)	Medium verbosity
	Contracted negation (e.g., ‘I can’t tell you...’) Informal lexicon (slang and swear words)
	Impolite form of address (in-group markers, e.g., ‘I get you, my pal.’)

CAs’ communication styles were manipulated to convey aggressive or cooperative intentions. To do so, hostility and agreeableness markers based on communication theory (Infante, 1995; Pennebaker and King, 1999; Mairesse and Walker, 2007; Mairesse and Walker, 2009) have been implemented in the communication content to affect participants’ evaluation of aggressivity and cooperation. On the one hand, CAs in their aggressive form had less verbosity, used personal attacks, and had negative content polarization. Moreover, they had aggressive sentences toward the participants rather than answering their inquiries (e.g., ‘Do you even know what you are talking about?’). On the other hand, CAs in their cooperative form had more verbosity and answered pedagogically to the detective’s questions. In addition, agents with a cooperative communication style requested confirmation for the relevance of their answers (e.g., ‘I hope my answers will help

you solve this affair.’) and used consilience markers (e.g., ‘sir’ or ‘detective’). To control the form of CAs’ communication style, a Wizard of Oz (WoZ) method was used. Precisely, the WoZ used a working sheet for each identity (see Table 2 for an example of Anthony’s identity). This working sheet was a list of the detective’s potential questions and the content’s communication declension (see Table 2). In particular, the content communication was declined in two styles (aggressive or cooperative) that the WoZ followed based on the experimental condition the discussion was set in (e.g., an aggressive or a cooperative CA). The potential questions asked by participants were listed based on the intention associated and involved specific situations such as ‘Initial contact’, ‘Backstory information’ and ‘Accusation’. If participants asked follow-up questions about a specific topic, the WoZ either rephrased their answer in the cooperative form condition or made the answer more aggressive (i.e., the Wizard of Oz answers the question and adds impatience markers such as ‘as I already said’, ‘Your questions are annoying’). The CAs’ roles in the scenario are closely tied to their context (i.e., a witness or a suspect). For instance, in a crime-solving game, witnesses can be expected to act as cooperative agents who assist participants in solving the crime through their communication, while suspects can be expected to be more hostile and convey their motivation in the game (i.e., indicating whether they are guilty or not). Before each discussion, participants were given a brief description of the CA’s identity content (i.e., the same description of the experiment’s introduction) and their role in the scenario (e.g., ‘Witness: Enzo, 32 years old, a tormented bartender who loads up the drinks to disinhibit the customers.’).

Table 2 – Example of the Wizard of Oz’s working sheet: Co-manager’s answers.

Detective’s potential question	Question’s intent	Cooperative answers	Aggressive answers
<i>‘Hello.’</i>	Initial contact	<i>‘Hello sir.’</i>	<i>‘Hello sir.’</i>
<i>‘How are you?’</i>	Initial contact	<i>‘Personally, I am doing very well. Even though the disappearance of such a large sum of money is on my mind.’</i>	<i>‘Let’s get to the point, why am I being questioned about the disappearance of 400,000 euros in my own casino?’</i>
<i>‘How long have you been working here?’</i>	Backstory information	<i>‘It has been 2 years since I was hired to modernize the casino «Le Prestige». The old casino was obsolete and this modernization with the addition of the hotel complex makes it possible to make this place a splendid jewel for tourism.’</i>	<i>‘I’ve been the co-manager of the casino for 2 years. But if the goal is to have such basic information, go to HR, you will waste less time.’</i>
<i>‘Have you stolen the money?’</i>	Suspicion	<i>‘Of course not. I am the one who warned the police about the discrepancy in the accounts, I think that clears me. But if some elements are not clear to you, I am ready to answer all of your questions.’</i>	<i>‘This accusation is outrageous! Do your job seriously before you come and waste my time.’</i>
<i>‘Are you the culprit?’</i>	Suspicion	<i>‘I am not. On the contrary, I’m here to help you find out who the real culprit is.’</i>	<i>‘Of course not! I’m looking for it, just like you. So do your job!’</i>
<i>‘I have some evidence against you!’</i>	Accusation	<i>‘What evidence are you talking about? In any case, I am ready to tell you everything I know.’</i>	<i>‘This attempt at destabilization is ridiculous. What evidence? Be specific!’</i>
<i>‘You look nervous.’</i>	Follow-up	<i>‘I’m not. As an exemplary leader, I work under constant pressure. What you call nervousness, I call responsiveness.’</i>	<i>‘I’m just annoyed by the level of your questions. I have several meetings today so make it quick and better.’</i>
<i>‘What do you think of Mathieu?’</i>	Social interaction	<i>‘He seems to be impeccable. Clients like him and he has adapted well to the change in staff.’</i>	<i>‘He seems experienced.’</i>
<i>‘Thanks for your answers.’</i>	Gratitude	<i>‘Thanks to you. I remain at your service.’</i>	<i>‘I wouldn’t say that the pleasure is shared.’</i>

6. Results

In the experiment, participants interacted with four CAs and were required to rate their experience with each one. A pseudo-randomized design was employed to manipulate the roles, communication style, and order of presentation across participants. The analysis of participants' experience encompassed both CAs' evaluation through rating scales and behavioral measures of the conversations. The CAs' evaluation included participants' ratings of each of the items on perceived intelligence and believability scales. Additionally, participants indicated their judgment of CA's warmth and aggressiveness through specific items. The behavioral measures included the input length of the participants, the delay, and the number of turns during their conversations. As a summary of participants' narrative experience, the culprit's designation and preference's ranking were analyzed too.

A three-way ANOVA was conducted to analyze participants' ratings of their conversations with CAs. Due to the multiple items involved in the scales, only the statistically significant ones are reported here. The factor 'role' had a significant impact on the item 'Visual impact' of the believability scale (i.e., '< X >'s discourse draws my attention'). When CAs were introduced as witnesses, participants were significantly more attentive during the conversation compared to suspects ($F(1, 124) = 5.147, p = 0.025$). Furthermore, when examining the different levels of the 'communication style' factor, a simple effect analysis of the 'role' factor revealed that participants rated their attention significantly lower when suspects exhibited a cooperative communication style compared to an aggressive communication style. ($p = 0.016$). The 'communication style' factor had a significant impact on participants' rating of warmth ($F(1, 124) = 34.086, p < 0.001$) and aggressivity ($F(1, 124) = 258.903, p < 0.001$). The interaction between the 'role' and 'order' factors had a significant effect on participants' ratings. Precisely, there were significant differences in the evaluations of the first and last encountered CAs. The analysis indicated that participants rated their attention higher when CAs were introduced as witnesses ($F(1, 28) = 4.773, p = 0.037$) and attributed them with more personality than suspects ($F(1, 28) = 10.817, p = 0.003$). Suspects were evaluated as more competent ($F(1, 28) = 9.789, p = 0.004$), knowledgeable ($F(1, 28) = 18.640, p < 0.001$), intelligent ($F(1, 28) = 14.497, p < 0.001$), sensible ($F(1, 28) = 7.846, p = 0.009$) and responsible ($F(1, 28) = 4.443, p = 0.045$). The interaction between order and communication style had no significant impact on participants' ratings.

To analyze participants' behaviors, each of the three measures described above was analyzed separately using three-way ANOVAs following the same approach of the participants' ratings of the items' scales. Regarding the input length, the results showed a significant effect of the factor 'communication style' ($F(1, 1067) = 5.017, p = 0.025$). Participants made longer inputs (in terms of sentence length) when they were interacting with aggressive CAs. There was a significant interaction between the factors 'role' and 'order' ($F(3, 1067) = 3.829, p = 0.010$). Participants wrote sentences with more characters when interacting with suspects during the first and second conversations, while it is the opposite for the last conversations. There was also a significant interaction between the factors 'role', 'communication style', and 'order' ($F(3, 1067) = 3.287, p = 0.020$). Precisely, the order of the conversation had a significant impact on participants' input length for aggressive suspects ($p = 0.029$). Regarding the delay of the inputs, the analysis excluded participants' first message, as it initiated their conversation. The results of the three-factor ANOVA indicated a significant interaction between the factors 'role' and 'order' ($F(3, 1067) = 8.071, p < .001$). Participants took longer to write their inputs when they faced suspects during the first and second conversations, while it is the opposite during the last conversations. There was a significant effect between the factors 'role', 'communication style', and 'order' ($F(3, 1067) = 6.216, p < .001$). Simple main effects analyses showed that the order of the conversation had a significant impact on participants' delay for aggressive suspects ($p = 0.002$), cooperative suspects ($p < .001$), and aggressive witnesses ($p = 0.009$), but no significant effect was observed for cooperative witnesses ($p = 0.370$). Finally, the analysis of the number of turns during the conversations revealed a significant impact of communication style ($F(1, 1064) = 64.494, p < .001$) on conversations. Conversations with aggressive CAs were significantly longer compared to cooperative ones. There was an effect of order ($F(3, 1064) = 6.574, p < .001$), with participants having longer conversations by the last encountered CAs. There was a significant interaction between order and role ($F(3, 1064) = 10.546, p < .001$) as conversations with suspects were longer when they were encountered last. There was a significant interaction between order and communication style ($F(3, 1064) = 4.762, p = 0.003$) highlighting a significant difference between aggressive and cooperative CAs through the conversations. Lastly, there was a significant interaction between order, role, and style ($F(3, 1064) = 10.598, p < .001$). Simple main effects revealed that the order of interaction influenced the conversation length for aggressive suspects ($p < .001$), cooperative suspects ($p < .001$), and

aggressive witnesses ($p = 0.011$). Notably, conversations with cooperative witnesses did not appear to be affected by the order of the conversation.

Regarding participants' indication of the culprit and their ranking of the conversations. A contingency table was used to analyze the distribution of participants' indications of the culprit across the four conditions. A chi-squared test was conducted, and the results indicated that there was no significant difference between witnesses and suspects ($\chi^2 = 0.439$, $p = 0.508$) but a significant effect of the 'communication style' ($\chi^2 = 4.176$, $p = 0.029$). In addition, a linear regression analysis was performed on participants' certainty scores to identify predictors of their choice of the culprit. The regression only identified the aggressivity score as a significant predictor of participants' certainty ($r = 0.553$, $p = 0.009$). A log-linear regression was conducted to analyze the relationship between 'role', 'communication style', and participants' ranking of conversations. The analysis only indicated a significant association between 'communication style' and participants' preference ($r = 0.170$ ($SE = 0.057$), $z = 3.004$, $p = 0.003$). A MANOVA was conducted on participants' ratings and behavioral measures, with the ranking of the conversation as a factor. The analysis indicated a significant difference in the scores of the items 'predictability' ($F(1, 124) = 2.813$, $p = 0.042$) and 'aggressiveness' ($F(1, 124) = 5.143$, $p = 0.002$). The most and least preferred conversations were rated as highly predictable, with the preferred one rated as less aggressive than the other. The length of the conversation was the only significant effect on participants' ranking ($F(3, 1079) = 3.621$, $p = 0.013$). The preferred conversation was significantly longer compared to the other ones.

7. Discussion

The experiment's results provide insights into users' experience when interacting with CAs in a detective computer game. By combining ratings of perceived intelligence and believability with participants' behaviors during conversations, a comprehensive understanding of how CAs' design influences user experience was obtained. The analysis revealed that the design parameters of CAs have a distinct impact on participants' ratings. Participants' attention was rated higher when they encountered witness CAs, while the communication style influenced their perception of aggressiveness and warmth. The factor 'role' significantly influenced the evaluation of perceived

intelligence and believability of the first encountered CA. The CAs' communication style had a significant impact on the participants' behaviors during the conversations. Participants wrote longer inputs and had larger conversations when confronted with aggressive CAs. In summary, the identification of the culprit was influenced by the perceived aggressiveness of the CA, regardless of its role in the narrative. Regarding users' preference between the conversations, cooperative ones were preferred. In the following, the results of the factors 'role' and 'communication style' are discussed separately.

The initial result that emerged from the experiment is the influence of the roles of CAs on users' ratings of perceived intelligence and believability, regardless of their communication style. Explicit roles aim to activate stereotypes in users' minds, and thus generate expectations. In a police investigation, suspects and witnesses are recognized to engage in distinct types of interactions with investigators. Interactions with suspects typically involve highly challenging and argumentative conversations, as they are expected to defend their alibis. On the contrary, witnesses readily provide crucial information to facilitate the progression of the investigation. In the experiment, roles were explicitly communicated to participants before the interactions, allowing them to anticipate the conversations and consequently determine their conversational strategy towards the CA. This included predicting the topic of the detective's inquiries and adjusting their approach accordingly. The strategy adopted by the participants may have influenced their evaluation of the encountered CAs, particularly during the first interaction. Using the item 'Visual impact' of the believability scale (c.f., the replacement of 'behavior' by 'discourse', participants rated their attention drawn by the discourse of the CA), participants indicated the significant difference of their attention when they were interacting with witnesses. Participants were more attentive to the discourse of witnesses as they likely anticipated them to provide valuable information that could help identify the culprit. Subsequently, participants rated the perceived intelligence of suspects significantly higher than that of witnesses. When conversing with suspects, participants were more inclined towards suspicion and accusatory inquiries, while conversations with witnesses tended to be more informative in nature. The different strategies adopted by participants reflect their underlying motivation during the interaction, as suspects are implicitly more likely to be identified as the culprit. Participants may have perceived the suspects as being more intelligent because they responded to accusations and suspicions, whereas the strategy employed towards witnesses involved informative inquiries,

which typically resulted in less argumentative responses. The distinction is observed in participants' behaviors as their input length and delay were affected by the roles and order of CAs. The difference between the two roles and the changes over time suggest different strategies based on CAs' roles but also the context of the conversation. Notably, participants exhibited a tendency to send longer inputs and took longer time to formulate their inquiries when encountering suspects initially. Conversely, participants adopted the opposite approach when engaging with witnesses during the last conversations. This shift in strategy suggests that participants initially prioritized interactions with suspects, anticipating more crucial information in order to accomplish their objective. As the conversations progressed, participants appeared to adjust their strategy and shift their focus towards gathering information from witnesses to uncover potential culprits. Therefore, CAs' explicit roles affected users' expectations which were reflected through participants' rating of the four CAs and their behaviors. Analysis of users' behaviors suggested they applied different strategies to achieve their goal (i.e., naming the culprit) based on CAs' role.

The second result of the experiment is the importance of the CA's communication style to users' perception of aggressivity and warmth and its impact on users' behaviors. CAs with an aggressive communication style were perceived as more aggressive and colder in comparison to cooperative ones. Although there was no significant effect of communication style on participants' perception of intelligence and believability, the communication style did impact the item 'Personality' of the believability scale regarding the suspects. While there was no significant effect of communication style on participants' perception of intelligence and believability, the communication style did have an impact on the item 'Personality' of the believability scale specifically related to the suspects. In the experiment, the personality of cooperative suspects was rated significantly lower compared to the other conditions. Participants were inclined to accuse and argue with suspects, but the cooperative suspects surpassed their expectations by defusing tensions and responding calmly to their accusatory inquiries. This finding is consistent with Magerko's perspective (2007) on defining believability as a metric for artificial agents, wherein observers evaluate them based on their expectations. Loyall (1997) further explains that the dimension of personality in believability attribution is not solely an assessment of the agent's behavior but also reflects users' acknowledgment of the agent's distinctiveness, serving as a strong motivator for user engagement. As a result, suspects with an unexpected communication style receive lower rankings in terms of

personality attribution, which could potentially hinder their effectiveness as engaging characters in a narrative experience. Participants' inputs and the number of turns during their conversations were also affected by CAs' communication style. Participants conversing with aggressive CAs wrote larger inputs and engaged in longer conversations. These findings diverge from the results of the experiment conducted by Ruane et al. (2021). In their study, participants tended to mimic the communication style of the chatbot while in the current experiment, aggressive CAs exhibited smaller outputs compared to cooperative ones. The difference between our results and those of Ruane et al. (2021) can be attributed to the varying contexts of the experiments. Here, the conversations took place within the context of a computer game (i.e., a narrative experience), whereas in Ruane et al. (2021) experiment, participants answered chatbots' questions regarding the experience they have about university life. In a narrative experience such as a detective game, aggressive CAs might trigger suspicion in users' minds, leading them to focus on their inputs, inducing longer sentences, and engaging in lengthier conversations to achieve their goals. Furthermore, the aggressive communication style emerged as the only predictor of participants' identification of the culprit, highlighting the stronger influence of communication style over initial expectations. Aggressive CAs were significantly more frequently identified as the culprit, indicating an implicit association between aggressiveness and guiltiness (Infante and Wigley, 1986).

Additionally, participants' preference for the conversations was influenced by CAs' communication style. Participants tended to rank conversations higher when CAs were cooperative regardless of their role. The analysis of participants' behaviors indicated that within cooperative conversations, the preferred ones tended to be longer. These results align with the hypothesis made by Ruane et al. (2021) regarding users' engagement and the length of the conversations with CAs. Furthermore, participants' ratings outlined the importance of CAs' predictability on their conversations' ranking. In narrative experiences, Loyall (1997) highlighted the importance of predictability for users to anticipate their interaction based on their expectations as it enhances their enjoyment. In the current experiment, preferred conversations were the ones perceived as highly predictable and cooperative. To enhance users' preferences, CAs' designers in narrative experience should manipulate CAs' roles and communication styles to reduce the gap between users' expectations and the conversation's tone.

The results of this experiment outline the importance of the role and communication style of CAs on both users' evaluation of CAs and their behaviors during the conversations. Participants perceived suspects as more intelligent than witnesses, but this result can be attributed to the different conversational strategies employed by users during their interactions. Furthermore, suspects with unexpected communication styles, such as being friendly, received lower ratings in terms of personality, potentially leading to a decrease in user engagement in the narrative experience. Aggressive communication style was highlighted as a significant predictor of being named as the culprit, regardless of the role. The difference was also observed through the different users' behaviors. Taken together, these findings suggest that considering users' expectations and perceptions by aligning the roles of CAs with their communication styles in a game, such as a detective game, can enhance user engagement and create a more captivating narrative experience. By ensuring a proper correspondence between the roles and communication styles of CAs, game designers can optimize user engagement and enhance the overall quality of the narrative experience.

8. Conclusion

This experiment highlights the importance of considering users' expectations in narrative experiences. The approach adopted here enables one to manipulate the roles and communication styles of different CAs and gain insight of their impact on users' behaviors and on their perception of intelligence and believability. The information gathered from the experiment is crucial for creating engaging CAs that effectively convey the narrative through their interactions with users. By understanding the impact of factors such as roles, communication styles, and user expectations, developers can design CAs that enhance the immersive and interactive nature of the narrative, leading to a more enjoyable and compelling user experience. Precisely, manipulating the parameters of role and communication style affected users' experience. The linguistic cues influenced users' behaviors and their perception of aggressiveness. These significant changes follow the studies made on the impact of personality traits on users' experience with CAs. The results highlight the significance of communication style, regardless of the role, in identifying the culprit in a detective game. However, the role itself plays a crucial role in shaping users' expectations and their attitude toward the CA. These findings have broader implications for the

design of CAs in different narrative contexts, outlining the importance of attentively considering users' expectations and perceptions. By carefully aligning roles and communication styles, developers can create more immersive and engaging experiences for users in various narrative scenarios. These results can be extended to other types of agents, such as embodied conversational agents. All types of agents, whether they are virtual assistants, chatbots, or even virtual characters in games, can greatly benefit from considering users' expectations regarding their nonverbal behaviors. By incorporating appropriate nonverbal cues, such as gestures, facial expressions, and body language, agents can enhance the user experience and increase engagement. When combined with suitable roles and communication styles, this holistic approach can create more believable and immersive interactions, leading to improved user satisfaction and enjoyment.

9. Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

10. Author Contributions

Conceptualization: RP, MLM, CP; MA; Design and Methodology: RP, MLM, CP, MA; Conduction of the study: RP; Statistical analysis and interpretation: RP, CP, MA; Writing – original draft preparation: RP; Writing – review and editing: RP, MLM, CP, MA; Resources: RP; Supervision: CP, MA.

11. Funding

Remi Poivet's research is funded by the French ANRT and the company Ubisoft.

12. Acknowledgments

We thank Ubisoft's writer Joakim Iboudghacene for creating the different characters' personalities of the experiment and kindly endorsed the Wizard of Oz's role.

13. References

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

Brahnam, S. and De Angeli, A., 2012. Gender affordances of conversational agents. *Interacting with Computers*, 24(3), pp.139-153.

Følstad, A. and Skjuve, M., 2019, August. Chatbots for customer service: user experience and motivation. In *Proceedings of the 1st international conference on conversational user interfaces* (pp. 1-9).

Gomes, P., Paiva, A., Martinho, C. and Jhala, A., 2013. Metrics for character believability in interactive narrative. In *Interactive Storytelling: 6th International Conference, ICIDS 2013, Istanbul, Turkey, November 6-9, 2013, Proceedings 6* (pp. 223-228). Springer International Publishing.

Heylighen, F. and Dewaele, J.M., 2002. Variation in the contextuality of language: An empirical measure. *Foundations of science*, 7, pp.293-340.

Infante, D.A. and Wigley III, C.J., 1986. Verbal aggressiveness: An interpersonal model and measure. *Communications Monographs*, 53(1), pp.61-69.

Infante, D.A., 1995. Teaching students to understand and control verbal aggression. *Communication Education*, 44(1), pp.51-63.

Isbister, K. and Nass, C., 2000. Consistency of personality in interactive characters: verbal cues, non-verbal cues, and user characteristics. *International journal of human-computer studies*, 53(2), pp.251-267.

Klüwer, T., 2011. From chatbots to dialog systems. In *Conversational agents and natural language interaction: Techniques and Effective Practices* (pp. 1-22). IGI Global.

Koda, T. and Maes, P., 1996, November. Agents with faces: The effect of personification. In *Proceedings 5th IEEE international workshop on robot and human communication. RO-MAN'96 TSUKUBA* (pp. 189-194). IEEE.

Komatsu, T. and Yamada, S., 2011. Adaptation gap hypothesis: How differences between users' expected and perceived agent functions affect their subjective impression. *Journal of Systemics, Cybernetics and Informatics*, 9(1), pp.67-74.

Loyall, A.B., 1997. *Believable agents: building interactive personalities* (Doctoral dissertation, Carnegie Mellon University.).

Magerko, B., 2007. Evaluating Preemptive Story Direction in the Interactive Drama Architecture. *J. Game Dev.*, 2(3), pp.25-52.

Mairesse, F. and Walker, M., 2007, June. PERSONAGE: Personality generation for dialogue. In *Proceedings of the 45th annual meeting of the Association of Computational Linguistics* (pp. 496-503).

Mairesse, F. and Walker, M.A., 2009. *Can conversational agents express big five personality traits through language?: Evaluating a psychologically-informed language generator*. Cambridge & Sheffield, United Kingdom: Cambridge University Engineering.

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

Mou, Y. and Peng, W., 2009. Gender and racial stereotypes in popular video games. *Handbook of research on effective electronic gaming in education*, pp.922-937.

Moussawi, S., Koufaris, M. and Benbunan-Fich, R., 2021. How perceptions of intelligence and anthropomorphism affect adoption of personal intelligent agents. *Electronic Markets*, 31, pp.343-364.

Nag, P. and Yalçın, Ö.N., 2020, October. Gender stereotypes in virtual agents. In Proceedings of the 20th ACM International conference on intelligent virtual agents (pp. 1-8).

Pennebaker, J.W. and King, L.A., 1999. Linguistic styles: language use as an individual difference. *Journal of personality and social psychology*, 77(6), p.1296.

Resendez, V., 2020. A very formal agent: how culture, mode of dressing and linguistic style influence the perceptions toward an Embodied Conversational Agent? (Master's thesis, University of Twente).

Ruane, E., Farrell, S. and Ventresque, A., 2021. User perception of text-based chatbot personality. In *Chatbot Research and Design: 4th International Workshop, CONVERSATIONS 2020, Virtual Event, November 23–24, 2020, Revised Selected Papers 4* (pp. 32-47). Springer International Publishing.

de Sá Siqueira, M.A., Müller, B.C. and Bosse, T., 2023. When do we accept mistakes from chatbots? The impact of human-like communication on user experience in chatbots that make mistakes. *International Journal of Human–Computer Interaction*, pp.1-11.

Schaffer, O. and Fang, X., 2019. Digital game enjoyment: a literature review. In *HCI in Games: First International Conference, HCI-Games 2019, Held as Part of the 21st HCI International Conference, HCII 2019, Orlando, FL, USA, July 26–31, 2019, Proceedings 21* (pp. 191-214). Springer International Publishing.

Simpson, J., Gaiser, F., Macík, M. and Breßgott, T., 2020, July. Daisy: a friendly conversational agent for older adults. In *Proceedings of the 2nd conference on conversational user interfaces* (pp. 1-3).

Völkel, S.T., Schödel, R., Buschek, D., Stachl, C., Winterhalter, V., Bühner, M. and Hussmann, H., 2020, April. Developing a personality model for speech-based conversational agents using the psycholexical approach. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1-14).

Warner, R.M. and Sugarman, D.B., 1986. Attributions of personality based on physical appearance, speech, and handwriting. *Journal of personality and social psychology*, 50(4), p.792.

CHAPTER 5: General Discussion

This thesis investigates the role of NPCs' design in players' experiences, specifically focusing on their evaluations of NPCs' perceived intelligence and believability — two dimensions crucial for creating engaging virtual agents in narrative experiences. The main hypothesis posits that players would evaluate NPCs more positively when their design aligns with players' expectations. The first experiment conducted on players' evaluations of NPCs focused on hostility and how design parameters, namely appearance and behavior, affect players' assessments of NPCs' personality attributes and hostility. The second experiment examined how the consistency and coherence of NPCs' design parameters influence players' expectations and their assessments of NPCs' perceived intelligence and believability. The final experiment explored an additional modality of interactions with NPCS through the relationship between their role in the game and their communication style in textual computer games. It focused on how these factors affect players' evaluations of perceived intelligence and believability. After summarizing the results briefly, the discussion will center on the thesis's contribution to human evaluations of virtual agents, particularly NPCs, and its implications for game developers.

Insights on players' evaluation of NPCs' hostility

Before investigating the relationship between NPCs' design and players' expectations, we needed to understand how players perceive NPCs' roles in the game. We chose to focus on the role of enemies because their function is crucial in shaping the gaming experience and represents one of the most common forms of interaction in video games. Game designers face the challenge of selecting parameters to convey NPCs' hostility efficiently. Therefore, our first study explored players' evaluations of hostility in a military shooter game setting, with a specific focus on NPCs' design and its impact on personality trait ascriptions. Drawing inspiration from the metaphorical design of virtual agents, which relies on visual cues in their appearance to convey their function to users, we hypothesized that players would assess NPCs' hostility based on the perceived threat

conveyed by their appearance and the aggressiveness of their behaviors. The results revealed that behaviors played a significantly stronger role than appearance in the evaluation of hostility within NPCs. NPCs exhibiting aggressive behaviors were consistently rated as hostile, regardless of their appearance. Hostile NPCs were also perceived as unlikable, harmful, and knowledgeable.

Contribution to humans' evaluation of virtual agents

Study 1 provides a deeper understanding of how users gather information from the design of virtual agents when assessing their hostility in video games. Similar to research conducted on the influence of virtual agents' appearance and behaviors on users' evaluation (Dai & MacDorman, 2018; Warpefelt, 2015), we observe that evaluation of NPCs' threat, aggressiveness and hostility is influenced by the cues from their appearance and behaviors.

On the one hand, the evaluation of a virtual agent's threat refers to users' anticipation of potential negative outcomes in interactions. While virtual agents have no direct consequences on users, as mentioned by Bosse et al. (2018), NPCs do impact users through the interaction with players' avatars in the game experience. NPCs using their abilities to harm players' avatars can negatively affect the gaming experience. Consistent with previous findings in the literature (Rogers et al., 2018), we note the effect of visual cues from appearance on personality trait ascription. Players evaluating threats in NPCs also perceive them as more competent, dangerous, and antipathetic. Thus, similar to Warpefelt's results (2015) describing players' identification of NPCs' role and function in the game based on visual cues, players in our study ascribe personality traits associated with their expectations of enemies in a military shooter game.

On the other hand, we observe that aggressiveness is the primary factor conveying hostility, with no significant effect of appearance threat. These findings confirm that in the context of video games, especially military shooter games, players' evaluations of hostility are primarily based on NPCs' aggressive behaviors. In players' mental models, enemies are reduced to their functions: convey the challenge of the game experience. Our results confirm that such role can be identified through NPCs' behaviors only. Hence, an NPC with a non-threatening appearance that does not rely on items from their appearance to convey threat can still be identified as an enemy if their behaviors are aggressive. While the evaluation of other agents' threats always occurs during social interactions (Blascovich et al., 2001), these findings shed light on the crucial role of behaviors over appearance on hostility evaluation during players' interactions with NPCs.

Contribution to game designers

One of the challenges for game designers is to effectively convey the role of NPCs in the gaming experience. While we initially posited that a combination of appearance and behaviors would influence hostility perception, it became apparent that behaviors are the primary factor. Game designers can draw upon this study to make more informed decisions when crafting NPCs' parameters.

Although NPCs' threat conveyed by their appearance does not impact hostility perception, it may influence the attribution of other traits. For instance, creating enemies that possess threatening appearance to convey the amount of challenge in the game experience could help players make better decisions to interact. Players could hierarchize enemies following NPCs' threat and base their order of interaction from it. Another trait ascription is the charisma of the character. Take the iconic character of Darth Vader as an example in the movie industry. His appearance cleverly draws from samurai and other antagonistic elements in popular culture, including his distinctive cape, which adds an air of malevolence to his presence. This alignment with players' expectations doesn't merely enhance hostility evaluation; it also contributes to his undeniable appeal (Rojcewicz., 1987). Darth Vader's appearance goes beyond being perceived solely as an antagonist; it becomes a source of fascination and charisma, showcasing the power of appearances' design in shaping player perceptions. In contrast, in the tactical shooter game *Ready or Not* (Interactive VOID, 2021), enemies do not always have threatening appearances (see figure 13). In this game, players assume the role of the leader of a police tactical unit engaged in rescue missions in hostile environments. The game experience relies on players' need to make accurate decisions about shooting at NPCs, as any delay in shooting at enemies can be fatal. Game designers convey hostile roles using only NPCs' behaviors to create a tense game experience, forcing players to make quick decisions when encountering an NPC. The tension in the game experience relies on the lack of information conveyed by NPCs' appearance when they are not threatening.

Consequently, our study can be used by game designers as confirmation that behaviors shape players' evaluation of hostility in a military shooter game. However, it does not suggest setting aside the consideration of appearance, as it still contributes to players' overall game experience.



Figure 13: In the game *Ready or not* (Interactive VOID, 2021), players have to make quick and accurate decisions when facing an NPC.

Insights on players' evaluation of unexpected NPCs

When investigating the evaluation of NPC's hostility, we manipulated two key parameters of their design: appearance and behavior. While behavior emerged as the primary factor influencing hostility perception, we hypothesized that appearance still plays a crucial role in establishing accurate players' expectations when evaluating NPCs' behaviors. Therefore, in study 2, we explored how the violation of initial players' expectations affected their evaluations of NPCs' perceived intelligence and believability. Additionally, we analyzed how players can adapt their expectations of NPCs when confronted with multiple instances of unexpected behaviors during a gaming experience. The results shed light on the respective roles of coherence and consistency in NPCs' design parameters on player evaluations. Specifically, NPCs that align with players' expectations (i.e., that are coherent) receive more positive evaluations, but the consistency of their design can be used to modulate players' expectations to align with initially unexpected NPC designs.

Contribution to humans' expectations

One of the challenges for players is to efficiently engage with NPCs, particularly hostile ones, within the gaming experience. Before starting a gaming experience, players form expectations about the various roles within a particular game type. In the context of a military shooter game, players may expect to defeat NPCs with soldier-like appearances to protect NPCs with civilian-like appearances. In study 2, we manipulated NPCs' design parameters (i.e., appearance and behaviors) to convey false information and disrupt players' expectations. Our study used both direct and indirect measures of players' expectations, including explicit ratings of NPCs' appearance, players' behaviors, and physiological measures during interactions. Our approach revealed that interactions with NPCs that align with initial expectations have a positive impact on the gaming experience. Players make better decisions and have more confidence in their expectations. NPCs' design that does not align with players' initial expectations affect players' evaluation of them. However, the consistency of NPCs' design during the game experience enables players to adjust their expectations accordingly. Following the Meaning Maintenance and Violation of Expectations Models (Gollwitzer et al., 2018; Doering et al., 2018; Proulx et al., 2012; Proulx & Inzlicht, 2012), players gather information about encountered NPCs that challenge their expectations to maintain their understanding of the gaming experience. Players interacting with consistent NPCs that defy their initial expectations are capable of accommodating their mental models. However, we observed that for players to process new information that contradicts initial expectations, they must identify a pattern. In our context, players facing NPCs that are consistently incoherent understood after a few interactions that characters with a civilian appearance were enemies for instance, allowing them to adjust their expectations accordingly. In comparison, players facing inconsistent disruptions of their expectations were unable to adjust, resulting in a lack of learning during their gaming session, particularly regarding the distance they kept from enemies. Additionally, the higher level of arousal during interactions with NPCs suggests that the uncertainty induced by their inability to adjust their expectations conveyed more stress in the game experience (Starcke & Brand, 2016). Game designers can leverage this finding to align with their intentions when selecting NPCs' parameters for game experiences that rely on tension within hostile interactions. Similar to *Ready or Not* (Interactive VOID, 2021), altering NPCs' appearances to disrupt players' expectations and reduce the amount of conveyed information can be an

intriguing strategy. This approach compels players to base their decisions more on reactivity than anticipation. Consequently, players encountering consistent NPCs with specific design parameters can form efficient expectations about interaction outcomes, regardless of the coherence of those parameters. These findings provide fresh insights into the dynamics of human expectations in the context of video games, namely that players gain from the limited behaviors of NPCs as they identify patterns within their design to make accurate expectations about them.

Contribution to perceived intelligence

Within the gaming experience, NPCs are assessed for their intelligence, despite their inherently artificial nature. To convey the illusion of intelligence, game designers must meticulously manipulate specific design parameters. In this manuscript, we align NPCs' perceived intelligence with the concept defined in existing literature (Bartneck et al., 2009). Perceived intelligence, as we use it here, encapsulates users' assessments of the apparent logic of artificial agents' behaviors and their performance in executing tasks. We posited that players would rely on their mental models to form expectations to assess NPCs' perceived intelligence during their interactions. Consequently, NPCs aligning with players' expectations would receive positive evaluations. Study 2 confirms our primary hypothesis: players possess initial expectations that contribute to their evaluation of NPCs. When the information conveyed by NPCs' appearance does not align with their behaviors, it disrupts players' expectations. In such cases, players are less capable of explaining NPCs' behaviors, resulting in a decrease in their evaluation of NPCs' intelligence as it affects their apparent logic.

Comparing our findings to the research by Butcher and Griesemer (2004) on players' assessments of enemies' intelligence, we observed that civilians who exhibited aggressive behaviors were not perceived as intelligent when compared to their counterparts who displayed the expected submissive behaviors. Butcher and Griesemer emphasize that enemies' intelligence is primarily conveyed through the level of challenge they introduce into the gaming experience. Interestingly, in the context of aggressive civilians, despite presenting a level of challenge similar to aggressive soldiers, they did not evoke the same perception of intelligence. Furthermore, players who encountered only aggressive civilians and formed expectations about them during their gaming experience evaluated submissive civilians more favorably because they conformed to their initial expectations. Additionally, players who encountered inconsistent designs, encompassing both

NPCs that aligned and disrupted their initial expectations, were inclined to assess more positively NPCs with design parameters aligning with their initial expectations. These results indicate that perceived intelligence of NPCs is influenced by players' initial expectations and cannot be enhanced through the accommodation of their mental models. It is important to note that our results are constrained by the military shooter game setting, where interactions with NPCs were limited to shooting them or not. Therefore, in another type of game, such as an RPG, interactions can be more complex, and NPCs' perceived intelligence may vary accordingly.

Contribution to Believability

In this manuscript, we have chosen to embrace the approach of defining believability as players' investment of belief in the NPCs they encounter, which involves accepting their stereotyped behaviors. Several dimensions were listed by Gomes et al. (2013) as contributing to the creation of believable characters in video games, and we posited that one of them, coherence, was crucial for players to believe in NPCs. Therefore, we hypothesized that creating NPCs with design parameters aligning with players' expectations would enhance their believability. The study confirmed our hypothesis, as the disruption of players' initial expectations led to lower ratings regarding the dimensions of believability. Players were less inclined to believe in NPCs with unexpected behaviors. Disrupting their expectations triggers surprise in players' minds, which could have an interesting effect on the game narratives (Lavik, 2006). However, it also influences how players react to NPCs. In their discussion of these dimensions, Gomes et al. (2013) emphasized the importance of predictability in modulating a character's believability. Characters that are either overly predictable or not predictable enough may not be perceived as believable (Loyall, 1997). In our context, players rely on their mental model of enemies and innocents to predict NPCs' behaviors. Hence, players' initial expectations appear to be crucial in their evaluation of NPCs' believability, as they determine what is or is not believable from their point of view. NPCs with initially unexpected behaviors were not considered believable, but certain dimensions, such as visual impact and personality ascription, were positively affected. In this sense, breaking players' initial expectations could benefit the narrative by increasing players' attention towards NPCs and making them stand out when compared to others (Loyall, 1997). Furthermore, our study revealed that players facing only aggressive civilians had similar evaluations, underscoring the importance of initial expectations. This resulted in a more positive evaluation of submissive civilians.

However, players who interacted solely with aggressive civilians rated their believability more positively compared to those who encountered only submissive ones. Players who exclusively interacted with NPCs that disrupted their initial expectations tended to adjust their mental model and expectations based on the identification of associative patterns between appearance and behaviors. Our findings emphasize the pivotal role of players' initial expectations in evaluating NPCs' believability. Interestingly, players' assessment of believability can be enhanced through consistent interactions with NPCs that challenge their initial expectations. Game designers can leverage these results to innovate in their character design by exploring unexpected combinations of NPCs' parameters, knowing that players can adapt and learn from their consistent interactions with such characters.

Insights on players' evaluation of Conversational Agents

To delve deeper into players' interactions with NPCs, we explored their evaluation of perceived intelligence and believability in a different modality of interaction: textual interactions. Conversations with NPCs are constrained by the technological limitations of generating text-based dialogues. Therefore, similar to non-verbal behaviors, NPCs' conversations are limited by their role in the narrative. The recent advancements in large language models present an opportunity for game designers to create less restricted and more engaging NPC dialogues. We explored this possibility using a Wizard of Oz approach in the setting of a computer text-based game to compare how expectations and conversational agents' (CAs) design affect players' evaluations of perceived intelligence and believability. We posited that players would use the communicated role of the characters to form expectations about their communication style similarly to NPCs' appearance and behaviors. Study 3 revealed that players do rely on the conversational roles within the narrative to shape their expectations. However, it wasn't crucial for the communication style to match players' expectations to receive positive assessments regarding their perceived intelligence and believability.

Contribution to perceived intelligence and believability

Based on the results obtained in our investigation of non-verbal interactions, we would have anticipated that players rely on their expectations to evaluate the perceived intelligence and believability of CAs. However, this wasn't the case. In that sense, players in the computer textual

game study did use the role of conversational agents in the narrative to form expectations, but their communication style did not impact perceived intelligence nor believability. Players relied on their mental models associated with CAs' roles to form expectations that shaped their strategies during their conversations and influenced their perception of CAs' intelligence. This finding is similar to players using appearances to form expectations during non-verbal interactions. These findings shed light on the possibilities of creating natural conversations in video games. Players interacting with CAs in the narrative were inclined to believe in their interactions and perceived intelligence regardless of the manipulated design. In comparison to players' mental models of NPCs and their expectations of the limitations of their behaviors during interactions, CAs' behaviors in our study were enhanced through the use of WoZ. In that sense, players' evaluation of perceived intelligence and believability was influenced by their observation of the coherent answers of the CAs during their conversation. Additionally, the quality of the characters' dialogues were better compared to what is currently observed in video games, hence it might have positively surprised players (Burgoon, 2015), which induced overall positive evaluation of the CAs regardless of their aggressiveness or cooperation during the conversations. Our findings suggest that players interacting with NPCs that are able to generate coherent conversation have a significant impact on their evaluation of perceived intelligence and believability. As this technology continues to grow within the game industry, game designers have the opportunity to use this additional modality of interaction to create new forms of gaming experiences that contribute to a more immersive and engaging style of video games.

Conclusion and opportunities for future works

In conclusion, this thesis has uncovered key insights into players' interactions with NPCs in video games, focusing on their assessments of NPCs' perceived intelligence and believability. Our research, spanning multiple experiments, highlighted the pivotal role of NPCs' behaviors, especially aggressiveness, in shaping players' evaluations of hostility. This finding underscores the significance of behavior over appearance in conveying NPCs' roles and intentions, offering valuable guidance to game designers. Furthermore, our work emphasized the importance of players' initial expectations in shaping their perceptions. NPCs that aligned with these expectations

were favorably evaluated, emphasizing the need for design coherence and consistency to enhance believability and perceived intelligence.

These findings provide a strong foundation for future research. Expanding on our investigations of NPCs' appearance and behaviors for hostility, it would be valuable to explore these dynamics in different gaming contexts. For instance, examining how players perceive and assess hostility in cartoon-like games such as *Mario + Rabbids Sparks of Hope* (Ubisoft, 2023) could provide insights into how these design parameters function across diverse genres and art styles. Moreover, building on our exploration of the variation in NPCs' appearance and behaviors and their impact on perceived intelligence and believability, future studies could delve deeper into the nuances of this relationship. Investigating how subtle alterations in NPCs' design parameters influence players' assessments of believability and perceived intelligence could offer a deeper understanding of the relationship between these dimensions. Additionally, within the context of verbal interactions with NPCs, there is potential for research on NPCs' trustworthiness. Trustworthiness in this context would refer to the reliability of the text generated by NPCs (Brahnam, 2009). Examining how players perceive the trustworthiness of information provided by generated text in conversations with NPCs could open avenues for enhancing players' trust in virtual agents. Understanding the factors that contribute to players' willingness to rely on NPCs for critical in-game information would be a valuable contribution to both the gaming and AI communities. In line with our research on novel game design approaches, researchers and game designers alike could continue to experiment with innovative methods of enhancing players' experiences. Similar to our investigations into conversational agents, future studies might explore how multimodal evaluations encompassing appearance, verbal, and non-verbal behaviors collectively affect players' evaluations of believability and perceived intelligence. This approach could provide a holistic view of how different design elements synergize to shape player perceptions and their engagement in the interactions with NPCs.

In conclusion, this research offers a solid foundation for future investigations in the field of players' interactions with NPCs, their evaluations of virtual agents, and the design elements that contribute to engaging gaming experiences. By exploring these avenues, researchers and game designers can continue to enhance the quality of interactions with NPCs and players' experiences in the dynamic world of video games.

REFERENCES

Afonso, N., & Prada, R. (2009). Agents that relate: Improving the social believability of non-player characters in role-playing games. In *Entertainment Computing-ICEC 2008: 7th International Conference, Pittsburgh, PA, USA, September 25-27, 2008. Proceedings 7* (pp. 34-45). Springer Berlin Heidelberg.

André, E., Müller, J., & Rist, T. (1996, May). The PPP persona: a multipurpose animated presentation agent. In *Proceedings of the workshop on Advanced visual interfaces* (pp. 245-247).

Adamopoulou, E., & Moussiades, L. (2020). An overview of chatbot technology. In IFIP international conference on artificial intelligence applications and innovations (pp. 373-383). Springer, Cham.

Bailey, C., & Katchabaw, M. (2008, November). An emergent framework for realistic psychosocial behaviour in non player characters. In *Proceedings of the 2008 Conference on Future Play: Research, Play, Share* (pp. 17-24).

Bartneck, C., Kanda, T., Mubin, O., & Al Mahmud, A. (2009). Does the design of a robot influence its animacy and perceived intelligence?. *International Journal of Social Robotics, 1*, 195-204.

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

Bavelas, J., Gerwing, J., & Healing, S. (2014). Hand and facial gestures in conversational interaction.

Bell, S., Wood, C., & Sarkar, A. (2019, May). Perceptions of chatbots in therapy. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1-6).

Biancardi, B., Wang, C., Mancini, M., Cafaro, A., Chanel, G., & Pelachaud, C. (2019, September). A computational model for managing impressions of an embodied conversational agent in real-time. In *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)* (pp. 1-7). IEEE.

Biocca, F., Kim, J., & Choi, Y. (2001). Visual touch in virtual environments: An exploratory study of presence, multimodal interfaces, and cross-modal sensory illusions. *Presence: Teleoperators & Virtual Environments*, *10*(3), 247-265.

Bosse, T., Hartmann, T., Blankendaal, R. A., Dokter, N., Otte, M., & Goedschalk, L. (2018, July). Virtually bad: a study on virtual agents that physically threaten human beings. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems* (pp. 1258-1266).

Brahnam, S. (2009). Building character for artificial conversational agents: Ethos, ethics, believability, and credibility. *PsychNology Journal*, *7*(1).

Brown, E., & Cairns, P. (2004, April). A grounded investigation of game immersion. In *CHI'04 extended abstracts on Human factors in computing systems* (pp. 1297-1300).

Brusk, J. (2014). Steps Towards Creating Socially Competent Game Characters.

Burgoon, J. K., Jensen, M. L., Kruse, J., Meservy, T. O., & Nunamaker, J. F. (2007). Deception and intention detection. *Handbooks in information systems*, *2*, 187-208.

Burgoon, J. K. (2015). Expectancy violations theory. *The international encyclopedia of interpersonal communication*, 1-9;

Burleigh, T. J., Schoenherr, J. R., & Lacroix, G. L. (2013). Does the uncanny valley exist? An empirical test of the relationship between eeriness and the human likeness of digitally created faces. *Computers in human behavior*, 29(3), 759-771.

Cafaro, A., Vilhjálmsón, H. H., & Bickmore, T. (2016). First impressions in human--agent virtual encounters. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 23(4), 1-40.

Chowanda, A., Flintham, M., Blanchfield, P., & Valstar, M. (2016). Playing with social and emotional game companions. In *Intelligent Virtual Agents: 16th International Conference, IVA 2016, Los Angeles, CA, USA, September 20–23, 2016, Proceedings 16* (pp. 85-95). Springer International Publishing.

Coleridge, S. T., Bonnamy, Y., Burton, R., Devlin, W., & Neville, J. (1985). *Samuel Taylor Coleridge* (p. 515). Oxford: Oxford University Press.

Dai, Z., & MacDorman, K. F. (2018). The doctor's digital double: how warmth, competence, and animation promote adherence intention. *PeerJ Computer Science*, 4, e168.

Dechant, M. J., Welsch, R., Frommel, J., & Mandryk, R. L. (2022, April). (Don't) stand by me: How trait psychopathy and NPC emotion influence player perceptions, verbal responses, and movement behaviours in a gaming task. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (pp. 1-17).

Desai, N., Zhao, R., & Szafron, D. (2016). Effects of gender on perception and interpretation of video game character behavior and emotion. *IEEE Transactions on Computational Intelligence and AI in Games*, 9(4), 333-341.

Dermouche, S., & Pelachaud, C. (2019, October). Generative model of agent's behaviors in human-agent interaction. In *2019 International Conference on Multimodal Interaction* (pp. 375-384).

Doering, B. K., Glombiewski, J. A., & Rief, W. (2018). Expectation-focused psychotherapy to improve clinical outcomes. *International Review of Neurobiology*, *138*, 257-270.

Drennan, P., Viller, S., & Wyeth, P. (2004). Engaging game characters: Informing design with player perspectives. In *Entertainment Computing–ICEC 2004: Third International Conference, Eindhoven, The Netherlands, September 1-3, 2004. Proceedings 3* (pp. 355-358). Springer Berlin Heidelberg.

Ekman, P. (2004). Emotional and conversational nonverbal signals. In *Language, knowledge, and representation: Proceedings of the sixth international colloquium on cognitive science (ICCS-99)* (pp. 39-50). Dordrecht: Springer Netherlands.

Ermi, L., & Mäyrä, F. (2005, June). Fundamental components of the gameplay experience: Analysing immersion. In *DiGRA Conference*.

Even, C., Bosser, A. G., & Buche, C. (2018, October). Bot believability assessment: A novel protocol & analysis of judge expertise. In *2018 International Conference on Cyberworlds (CW)* (pp. 96-101). IEEE.

Ferstl, Y., Kokkinara, E., & McDonnell, R. (2017). Facial features of non-player creatures can influence moral decisions in video games. *ACM Transactions on Applied Perception (TAP)*, *15*(1), 1-12.

Ferstl, Y., & McDonnell, R. (2018, November). A perceptual study on the manipulation of facial features for trait portrayal in virtual agents. In *Proceedings of the 18th International Conference on Intelligent Virtual Agents* (pp. 281-288).

Fleming, R., Mohler, B. J., Romero, J., Black, M. J., & Breidt, M. (2016, February). Appealing Female Avatars from 3D Body Scans: Perceptual Effects of Stylization. In *VISIGRAPP (1: GRAPP)* (pp. 335-345).

Friston, K., Samothrakis, S., & Montague, R. (2012). Active inference and agency: optimal control without cost functions. *Biological cybernetics*, 106, 523-541.

Fukayama, A., Ohno, T., Mukawa, N., Sawaki, M., & Hagita, N. (2002, April). Messages embedded in gaze of interface agents---impression management with agent's gaze. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 41-48).

Giusti, R., Hullett, K., & Whitehead, J. (2012, May). Weapon design patterns in shooter games. In *Proceedings of the First Workshop on Design Patterns in Games* (pp. 1-7).

Gomes, P., Paiva, A., Martinho, C., & Jhala, A. (2013). Metrics for character believability in interactive narrative. In *Interactive Storytelling: 6th International Conference, ICIDS 2013, Istanbul, Turkey, November 6-9, 2013, Proceedings 6* (pp. 223-228). Springer International Publishing.

Gómez-Gauchía, H., & Peinado, F. (2006, December). Automatic customization of non-player characters using players temperament. In *International Conference on Technologies for Interactive Digital Storytelling and Entertainment* (pp. 241-252). Berlin, Heidelberg: Springer Berlin Heidelberg.

Green, R. D., MacDorman, K. F., Ho, C. C., & Vasudevan, S. (2008). Sensitivity to the proportions of faces that vary in human likeness. *Computers in Human Behavior*, 24(5), 2456-2474.

Griesemer J. & Butcher C, "The Illusion of Intelligence", Game Developer Conference, San Jose, 2002.

Gollwitzer, M., Thorwart, A., & Meissner, K. (2018). Psychological responses to violations of expectations. *Frontiers in Psychology*, 8, 2357.

Hamdy, S., & King, D. J. (2017, August). Affect and believability in game characters: A review of the use of affective computing in games. In GAME-ON'2017, 18th annual Conference on Simulation and AI in Computer Games (pp. 90-97). EUROSIS.

Headleand, C. J., Jackson, J., Williams, B., Priday, L., Teahan, W. J., & Ap Cenydd, L. (2016). How the perceived identity of a npc companion influences player behavior. *Transactions on Computational Science XXVIII: Special Issue on Cyberworlds and Cybersecurity*, 88-107.

Heylen, D. (2005). Challenges ahead: Head movements and other social acts in conversations. *Virtual Social Agents*, 45-52.

Johansson, M., Strååt, B., Warpefelt, H., & Verhagen, H. (2013, June). Analyzing the social dynamics of non-player characters. In International Simulation and Gaming Association Conference (pp. 173-187). Cham: Springer International Publishing.

Hoorens, V., & Ramachandran, V. (2012). Expectation.

Isla, D. "Handling Complexity in the Halo 2 AI". In Game Developer Conference, San Francisco, 2005.

Junuzovic, S., Inkpen, K., Tang, J., Sedlins, M., & Fisher, K. (2012, October). To see or not to see: a study comparing four-way avatar, video, and audio conferencing for work. In *Proceedings of the 2012 ACM International Conference on Supporting Group Work* (pp. 31-34).

Kang, S. H., Sidner, C., Gratch, J., Artstein, R., Huang, L., & Morency, L. P. (2011). Modeling nonverbal behavior of a virtual counselor during intimate self-disclosure. In *Intelligent Virtual Agents: 10th International Conference, IVA 2011, Reykjavik, Iceland, September 15-17, 2011. Proceedings 11* (pp. 455-457). Springer Berlin Heidelberg.

Koda, T., Ruttkay, Z., Nakagawa, Y., & Tabuchi, K. (2010). Cross-cultural study on facial regions as cues to recognize emotions of virtual agents. *Culture and Computing: Computing and Communication for Crosscultural Interaction*, 16-27.

Lakin, J. L., & Chartrand, T. L. (2003). Using nonconscious behavioral mimicry to create affiliation and rapport. *Psychological science*, 14(4), 334-339.

Lavik, E. (2006). Narrative structure in The Sixth Sense: a new twist in» twist movies»??. *The Velvet Light Trap*, 58(1), 55-64.

Lee, M. S., & Heeter, C. (2008, May). Computer science and communication perspectives on character believability in games. In Annual Meeting of the International Communication Association.

Lee, M. S., & Heeter, C. (2012). What do you mean by believable characters?: The effect of character rating and hostility on the perception of character believability. *Journal of Gaming & Virtual Worlds*, 4(1), 81-97.

Lester, J. C., & Stone, B. A. (1997, February). Increasing believability in animated pedagogical agents. In *Proceedings of the first international conference on Autonomous agents* (pp. 16-21).

Lester, J. C., Towns, S. G., Callaway, C. B., Voerman, J. L., & FitzGerald, P. J. (2000). Deictic and emotive communication in animated pedagogical agents. *Embodied conversational agents*, 123-154.

Linderoth, J. (2013). Beyond the digital divide: An ecological approach to gameplay. *Transactions of the Digital Games Research Association*, 1(1).

Livingstone, D. (2006). Turing's test and believable AI in games. *Computers in Entertainment (CIE)*, 4(1), 6-es.

Loyall, A. B. (1997). *Believable agents: building interactive personalities* (Doctoral dissertation, Carnegie Mellon University.).

Lugrin, B. (2021). Introduction to socially interactive agents. In *The Handbook on Socially Interactive Agents: 20 years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics Volume 1: Methods, Behavior, Cognition* (pp. 1-20).

MacDorman, K. F., Green, R. D., Ho, C. C., & Koch, C. T. (2009). Too real for comfort? Uncanny responses to computer generated faces. *Computers in human behavior*, 25(3), 695-710.

Mateas, M. (2001). An Oz-centric review of interactive drama and believable agents. In *Artificial intelligence today: Recent trends and developments* (pp. 297-328). Berlin, Heidelberg: Springer Berlin Heidelberg.

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

McDonnell, R., & Mutlu, B. (2021). Appearance. *The handbook on socially interactive agents: 20 years of research on embodied conversational agents, intelligent virtual agents, and social robotics volume 1: Methods, behavior, cognition*, 105-146.

Mitchell, E., & Mamykina, L. (2021, May). From the Curtain to Kansas: Conducting Wizard-of-Oz Studies in the Wild. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1-6).

Moussawi, S., Koufaris, M., & Benbunan-Fich, R. (2021). How perceptions of intelligence and anthropomorphism affect adoption of personal intelligent agents. *Electronic Markets*, 31, 343-364.

Moray, N. (1996, October). A taxonomy and theory of mental models. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 40, No. 4, pp. 164-168). Sage CA: Los Angeles, CA: SAGE Publications.

Nass, C. I., Moon, Y., Morkes, J., Kim, E. Y., & Fogg, B. J. (1997). Computers are social actors: A review of current research. *Human values and the design of computer technology*, 137-162.

Niewiadomski, R., Demeure, V., & Pelachaud, C. (2010). Warmth, competence, believability and virtual agents. In *Intelligent Virtual Agents: 10th International Conference, IVA 2010, Philadelphia, PA, USA, September 20-22, 2010. Proceedings 10* (pp. 272-285). Springer Berlin Heidelberg.

Normoyle, A., Badler, J. B., Fan, T., Badler, N. I., Cassol, V. J., & Musse, S. R. (2013). Evaluating perceived trust from procedurally animated gaze. In *Proceedings of Motion on Games* (pp. 141-148).

Ochanine, D. (1978). Le rôle des images opératives dans la régulation des activités de travail, *Psychologie et éducation*, n 2, mai 1978, 63-72. *Toulouse le Mirail*.

Oertel, C., Castellano, G., Chetouani, M., Nasir, J., Obaid, M., Pelachaud, C., & Peters, C. (2020). Engagement in human-agent interaction: An overview. *Frontiers in Robotics and AI*, 7, 92.

Ohrberg, S. (2019). *Recreating Believability In NPCs: The Effects Of Visual And Logical Behaviour*.

Oliver, R. L. (1980). A cognitive model of the antecedents and consequences of satisfaction decisions. *Journal of marketing research*, 17(4), 460-469.

Oliver, R. L. (2010). Customer satisfaction. *Wiley international encyclopedia of marketing*.

Paradedda, R., Ferreira, M. J., Oliveira, R., Martinho, C., & Paiva, A. (2019, October). The role of assertiveness in a storytelling game with persuasive robotic non-player characters. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play* (pp. 453-465).

Pelachaud, C., Busso, C., & Heylen, D. (2021). Multimodal behavior modeling for socially interactive agents. In *The Handbook on Socially Interactive Agents: 20 Years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics Volume 1: Methods, Behavior, Cognition* (pp. 259-310).

Peters, C., Castellano, G., & De Freitas, S. (2009, November). An exploration of user engagement in HCI. In *Proceedings of the International Workshop on Affective-Aware Virtual Agents and Social Robots* (pp. 1-3).

Pradantyo, R., Birk, M. V., & Bateman, S. (2021). How the visual design of video game antagonists affects perception of morality. *Frontiers in Computer Science*, 3, 531713.

Proulx, T., Inzlicht, M., & Harmon-Jones, E. (2012). Understanding all inconsistency compensation as a palliative response to violated expectations. *Trends in cognitive sciences*, 16(5), 285-291.

Proulx, T., & Inzlicht, M. (2012). The five “A”s of meaning maintenance: Finding meaning in the theories of sense-making. *Psychological inquiry*, 23(4), 317-335.

Rabardel, P. (1995). *Les hommes et les technologies; approche cognitive des instruments contemporains* (p. 239). Armand colin.

Ransom, M., Fazelpour, S., Markovic, J., Kryklywy, J., Thompson, E. T., & Todd, R. M. (2020). Affect-biased attention and predictive processing. *Cognition*, 203, 104370.

Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people*. Cambridge, UK, 10(10).

Roese, N. J., & Sherman, J. W. (2007). Expectancy. In A. W. Kruglanski & E. T. Higgins (Eds.), *Social psychology: Handbook of basic principles* (pp. 91–115). The Guilford Press.

Rogers, K., Aufheimer, M., Weber, M., & Nacke, L. E. (2018, January). Towards the Visual Design of Non-Player Characters for Narrative Roles. In *Graphics Interface* (pp. 154-161).

Rojcewicz, S. (1987). Darth Vader: Masks, Power, and Meaning. *Journal of Poetry Therapy, 1*(1), 23-30.

Rowling, J.K. (1997). *Harry Potter and the Philosopher's Stone*. Bloomsbury.

Saygin, A. P., Chaminade, T., Ishiguro, H., Driver, J., & Frith, C. (2012). The thing that should not be: predictive coding and the uncanny valley in perceiving human and humanoid robot actions. *Social cognitive and affective neuroscience, 7*(4), 413-422.

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

Shiban, Y., Schelhorn, I., Jobst, V., Hörnlein, A., Puppe, F., Pauli, P., & Mühlberger, A. (2015). The appearance effect: Influences of virtual agent features on performance and motivation. *Computers in Human Behavior, 49*, 5-11.

Soni, B., & Hingston, P. (2008, June). Bots trained to play like a human are more fun. In *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)* (pp. 363-369). IEEE.

Szafir, D., & Mutlu, B. (2012, May). Pay attention! Designing adaptive agents that monitor and improve user engagement. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 11-20).

Thomas, F., & Johnston, O. (1981). Disney animation: The illusion of life.

Tjokrosetio, A. N., & Chowanda, A. (2021). Character believability enhancement using facial expression recognition to improve the player's immersive experience. *ICIC Express Letters*, 15, 1235-1242.

Togelius, J., Yannakakis, G. N., Karakovskiy, S., & Shaker, N. (2012). Assessing believability. *Believable Bots: Can Computers Play Like People?*, 215-230.

Veletsianos, G. (2010). Contextually relevant pedagogical agents: Visual appearance, stereotypes, and first impressions and their impact on learning. *Computers & Education*, 55(2), 576-585.

Volante, M., Babu, S. V., Chaturvedi, H., Newsome, N., Ebrahimi, E., Roy, T., Daily, SB., & Fasolino, T. (2016). Effects of virtual human appearance fidelity on emotion contagion in affective inter-personal simulations. *IEEE transactions on visualization and computer graphics*, 22(4), 1326-1335.

Vosinakis, S., & Panayiotopoulos, T. (2001). SimHuman: A platform for real-time virtual agents with planning capabilities. In *Intelligent Virtual Agents: Third International Workshop, IVA 2001 Madrid, Spain, September 10–11, 2001 Proceedings 3* (pp. 210-223). Springer Berlin Heidelberg.

Walker-Smith, G. J., Gale, A. G., & Findlay, J. M. (2013). Eye movement strategies involved in face perception. *Perception*, 42(11), 1120-1133.

Wang, Y., Geigel, J., & Herbert, A. (2013, September). Reading personality: Avatar vs. human faces. In *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction* (pp. 479-484). IEEE.

Wang, I., & Ruiz, J. (2021). Examining the use of nonverbal communication in virtual agents. *International Journal of Human–Computer Interaction*, 37(17), 1648-1673.

Warpefelt, H., & Strååt, B. (2013). Breaking immersion by creating social unbelievability. In *Proceedings of AISB 2013 Convention. Social Coordination: Principles, Artefacts and Theories (SOCIAL. PATH)* (pp. 92-100).

Warpefelt, H. (2015, May). Cues and insinuations: Indicating affordances of non-player characters using visual indicators. In DiGRA Conference.

Warpefelt, H., & Verhagen, H. (2015). Towards an updated typology of non-player character roles. In *Proceedings of the international conference on game and entertainment technologies* (pp. 1-9).

Warpefelt, H. (2016). *The Non-Player Character: Exploring the believability of NPC presentation and behavior* (Doctoral dissertation, Department of Computer and Systems Sciences, Stockholm University).

Zibrek, K., Kokkinara, E., & McDonnell, R. (2018). The effect of realistic appearance of virtual characters in immersive environments-does the character's personality play a role?. *IEEE transactions on visualization and computer graphics*, 24(4), 1681-1690.