

Let's play, I spy...



About the book

Game companies are collecting and analysing player gameplay actions and use it to their own advantage, but, how far can they model players?

In this book we examine the games industry and how they evaluate their players, explore how games can be designed to model specific player traits, and analysed how interruptions affect player experience.

This work also describes Games User Research as a Service an in-game behaviour data telemetry platform that can serve as a stepping stone for the improved collaboration between game developers and researchers.

About the author

Carlos Pereira Santos was the technical responsible for the development of large international solutions in Critical Software for the Navy and the European Space Agency.

Currently he is a Research and Development Coordinator and Lecturer in the Breda University of Applied Sciences, where he is technical responsible for the development of multiple games and other entertainment projects involving cutting-edge research and technology like Games, Mixed Reality, and User Experience.



I spy... game analysis

Looking into how you play the game, may provide some clues about your personality.

Which symbols did you pick first and last? Why?

Did you write the results on the book?

Did you follow a consistent method?

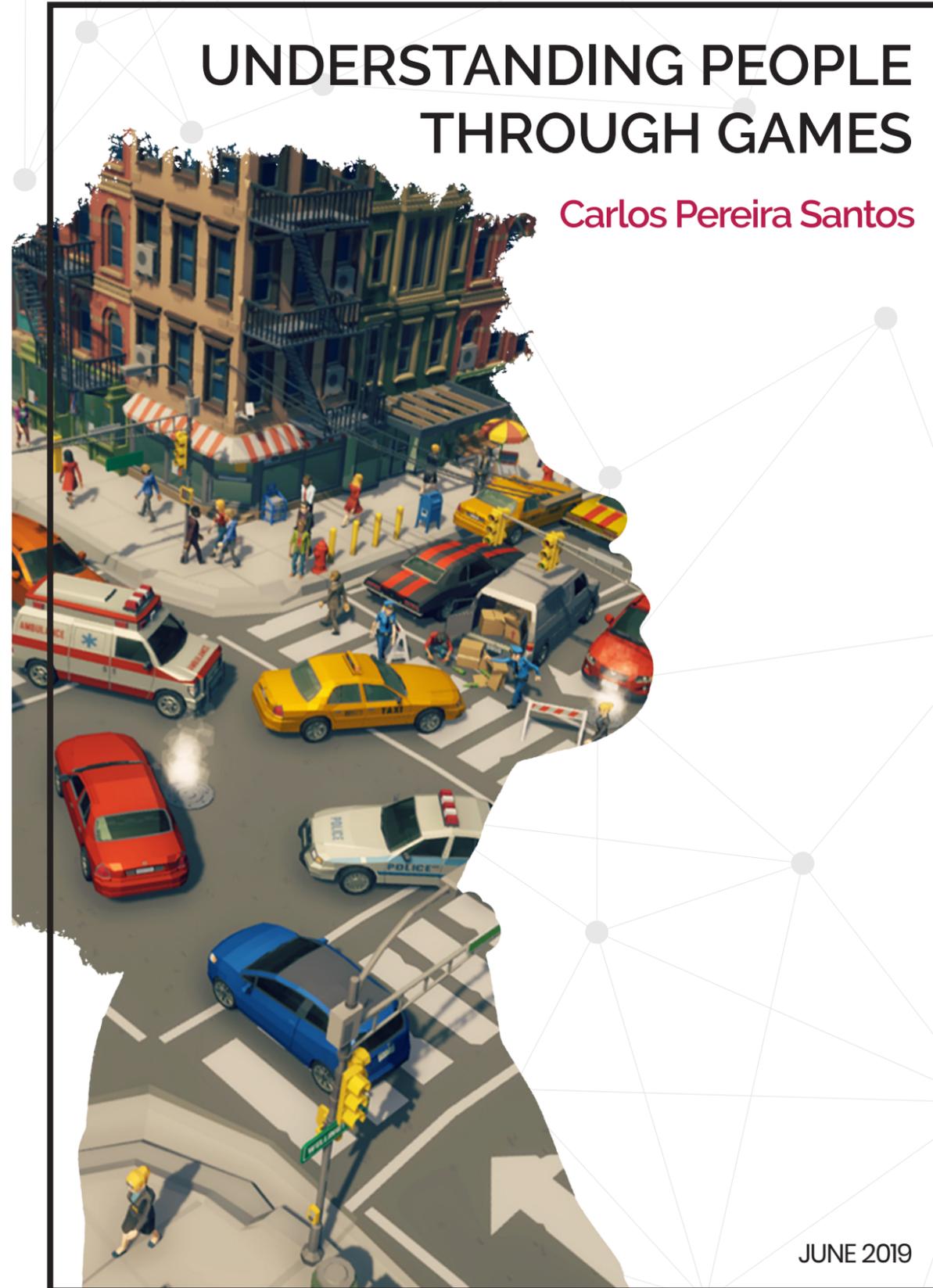
Did you compare results with others?

Did you read these questions before you play?



Carlos Pereira Santos

Understanding people through games



UNDERSTANDING PEOPLE THROUGH GAMES

Carlos Pereira Santos

JUNE 2019

UNDERSTANDING PEOPLE THROUGH GAMES

Carlos Pereira Santos

Understanding People through Games

Carlos Pereira Santos

Understanding People through Games

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Technische Universiteit Eindhoven, op gezag van de rector magnificus prof.dr.ir. F.P.T. Baaijens, voor een commissie aangewezen door het College voor Promoties, in het openbaar te verdedigen op maandag 24 juni 2019 om 13:30 uur

door

Carlos Alberto Pereira Santos

geboren te Faro, Portugal

© Carlos Pereira Santos, 2019

Eindhoven University of Technology Library
ISBN: 978-90-386-4791-3.
Nederlandstalige Uniforme Rubrieksindeling: 983

Cover by: Joey Relouw

Dit proefschrift is goedgekeurd door de promotoren en de samenstelling van de promotiecommissie is als volgt:

voorzitter: prof.dr. L. Chen
1e promotor: prof.dr. P. Markopoulos
copromotor: dr.ir. V. J. Khan
leden: prof.dr. G. N. Yannakakis (University of Malta)
prof.dr. R. C. Veltkamp (Universiteit Utrecht)
prof.dr. R. Bernhaupt
prof.dr. N. Correia (Universidade Nova de Lisboa)
prof.dr.ir. P. H. M. Spronck (Universiteit Tilburg)
prof.dr. B. A. M. Schouten

To my girls Magali, Sofia and Diana

Het onderzoek of ontwerp dat in dit proefschrift wordt beschreven is uitgevoerd in overeenstemming met de TU/e Gedragscode Wetenschapsbeoefening.

Summary

Understanding People through Games

Video games are a mainstream form of entertainment sustained by a global creative industry. They support the spread of different ideas, cultural beliefs, political and ethical stances while being consumed and entertaining billions of persons around the world. It is only in the last few years that the scientific world has come to realize how games are changing the entertainment landscape and impacting individuals, enterprise, social constructs and culture.

Collecting and analysing player in-game behaviour data is a new promising method generating high interest, by improving player experience and developing representative models of players. This work focuses on in-game behaviour directly applied to the research fields of Player Modelling and Games User Research to evaluate and improve current practices, and increase the utility value of games.

In the first controlled experiments, three different video games were designed to model different aspects of the player, each with their own challenge: Need for Cognition is a dichotomic trait (high vs. low), Self-esteem is a reflective trait and not linked to an individual's abilities, and the player's ethical alignment is multi-dimensional. Those experiments successfully demonstrated the feasibility of purposely design games to profile players, and also allowed to propose a method for designing such games.

Moreover, such studies identified the necessity of scalable solutions to support the evaluation of player experiences and the development of player models. Henceforth, we proposed the GURaaS platform designed to improve the relationship between game developers and researchers by supporting the collection of player in-game behaviour and embed research instruments in games. The two follow up studies examined benefits and drawbacks of GURaaS. In the first study, we identified current gaps in the game companies' product development cycle, and review how they could be addressed by GURaaS. In the second study, we were able to measure the impact to the experience when players were interrupted by embedded research instruments.

Overall this work contributes by demonstrating the potential of developing games to profile players, and assess the feasibility of using such games as research instruments.

Acknowledgements

First and foremost, I want to thank the aid and support of Vasillis-Javed Khan, not only you introduced me to my promotor, but also to the Industrial Design Department in the Eindhoven University of Technology. You were my role model, and kept supporting and pointing me in the right direction, selfless, relentless, and what I can proudly call a great friend. You were there with me every step of the way, my research would have been impossible without you and I am in your debt forever.

To my promotor, Panos Markopolous. I'm grateful and appreciative of the work we developed together during this long endeavour, your patience and unconditional support were crucial. I just hope you feel as proud as me of the results we achieve together, since without your time, trust, guidance and help it wouldn't have been achievable.

My sincere thanks to my PhD committee, Prof. Dr. Georgious Yannakakis, Prof. Dr. Remco Veltkamp, Prof. Dr. Regina Bernhaupt, Prof. Dr. Nuno Correia, Prof. Dr. Ir. Pieter Spronck, Prof. Dr. Ben Schouten. I could not have asked for a better world-class expert group to review my work and lift the quality of this thesis to a next level, and I hope we can work together in the near future.

I also want to thank the wonderful people I had the privilege to work together, and directly or indirectly supported this thesis: Igor Mayer, Marnix Van Gisbergen, Mata Haggis, Josephine Lappia, Bruce Hancock, Daphne Heeroma, Frank Peters, Hans Bouwknecht, Aske Plaat, Max Louwerse, Harald Warmelink, Jessika Weber, Thomas Buijtenweg, Niels van Gaans, Heleen van Geelen, Rosalinde Kennis, Joyce Ridderhof, Martin Walker, Dave Reuling, Tino van der Kraan, Elwin Verploegen, Licia Calvi, Bojan Endrovski, Serena Lucrezi, Antonio Sarà, Fábio Campos, Gareth White, Ubaldo Pantaleo, David Wessman, Chris Rothwell, Marco Palma, Andrew Woodruff, Ubaldo Pantaleo, Bram Heijligers, Josse de Kievith, Frank Bloemendal, Mathijs van Laar, Mikel Agricola, Joost Meulenkamp, Paul Mertens, Jasper Koning, Xavier Razafindrakoto, Fábio Florencio, and Ruben van Oosterhout.

A special thanks to Jerry van Heuvel, Kevin Hutchinson, Jeroen van de Haterd, Wilco Boode, Joey Relouw and Phil de Groot. Your support was essential to breathe life into the many shared projects we worked on.

I want to thank generically to the Academy for Digital Entertainment from Breda University of Applied Sciences, to the managers, professors and researchers and the work they are doing there, and for their flexibility and encouragement to pursue this venture next side to a full-time job. I am profoundly grateful to Cradle R&D Lab and all its members, a creative space where ideas flow and come to life, and laughter lightens the mood. Thank you for listening, supporting, and making my life easier as a coordinator, while doing an amazing job.

I would also like to acknowledge the thousands of anonymous participants, which played my makeshift games, or accepted to be interviewed by me, and gave me the chance to look a bit inside of their daily life's.

Looking back, there are many, many friends from Faro, FCT/UNL, EST/IPS, Skysoft/GMV that I shared good and bad moments, that supported and helped me along my past, but due the circumstances of life we scattered and not kept in touch, to all those friends I want to say that you are still in my heart.

An immense thanks goes to my parents and grandparents for their wisdom, encouragement and life lessons. You define leading by example and showed what is to prevail over hard times. My thankfulness extends, to my brother João, for being an example of an unstoppable force, to his four kids and wife, and the adopted family of my wife, you all ground me to Earth by being my family.

For lifting my spirit to a magical, loving and wonderful place, all day, every day, I am profoundly grateful to the two special little girls, that came to be during this journey. You make me keep wanting to learn, and pushing myself to be a better person, and I pray to have the wisdom to support you on discovering your own bright futures.

My final thanks, and the most important one goes to my long life caring and loving partner, my wife. Magali you are a force of nature and keep our family together, physically and spiritually, I would like to show my profound appreciation and heartfelt thanks, for being my cornerstone, for your love, and all the big and small things you did for me. This work could only be done because of you.

Contents

Chapter 1: Introduction	1
Background and motivation.....	2
User Models	3
Player Profiling.....	4
Player Experience	6
Methodology and Approach	8
Contents of the Thesis	9
Study 1: Need for Cognition.....	9
Study 2: Self-Esteem.....	9
Study 3: Ethical Norms.....	9
GURaaS Platform Design.....	9
Study 4: Player Experience Evaluation Industry Practices.....	10
Study 5: Player Experience Interruption.....	10
Summary.....	11
References.....	11
Chapter 2: Profiling Personality Traits with Games	17
Abstract	17
Introduction.....	18
Measuring Personality Traits	18
Player Profiling.....	19
Games as a Personality Profiling Tool.....	21
Modelling Player's Need for Cognition: Study 1.....	22
Design of a Game to Model NfC.....	23
User Study.....	25
Results.....	27
Discussion.....	31
Modelling player's Self-esteem: Study 2.....	32
Motivation.....	32
Study Research Plan	34
Game Design Elements.....	35
Research Hypotheses	39

Hypothesis Review.....	46
Discussion.....	48
Systematic Method to Profile Player Traits.....	49
Player Profiles in the Games Industry	50
Player Profiles For Serious Applications.....	51
Directions	52
Conclusion.....	53
References.....	54
Chapter 3: Profiling Ethics Orientation through Play	61
Abstract	61
Introduction.....	62
Games as profiling tools	62
Multi-Dimensional Ethics Scale.....	63
Ethical Norms.....	64
Games and Ethics.....	67
Method	67
Branching Narrative.....	68
Participants.....	69
Process.....	70
Results.....	71
Raw scores	71
Machine Learning.....	72
Validation with a new data set	72
Discussion.....	73
Profiling moral judgment through play.....	73
Designing effective Choice/Actions identifiers.....	74
Designing Ethical Games.....	75
Opportunities	75
Conclusions.....	76
References.....	76

Chapter 4: GURaaS: An End-User Platform for Embedding Research Instruments into Games **81**

Abstract	81
Introduction.....	82
GURaaS Design.....	83
Players.....	84
Game Developers.....	87
Researcher	88
Architecture	89
Game User Research Kit.....	90
Game User Research Server	91
Game User Research Portal.....	92
Case Study.....	93
Conclusion.....	96
References.....	96

Chapter 5: Player Experience Evaluation in the Games Industry **101**

Abstract	101
Introduction.....	102
Background.....	102
Player Experience Evaluation	103
Game User Research-Focused Tools	103
Method	104
Participant Pool.....	104
Ethical Considerations	104
Interview Procedure	105
Games User Research as a Service.....	109
Data Analysis Procedure.....	109
Results.....	110
Referenced Methodologies	111
Tools for Player Experience Evaluation.....	114
Attitudes Towards the GURaaS Platform.....	115
Data Sharing and Player Privacy	116

Discussion.....	118
Player Experience Evaluation Attitudes.....	118
Exploring New Methodologies	119
Attitudes Towards the GURaaS Platform.....	119
Sharing of Player In-Game Data	120
Conclusion.....	120
Acknowledgment.....	120
References.....	121

Chapter 6: Effects of interruptions on the player experience **125**

Abstract	125
Introduction.....	126
Background.....	127
Relative vs. Absolute Ordinal Measurements.....	128
Measuring Game Experience.....	128
Interrupting an Experience.....	129
Ad Intrusiveness.....	130
Method	131
Experimental Conditions.....	132
Participants.....	135
Results.....	136
Game Performance.....	136
Game Core Experience	136
Game Post Experience	138
Perceived Intrusiveness.....	138
Reliability of the Demographic Questionnaire	140
Discussion.....	142
Detailed Evaluation of the Game Experience	142
Reliability of In-Game Questionnaires.....	143
Implications.....	144
Limitations and Opportunities.....	144
Conclusions.....	145
Acknowledgment.....	145
References.....	146

Chapter 7: Conclusion **151**

Profiling the Need for Cognition Case Study.....	151
Profiling the Self-Esteem Case Study.....	152
Profiling the Ethical Norms Case Study.....	152
Game User Research as a Service Platform	153
Player Experience Evaluation within the Games Industry.....	153
Player Experience Effects of Interruptions.....	154
Reflections on Methodology.....	155
Revision of Research Goals.....	156
Design games to infer specific player traits	156
Player experience evaluation within industry	156
Player-focused research through games tool.....	157
Impacts to the games industry.....	157
Future Challenges.....	158
Game Design.....	158
Ethical Grounds.....	159
GURaaS Research.....	160
Final Statements of the Thesis.....	160
Design profiling games.....	160
Games User Research as a Service.....	161
Desirability.....	161
References.....	163

Appendix A: Portuguese Summary **167**

Appendix B: List of Publications **169**

Thesis Related Publications.....	169
Other Publications.....	170

Chapter 1

Introduction

Video games have had a profound impact on the world's culture.

It is undeniable that video game consumption is prevalent on a global scale, spreading different ideas, cultural beliefs, political and ethical stances while being consumed by billions [13, 32, 42, 52].

The number of players is growing; as such, is not surprising that video games represent the fastest growing segment of the entertainment industry, with a total of 2.6 billion gamers and \$121.7 billion USD spent in 2017. Moreover, the demographics of players are also changing; not only are they evenly distributed across several age groups, but they have also become more gender neutral (45% of the US gamers are female). Furthermore, 55% of gamers say that video games help connect them with their friends [11].

Video games are defining their own culture [3, 40] and influencing existing art, for instance, by pushing new graphical designs and art styles from pixel and voxel art to digital 3D animations or by influencing musicians, whereby orchestras are composing and playing music from games [14]. People are not only just playing games, they also are finding entertaining to watch others play games and are willing to pay to see them do so. Being a professional video-game player is a real job, and eSports [5, 15] are an international phenomenon, attracting thousands to watch players perform/play live.

Games are also expanding their reach. They are becoming essential in the social construct of parenthood: 70% of parents believe video games have a positive influence on their children's lives, and 46% state that video games help their family spend time together [11]. There are games that have become well known because they contribute to education or pursue socially-desirable aims such as community development, sustainability or even the development of scientific knowledge. Recently, serious games have captivated players and contributed actively to specific goals. There are many examples of such games, including "Play to Cure: Genes in Space", which looks into analysing genetic data for cancer research [17] and yet another game that concentrates on the identification of malaria parasites [24]. The main advantage of these games is that through

careful design of the game elements and clear instructions, they keep players entertained and engaged while contributing to a research goal and quickly and cheaply crowd source specialized tasks which would otherwise be performed by trained professionals.

There are also many examples of serious games assisting stakeholders from government agencies and large industry sectors in important strategic decisions, e.g., for exploring participatory citizenship and city planning [36] or international cooperation and stakeholder engagement for maritime spatial planning [26].

Definition

Within the scope of this thesis, serious games, also known as applied games or games with a purpose, are games with enjoyable and/or engaging elements, but their primary goal is not related to entertainment but instead to, for example, educate, train, test and/or contribute directly to society [2, 27, 43].

Background and motivation

The history of the video game industry is relatively short, with the first widely publicly available video game appearing less than 50 years ago, and, although it is an exciting, creative new business area full of profitable opportunities, it is also prone to failure and far from stable. At the moment, there are several large shifts in the industry: large emerging markets in Asia, a general shift to mobile games, uncertainty regarding the role of augmented and virtual reality, the presence of new, online digital stores focusing on downloadable content (reducing the role of the traditional publishers), different independent financing schemes (e.g., crowdsourcing, angel investors), free-to-play games and micro-transactions, and an ever growing availability of online resources, including knowledge, assets, freelancers and tools.

It is my belief that the video-game landscape is far from defined and only now are we realizing how games are changing the entertainment landscape and impacting individuals, social constructs (families, friendship, and so forth), industry and culture. There is still a lot of exploration and research to be done at a social, cultural, didactic and psychological levels regarding how video games are impacting society. To be able to conduct such research, we need to better understand the dynamics between the player and the games. Clarifying those dynamics is my main motivation; specifically, I want to understand how games can be used as an instrument to study players.

Although there are (serious) games capable of aligning entertainment with specific research goals, they are focused mainly on task-oriented outcomes, i.e., they contribute through engaging players in performing impactful activities, thus falling more into the field of computer-supported cooperative work and social

computing. But, what if we want to design a game that focuses on studying the player? How can games be used to profile players' characters, traits, experiences and social behaviour?

User Models

In today's media landscape, there are large number of online consumer products, ranging from search engines, social networks and streaming services to online shopping Recommender Systems which are developing a profound research and knowledge base around User Models [18, 34]. The overall objective of a Recommender System is to identify items (e.g., products, websites, applications, movies, songs, and so on) for which users have a high affinity, i.e., to find items that the users are interested in. Since they are able to generate rich user profiles from consumer behaviour [16, 29, 46], services are actively using these systems to provide better experiences and generate more profit.

There are clear differences between such services and games; in general, when playing a video game, the involvement and frequency of player interactions are much higher, and every interaction within a video game can be logged and analysed. Telemetry and analytics of game metrics have become common practice in the industry, with large game companies tracking players actively with the intent to improve game development, content and revenue [10].

Definition

Within the scope of this thesis, Game Metrics are interpretable and collectable measurements of player interactions within the context of a game [50]. Generally, game metrics include player usage information (e.g., acquisition, retention, location, platform, playing time, frequency or expenditure), and in-game player behaviour (e.g., success rate, frequency or sequence of actions, item selection, accuracy, skill set).

Game companies are using such metrics to support the product development process, namely, in the areas of game authoring and content creation, by allowing designers to understand and perfect the player experience. In addition, they use it to drive business analyses, such as sales figures or the expected number of players. Multiple researchers are having an important impact on the industry through the development of specific methodologies with which to measure, represent and interpret different game metrics [10].

It is my conviction that evaluating in-game behaviour has great potential and is in line with my goal and motivation. My work in this research field contributes directly to existing literature and can help clarify the potential of using game metrics to understand and model intrinsic characteristics of players, namely, by using Player Modelling.

Definition

Player Modelling [1, 50, 54] is a new research field under the general umbrella of Computational Intelligence which is applied directly to games and explores the detection, modeling, prediction and expression of human player characteristics which are manifested through cognitive, affective and behavioral patterns.

There are multiple and great examples of player models that collect and analyse players' behavior to develop representative valuable models. Some examples can showcase how different, and important they are becoming to the industry.

For example, improve significantly the effectiveness of the game opponent AI on a complex RTS game modelling opponents [58]. Another example supports game designers by identifying and constructing models of player types. This allows to support the automation of testing procedures and evaluate if the intended game design matches the intended design [59]. Yet another good example of using video footage of human and AI player of a platformer to study the different elements of the level design and play style which may contribute to the believability of an agent and can support future designers to create better levels [60].

Player Profiling

Past research has already proven that it is possible to develop accurate user profiles based on in-game behaviour. Multiple publications have documented the accurate identification of specific player traits, such as age [44], skill [47, 49, 53], experience [31, 39], the big five personality traits (i.e. openness, conscientiousness, extraversion, agreeableness and neuroticism) [21, 22, 51] or HEXACO (i.e. honesty, emotionality, extraversion, agreeableness, conscientiousness and openness) [48].

Collectively, what these studies have shown is that although videogames create alternative realities which encourage a player to explore without any implications in real life, a player's individuality and personality still emerges in their in-game behaviour. In short, it is possible through data analysis to derive a player's real nature from their in-game behaviour.

Despite the extensive and successful work developed in the field, I noticed a missing link. All of the above-mentioned case studies have already shown that games can be used to infer and even predict specific player traits, but, is it possible to purposely design a game to do that? That led me to define my first research question.

Research Question 1

To what extent is it possible to purposely design a game which would be able to infer or predict a specific player trait?

In my opinion, the contribution to the field of answering such a question would be an increase in the utility value of games [12], i.e., allow game developers to design personalized game experiences based on players' specific traits. For example, the player experience could be improved by procedurally generated levels [23] focused mainly on exploration for players that avoid harmful situations [8].

Procedural content generation [55] is software through which (media) content is created/generated using parametric algorithms. Some game companies are using procedural content generation to aid their development pipeline to create better, bigger or more believable content, and even, embed such algorithms into their games exposing directly their players' content generation without any further design input.

Nowadays, it is possible to develop games with procedural content generation features which contemplate parameters linked with specific player models, like player preferences and traits. This would allow to adapt and tailor the game elements and experiences to the players' needs and interests. It is my belief that Experience-Driven Procedural Content Generation [56] will become increasingly more important for games to close the affective loop [57].

Furthermore, the utility could also be increased by allowing serious games to be purposely designed to profile players for other fields, including Psychology, Human Resources, Medical Diagnostics, or Human-Computer Interaction (HCI). An example would be to develop a game able to monitor regularly the self-esteem [35] of players with signs of depression.

This line of exploration led me to wonder if I could augment the link between games and research even further by considering what it would take for entertainment games to be used as a support mean to conduct player-focused research.

If games would allow researchers to target specific target audiences and embed research instruments, then researchers could not only understand games and experiences better, but this activity would also increase gaming's contribution to society.

Exploring such areas is obviously a complex topic, and there are multiple cases of such undertakings that would satisfy this premise. One, companies are using their own games to conduct research on the improving player experience or to define marketing strategies. Two, entertainment games have and are being used in research by many researchers that use known commercial games to conduct their own experiments [9, 30, 41, 45]. Nevertheless, the concept goes beyond that by allowing any interested third-party researchers to embed their research instruments directly within existing games.

Considering this possibility leads to numerous other possible research questions: the technical feasibility of such concept, influences on the player experience, player perceptions, industry acceptance, social responsibility, research validity, ethical

and player privacy perspectives, just to name a few topics. Player experience was elected as a conducting line with which to evaluate how entertainment-focused games can be used as a player research platform.

Player Experience

Concerned with the effects of embedding research instruments into games on the player experience, I looked directly into the HCI side of games. This led me to the parallel research field of User Experience in Games and thoughts as to how it could be linked with Player Modelling.

Definition

Within the scope of this thesis, User Experience in Games, or Games User Research (GUR) is a research field concerned with any aspect that may influence the experience and perception of video games [61]. Which includes the development of tools and techniques to measure and evaluate players' perceptions and experiences [4, 28, 37]

By definition, Player Modelling and User Experience in Games are closely related research fields, and it is possible to find multiple studies that cross over between them [30, 39, 49]. Both focus on gathering game metrics and deriving representative player models. The main differences are:

- User Experience in Games focuses mainly on the HCI side of games, while Player Modelling representations may be used for other purposes, such as driving Artificial Intelligence (AI) agents or generating procedural content.
- Despite the similarities of the tools and methodologies of both fields, the goal of Player Modelling is to construct representative player models of affect, cognition and behaviour, while Player Experience goal is to measure and understand the effects of the interaction between players and games.

GUR goes beyond playtesting game it directly looks into physiological measures for game evaluation, allowing designers to measure, analyse and evaluate if the game is providing the intended experience [62]. With the advancement of research in the areas of recommender system, data analytics, biometrical sensors, and mixes-methods are thriving in the new domain of GUR.

The mixed methods [63] are especially relevant to validate approaches, for example the collection of specific game metrics, with self-reported questionnaires [64-66], or a multiple biometrical sensing like heart rate, electro-dermic, electroencephalogram activity [67-69]. Finding those correlations are extremely important, allowing to create automatic metrics from game metrics. Using data analytics information is allowing companies to develop better games, closer to the players' and responding directly to their needs [70].

Within the scope of this thesis, I am particularly interested in looking directly into in-game behaviour from the vantage points of both fields, GUR and Player

Modeling. In an attempt to explore the commonalities in the current practices, methodologies and tools of the two fields, I was faced with a surprising lack of information on the topic. Although game companies seem actively collecting millions of game metrics, they are focusing mainly on usage data and less on player in-game behaviour. In short, despite companies having a clear purpose, their processes and tools for player experience evaluation are ill-defined.

Research Question 2

What is the current role of player experience evaluation within game companies?

There are different off-the-shelf software packages which can be used to log, survey and annotate different data sets, and the overhead of maintaining and combining several data pools is creating unnecessary entropy; however, industry opinion suggests that the existing tools are weak and do not support the work of game designers or researchers [6, 10]. If we are able to evaluate the current role of player experience evaluation, then we will be able to better understand the limitations of the existing software and support game companies and educational institutes in defining common processes and tools related to player experience evaluation. This will strengthen and mature methodologies, which will then be reflected in product quality and production costs, and give a focused direction to improving off-the-shelf software for player evaluation purposes.

To validate my research, I went through the process of collecting and evaluating players' in-game behaviours, and this required a robust definition of a system which supports data collection and the analysis of player in-game behaviour. This process required both academic analysis of existing systems and a practical implementation of the proposed improvements to evaluate their effectiveness. This led me to formulate the following question:

Research Question 3:

How can we operationalize a tool which would allow games to support player-focused research?

Answering this question will support directly both to academic and industry professionals in three distinct areas. First, the definition of an architecture and design principles of a tool which would be able to collect and evaluate in-game behaviour. Secondly, doing so would also allow us to evaluate the technical feasibility of such a platform and tackle other problems like how can game companies expose game data safely to researchers, or how would researchers be able to collect large amounts of data.

Lastly, it would demonstrate that supporting player-focused research through games is not only possible but also a line of research worth exploring, with the potential for developers and researchers to gain better understandings of the users as well as their links to games and the entertainment industry, in general.

Methodology and Approach

To answer the aforementioned research questions, my methodology is based on the MacKay and Fayard HCI Triangulation Framework [25]. The framework proposes a combination of science and design to support the study of interactions between participants and developed artefacts. Specifically, I approached the process by defining a set of design, empirical, theoretical activities, as illustrated in Figure 1-1.

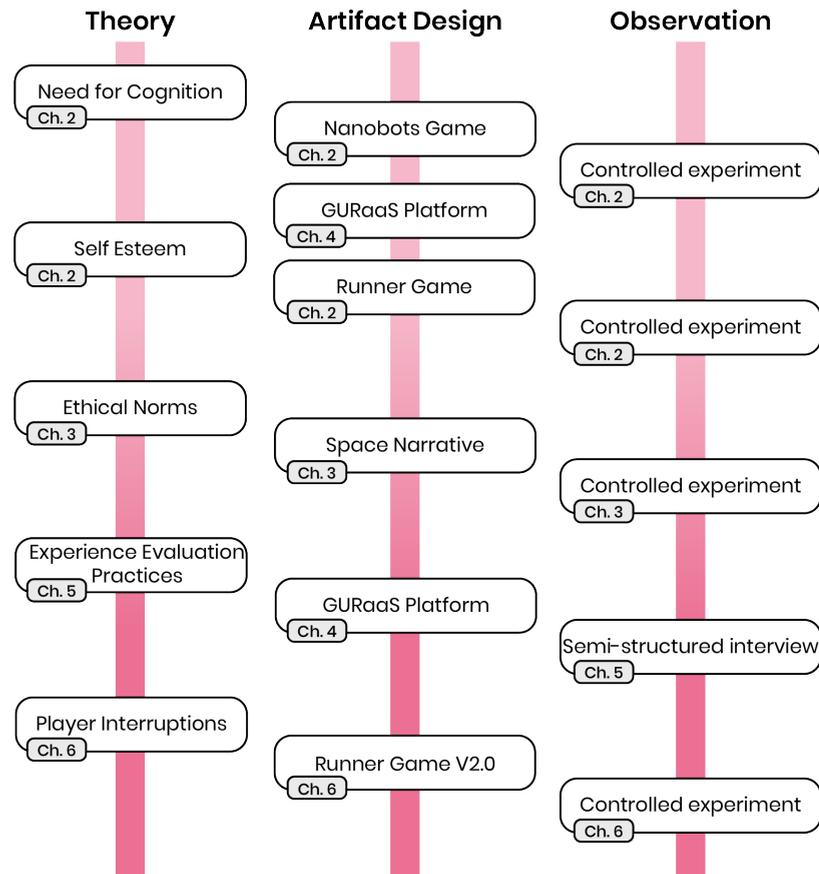


Figure 1-1: Overview of the triangulation framework activities of the thesis.

My research relied heavily on a consistent theoretical work from Psychology and Artificial Intelligence because it relates directly to modelling human traits. However, when it came to investigating player experience evaluation, I crossed into the field of HCI.

It is my expectation that my body of work contributes to multiple research fields by creating a bridge between them while increasing the value of games, supporting the industry by perfecting and maturing their processes, and, finally, improving games by enhancing the player experience.

Contents of the Thesis

This work is organized as a collection of peer reviewed articles, most of them published or accepted with the exception of the chapter 6, which was submitted but on the time of printing this book, the conference review was still out.

The collection reflects the work that I developed over time, and in this section details in more depth the different studies that were developed, and provide a conducting line and the link between the different presented chapters.

Study 1: Need for Cognition

In this study, I approach the first research question by conducting a controlled experiment in which I requested that a group of participants play Nanobots, a purposely-designed game capable of profiling the Need for Cognition (NfC) [7].

NfC is a simple, linear and stable psychological trait linked with reasoning. Due to its characteristics and that there has been no previous attempt to model it with games, NfC was an ideal trait to start with. The case study methodology, process and results can be found in Chapter 2.

Study 2: Self-Esteem

Also in Chapter 2, a second controlled experiment is described, which was conducted in a manner similar to that of the previous study but focused on Self-Esteem (SE) [35]. SE is a well-known and clear candidate for serious applications within the medical field. For this purpose, we designed and developed a new platform game – Runner.

Study 3: Ethical Norms

Since both NfC and SE are linearly-scaled psychological traits, I wanted to expand to a multidimensional trait. Further searching led me to Normative Ethics [33] and conducting a third controlled experiment still related to the first Research Question.

Chapter 3 documents comprehensively how a space-branching narrative was designed, developed and analysed using machine learning to be able to profile players' ethical alignments in multiple dimensions, namely, Contractualism, Utilitarianism and Moral Equity.

GURaaS Platform Design

During the development of the second controlled experiment, I noticed the need to separate the data collection and analysis system from the games being developed. Driven by the third Research Question, and lacking a proper off-the-shelf tool which satisfied my needs, I applied the design science method to

develop the Games User Research as a Service (GURaaS) platform, documented in detail in Chapter 4.

In that chapter, the major technical components to support data retrieval from games and player in-game behaviour on a large scale is documented. In addition, it introduces a novel idea of providing control to the internal or external researchers in order that they might focus on specific target audiences and customize research instruments directly within games.

The GURaaS platform was then used in multiple follow up studies, namely Studies 3, 4 and 5.

Study 4: Player Experience Evaluation Industry Practices

The second Research Question is tackled directly in Chapter 5, where the results of a semi-structured interview with game industry professionals from 11 different companies are presented. In an attempt to reach a global group, we interviewed companies from around the globe while targeting different game platforms (ranging from arcade virtual reality to mobile-based games) in different sizes, i.e., from small and independent developers to large multi-national companies.

The interviews had two distinct goals: first, to understand the current industry practices with regards to player experience evaluation, in terms of the tools and methods used. We explored deeply not only the results obtained but also what the motivations and limitations of their player experience evaluation practices were, including laws and privacy policies. Second, in the second half of the interview, the core concepts of GURaaS (detailed in Chapter 4) were introduced to the participants and then the revision of the features and concepts were discussed in order to understand the applicability of such a service in the game industry.

Study 5: Player Experience Interruption

From the several relevant outcomes obtained from Study 4 (Company Interviews), there was one that generated some buzz, namely the fact that researchers could introduce questionnaires in key moments of the game to players. Multiple participants questioned the impact of introducing questionnaires into the player experience.

Noticing the lack of documented information on how advertisements and questionnaire-based interruptions impact the player experience, I conducted a controlled experiment to measure the impact of the different types of interruptions. This process is described in Chapter 6.

To achieve my goal for this study, I repurposed the Runner platform developed for study 2 and measured the differences between three different conditions: no interruptions, advertisement interruptions and questionnaire-based interruptions.

Summary

In this thesis, I present a set of linked studies related to in-game behaviour and tie the research fields of Player Modelling and GUR together to answer the following questions:

- To what extent is it possible to purposely design a game which would be able to infer or predict a specific player trait?
- What is the current role of player experience evaluation within game companies?
- How can we operationalize a tool which would allow games to support player-focused research?

In the last chapter, we highlight the most important findings from each of the previous chapters as well as summarily answering the research questions, and discuss their limitations, and contributions.

References

- [1] Sander C.J. Bakkes, Pieter H.M. Spronck, and Giel van Lankveld. 2012. Player behavioural modelling for video games. *Entertainment Computing* 3, 3: 71–79.
- [2] Francesco Bellotti, Bill Kapralos, Kiju Lee, Pablo Moreno-Ger, and Riccardo Berta. 2013. Assessment in and of serious games: An overview. *Advances in Human-Computer Interaction* 2013.
- [3] Arthur Asa Berger. 2001. *Video Games: A Popular Culture Phenomenon*. Routledge, New York.
- [4] Regina Bernhaupt, Wijand Ijsselsteijn, Florian “Floyd” Mueller, Manfred Tscheligi, and Dennis Wixon. 2008. Evaluating user experiences in games. In *Proceeding of the twenty-sixth annual CHI conference extended abstracts on Human factors in computing systems - CHI '08*.
- [5] Bryce C Blum and Stephen D Fisher. 2014. THE RISE OF eSPORTS League of Legends Article Series. Foster Pepper.
- [6] Emily Brown. 2010. The Life and Tools of a Games Designer. In *Evaluating User Experience in Games*, Regina Bernhaupt (ed.). Springer, London, 73–87.
- [7] J T Cacioppo, R E Petty, and C F Kao. 1984. The efficient assessment of need for cognition. *Journal of personality assessment* 48, 3: 306–307.
- [8] C. Robert Cloninger, Dragan M. Svrakic, and Thomas R. Przybeck. 1993. A Psychological Model of Temperament and Character. *Archives of General Psychiatry*.
- [9] Ben Cowley. 2009. *Player Profiling and Modelling in Computer and Video Games*. University of Ulster, Coleraine.

- [10] Magy Seif El-Nasr, Anders Drachen, and Alessandro Canossa. 2013. Game Analytics: Maximizing the Value of Player Data.
- [11] ESA. 2018. Essential Facts about the Computer and Video Game Industry 2018. Entertainment Software Association 1. Retrieved 01/19/2019 from http://www.theesa.com/wp-content/uploads/2018/05/EF2018_FINAL.pdf
- [12] Peter C. Fishburn. 1970. Utility Theory for Decision Making. Research Analysis Corporation.
- [13] James Paul Gee. 2008. Video games and embodiment. *Games and Culture*.
- [14] Mata Haggis, Carlo Perrotta, Donatella Persico, Chris Bailey, Jeffrey Earp, Francesca Dagnino, Marcello Passarelli, Flavio Manganello, Francesca Pozzi, and Thomas Buijtenweg. 2018. A Manifesto for European Video Games. Rome.
- [15] Juho Hamari and Max Sjöblom. 2017. What is eSports and why do people watch it? *Internet Research*.
- [16] Rong Hu. 2010. Design and user issues in personality-based recommender systems. In *Proceedings of the fourth ACM conference on Recommender systems - RecSys '10*.
- [17] Nachiket Huilgol. 2014. Play to cure: Genes in space. *Journal of Cancer Research and Therapeutics* 10, 3: 792.
- [18] D Jannach, M Zanker, A Felfernig, and G Friedrich. 2010. *Recommender Systems: An Introduction*. Cambridge University Press.
- [19] F. Khatib, S. Cooper, M. D. Tyka, K. Xu, I. Makedon, Z. Popovic, D. Baker, and F. Players. 2011. From the Cover: Algorithm discovery by protein folding game players. *Proceedings of the National Academy of Sciences* 108, 18949–18953.
- [20] Firas Khatib, Frank DiMaio, Seth Cooper, Maciej Kazmierczyk, Mirosław Gilski, Szymon Krzywda, Helena Zabranska, Iva Pichova, James Thompson, Zoran Popović, Mariusz Jaskolski, and David Baker. 2012. Crystal structure of a monomeric retroviral protease solved by protein folding game players. *Nature Structural & Molecular Biology* 19, 364–364.
- [21] Giel Van Lankveld, Sonny Schreurs, and Pieter Spronck. 2009. Psychologically verified player modelling. In *GAMEON*, 12–19.
- [22] Giel van Lankveld, Pieter Spronck, Jaap van den Herik, and Arnoud Arntz. 2011. Games as personality profiling tools. *2011 IEEE Conference on Computational Intelligence and Games (CIG'11)*: 197–202.
- [23] Roland Van Der Linden, Ricardo Lopes, and Rafael Bidarra. 2014. Procedural generation of dungeons. *IEEE Transactions on Computational Intelligence and AI in Games*.
- [24] Miguel Angel Luengo-Oroz, Asier Arranz, and John Freen. 2012. Crowdsourcing malaria parasite quantification: an online game for analyzing images of infected thick blood smears. *Journal of medical Internet research* 14, 6: e167.
- [25] Wendy E. Mackay and Anne-Laure Fayard. 1997. HCI, Natural Science and Design: a Framework for Triangulation Across Disciplines. In *Proceedings of the conference on Designing interactive systems processes, practices, methods, and techniques - ACM DIS '97*, 223–234.

- [26] Igor Stefan Mayer, Qiqi Zhou, Lodewijk Abspoel, and Xander Keijser. 2014. Gaming the Future of the Ocean: The Marine Spatial Planning Challenge 2050. *SGDA 2014, Serious Games Development and Applications, LNCS 8778*: 150–162.
- [27] Igor Mayer, Harald Warmel'ink, and Qiqi Zhou. 2016. A frame-reflective discourse analysis of serious games. *British Journal of Educational Technology*.
- [28] Graham McAllister, Pejman Mirza-Babaei, and Jason Avent. 2013. Improving Gameplay with Game Metrics and Player Metrics. In *Game Analytics*. 621–638.
- [29] Alvaro Ortigosa, Rosa M. Carro, and José Ignacio Quiroga. 2014. Predicting user personality by mining social interactions in Facebook. In *Journal of Computer and System Sciences*.
- [30] C. Pedersen, J. Togelius, and G.N. Yannakakis. 2010. Modeling Player Experience for Content Creation. *IEEE Transactions on Computational Intelligence and AI in Games* 2, 1.
- [31] Chris Pedersen, Julian Togelius, and Georgios N. Yannakakis. 2009. Modeling player experience in Super Mario Bros. In *CIG2009 - 2009 IEEE Symposium on Computational Intelligence and Games*.
- [32] KA Pepler and YB Kafai. 2007. What Videogame Making Can Teach Us about Literacy and Learning: Alternative Pathways into Participatory Culture. In *Situated Play: Proceedings of the Third International Conference of the Digital Games Research Association (DiGRA)*.
- [33] R. E. Reidenbach and D. P. Robin. 1990. Toward the development of a multidimensional scale for improving evaluations of Business Ethics. *Journal of Business Ethics* 9, 8: 639–653.
- [34] Francesco Ricci, Lior Rokach, Bracha Shapira, and Paul B. Kantor. 2010. Introduction to Recommender Systems Handbook. In *CEUR Workshop Proceedings*.
- [35] M. Rosenberg. 1965. Rosenberg self-esteem scale (RSE). *Acceptance and commitment therapy., Measures package* 61: 52.
- [36] Ben Schouten. 2016. Playful empowerment, the role of game design innovation in participatory citizenship. *Lecture Notes in Computer Science Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics*.
- [37] Magy Seif El-Nasr, Heather Desurvire, Lennart Nacke, Anders Drachen, Licia Calvi, Katherine Isbister, and Regina Bernhaupt. 2012. Game user research. In *Proceedings of the 2012 ACM annual conference extended abstracts on Human Factors in Computing Systems Extended Abstracts - CHI EA '12*, 2679.
- [38] Sebastian Seung. 2013. Eyewire. Eyewire.
- [39] Noor Shaker, Stylianos Asteriadis, Georgios N. Yannakakis, and Kostas Karpouzis. 2013. Fusing visual and behavioral cues for modeling user experience in games. *IEEE Transactions on Cybernetics*.
- [40] Adrienne Shaw. 2010. What is video game culture? *Cultural studies and game studies. Games and Culture*.

- [41] Pieter Spronck, Iris Balemans, and Giel Van Lankveld. 2012. Player Profiling with Fallout 3. In *AIIDE*.
- [42] Jenny Sundén and Malin Sveningsson. 2012. Gender and sexuality in online game cultures: Passionate play.
- [43] Tarja Susi, Mikael Johannesson, and Per Backlund. 2007. Serious games—An overview. Skövde University of Skövde Technical Report HSIKITR07001: 28.
- [44] Shoshannah Tekofsky, Pieter Spronck, Aske Plaat, Jaap van den Herik, and Jan Broersen. 2013. Play style: Showing your age. In *2013 IEEE Conference on Computational Intelligence in Games (CIG)*, 1–8.
- [45] Christian Thureau and Christian Bauckhage. 2010. Analyzing the Evolution of Social Groups in World of Warcraft®. *Proceedings of the 2010 IEEE Conference on Computational Intelligence and Games*: 170–177.
- [46] Marko Tkalčič, Matevž Kunaver, Jurij Tasič, and Andrej Košir. 2009. Personality Based User Similarity Measure for a Collaborative Recommender System. *5th Workshop on Emotion in Human-Computer Interaction Real World Challenges*.
- [47] Julian Togelius, R De Nardi, and Sm Lucas. 2006. Making racing fun through player modeling and track evolution. *Optimizing Player Satisfaction in Computer and Physical Games*: 61.
- [48] Narnia C. Worth and Angela S. Book. 2014. Personality and behavior in a massively multiplayer online role-playing game. *Computers in Human Behavior* 38: 322–330.
- [49] Georgios N. Yannakakis and John Hallam. 2009. Real-time game adaptation for optimizing player satisfaction. *IEEE Transactions on Computational Intelligence and AI in Games*.
- [50] Georgios N. Yannakakis, Pieter Spronck, Daniele Loiacono, and Elisabeth André. 2013. Player Modeling. *Dagstuhl Follow-Ups* 6, 59.
- [51] Nick Yee, Nicolas Ducheneaut, Les Nelson, and Peter Likarish. 2011. Introverted Elves & Conscientious Gnomes: The Expression of Personality in World of Warcraft. *Proceedings of the Annual Conference on Human Factors in Computing Systems - CHI '11*: 753–762.
- [52] José P Zagal. 2009. Ethically Notable Videogames: Moral Dilemmas and Gameplay. *Proceedings of DiGRA 2009*: 9.
- [53] A E Zook and M O Riedl. 2012. A temporal data-driven player model for dynamic difficulty adjustment. In *Proceedings of the 8th AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment, AIIDE 2012*, 93–98.
- [54] Georgios N. Yannakakis, and Julian Togelius. 2018. *Artificial Intelligence and Games*, Springer, ISBN: 978-3-319-63519-4
- [55] Shaker, Noor, Julian Togelius, and Mark J. Nelson. *Procedural content generation in games*. Switzerland: Springer International Publishing, 2016.
- [56] Georgios N. Yannakakis and Julian Togelius. 2011. Experience-driven procedural content generation. *IEEE Transactions on Affective Computing* 2.3 (2011): 147–161.

- [57] Petra Sundström. 2006. Exploring the affective loop, Stockholm University, Technical Report.
- [58] Sander Bakkes, Pieter Spronck, and Jaap van den Herik. 2009. Opponent modeling for case-based adaptive game ai. *Entertainment Computing*, 1(1):27–37.
- [59] Anders Drachen, Alessandro Canossa, and Georgios N. Yannakakis. 2009. Player modeling using self-organization in Tomb Raider: Underworld. *IEEE Conference on Computational Intelligence and Games (CIG)*
- [60] Elizabeth Camilleri, Georgios N. Yannakakis, and Alexiei Dingli. 2016. Platformer level design for player believability. *2016 IEEE Conference on Computational Intelligence and Games (CIG)*.
- [61] Anders Drachen, Pejman Mirza-Babaei, Lennart Nacke. 2018. *Games User Research*. Oxford University Press
- [62] Lenart Nacke. 2015. *Games User Research and Physiological Game Evaluation*. Regina Bernhaupt Game User Experience Evaluation. Human-Computer Interaction Series. Springer
- [63] E. Calvillo-Gámez, P. Cairns, and A. Cox. 2015. Assessing the Core Elements of the Gaming Experience. *Game User Experience Evaluation*, Springer, pp. 37–62.
- [64] H. Desurvire and C. Wiberg. 2009. Game usability heuristics (PLAY) for evaluating and designing better games. *Lecture Notes in Bioinformatics*, 2009.
- [65] J. H. Brockmyer, C. M. Fox, K. A. Curtiss, E. Mcbroom, K. M. Burkhart, and J. N. Pidruzny, “Journal of Experimental Social Psychology The development of the Game Engagement Questionnaire : A measure of engagement in video game-playing,” *J. Exp. Soc. Psychol.*, 2009.
- [66] W. a. IJsselsteijn, Y. a. W. de Kort, and K. Poels, “Game Experience Questionnaire,” *FUGA - Fun Gaming*, 2013.
- [67] Cátia Mendes; Débora Vieira; Ana da Costa. 2018. Heart Rate, Electrodermic Activity and Player’s Experience with Virtual Games. *PsychTech & Health Journal*, p. 43-52, oct. 2018. ISSN 2184-1004
- [68] A. Plotnikov, N. Stakheika, A. de Gloria, C. Schatten, F. Bellotti, R. Berta, C. Fiorini, and F. Ansovini. 2012. Exploiting real-time EEG analysis for assessing flow in games. *IEEE 12th International Conference on Advanced Learning Technologies*, pp. 688-689. IEEE.
- [69] Stefano Gualeni, Dirk. Janssen, and Licia Calvi, 2012. How psychophysiology can aid the design process of casual games: A tale of stress, facial muscles, and paper beasts. *Proceedings of the International Conference on the Foundations of Digital Games*, pp. 149–155.
- [70] Anders Drachen, Alessandro Canossa, Sørensen J.R.M. (2013) *Gameplay Metrics in Game User Research: Examples from the Trenches*. Game Analytics. Springer.

Chapter 2

Profiling Personality Traits with Games

Abstract

Trying to understand a player's characteristics with regards to a computer game is a major line of research known as player modeling. The purpose of player modeling is typically the adaptation of the game itself. We present two studies that extend player modeling into player profiling by trying to identify abstract personality traits such as the need for cognition and self-esteem through a player's in-game behavior. We present evidence that game mechanics that can be broadly adopted by several game genres, such as hints and a player's self-evaluation at the end of a level, correlate with the aforementioned personality traits. We conclude by presenting future directions for research regarding this topic, discuss the direct applications for the games industry, and explore how games can be developed as profiling tools with applications to other contexts.

Summary:

Chapter 2 documents two studies which investigate the possibility of purposely developing games to predict personality traits and propose a systematic method.

Published in:

ACM Transactions on Interactive Intelligent Systems Journal (TiiS), Volume 9 Issue 2-3, Article No.: 11, March 2019, ACM New York, NY, USA

Authors:

Carlos Pereira Santos, Kevin Hutchinson, Vassilis-Javed Khan, Panos Markopoulos

Introduction

Personality refers to individual differences in characteristic patterns of thinking, feeling, and behaving [36]. Traditionally, psychologists have sought to explain such individual differences in terms of stable dispositional characteristics. The concept of traits is used in personality psychology to describe such dispositional characteristics as hostility, independence, self-esteem, helpfulness, and dominance, just to name a few [1].

For years psychological research has looked at predicting human behavior and attitudes on the basis of personality traits [1]. Assessing personality traits of individuals is best known for its direct application in clinical research for diagnosing and treating personality disorders [62]. The social and economic impact of such disorders are increasingly relevant, since the prevalence of subjects with borderline personality disorders is estimated to be between 15% and 25%, noting that there is a considerable lack of empirical research [40]. Key to diagnosing and treating such disorders is a reliable and efficient way of measuring personality traits.

However, personality assessment is very widely applicable outside clinical psychology as it can have applications in a variety of areas including education [73], management [10], direct marketing [52], media preferences [75] and persuasive communication [32]. Given this wide application of the concept of personality traits and their measurement, it is particularly interesting to find effective ways to measure.

Measuring Personality Traits

Self-report Based Measurements of Personality traits

There are multiple methods to measure personality traits. These can use stimulus, context, and situations to detect and measure attitude, behavioral and cognition manifestations; still most rely on self-reported verbal or written responses to questionnaire items [1]. Probably the best known questionnaire is the Revised NEO Personality Inventory (NEO-PI-R) [18], a 240-question, self-reported tool, which is considered the golden standard for measuring the Big-Five Personality Traits (neuroticism, extroversion, openness to experience, agreeableness, and conscientiousness) and demonstrates predictive validity in several domains. Because requesting participants to answer 240 questions is very time consuming (30-40 minutes), alternative questionnaires were developed and evaluated, such as the Ten-Item Personality Inventory (TIPI). These alternatives are typically shorter inventories and take less time to administer to participants, but are known to be less reliable [26].

In addition, there is a sense of unease that accompanies the use of such self-reported measuring tools, since responses may be systematically distorted by self-presentation bias or the reactivity phenomenon, especially if the personality

traits have strong social links or desirability that favor the participant. Several techniques have been developed and used to reduce or overcome these problems, like disguising the purpose of the inquiry or incorporating control items to evaluate the quality of the responses [1].

Implicit Measurements of Personality Traits

In recent years, new approaches to overcome self-representation biases have been uncovered, namely implicit measurements using priming of automatic responses to stimuli like words or pictures [28]. For example, a subject is asked to read out loud a word next to a picture. If both picture and word are positive (a “kitten” image and an “optimal” word), recognition and speed of pronunciation should be improved. On the other hand, if the image is replaced by a negative stimuli (e.g., “snake”), the new image will interfere with recognition and classification and affect the individual response time by a measurable amount. These new methods are known as Implicit Association Tests, and, on the one hand, they have demonstrated promising results [15], but, on the other hand, they have also drawn criticism, such as that they may not assess a person’s true trait, i.e., they may merely reflect learned associations that reflect cultural and social biases and not the individual’s own attitudes [11,34].

Analyzing user interactions to infer personality traits has been explored before in the domain of social networking systems. For example, it has been shown that linguistic features in a primarily text based medium (Twitter) [25] and visual features in an image and text based medium (Facebook) [15] seems that it could lead to a consistent decrease of the prediction errors for each personality trait.

Player Profiling

Implicit measures of personality and attitudes of users are typically applied in an experimental setting. In recent years, the rising popularity and availability of games has uncovered the ability to observe individual behaviors in the context of playing a game, an approach that is generally called player modelling [63].

Currently game developers apply player modelling primarily in order to adjust the game to the abilities and preferences of specific players [80]. This paper examines whether or not and how this technique of logging and analyzing player behavior can be systematically applied to modelling personality traits implicitly. Every single player interaction within a computer-game manifests player intent and behavior and the wealth of those interactions can be potentially catalogued and analyzed.

Gameplay Metric

A player’s sequence of movements, response times, avatar selection, customization, and even in-game purchases are all indicative examples of player interactions that are currently being logged and mined by game companies in order to strengthen their core business and design processes.

Definition

The term “Gameplay Metric” is defined as data-driven information that is collected within games and is in interpretable measures of in-game interactions by the players [21,72].

Such data supports predictive business considerations, such as expected sales or expected number of players, which are both valuable for commercial purposes in a very competitive market like the entertainment industry. The best and most documented case of the depth of the information that is being obtained is linked with former Eidos Interactive (now Square Enix¹), which published a series studies on player profiling, including how they used supervised learning algorithms to predict when a player will stop playing the game “Tomb Raider: Underworld” [43].

In addition to the application of game metrics for business purposes, it is becoming increasingly valuable to also track in-game player behavior to support game design, authoring, and content creation [20,43]. Simple metric, such as in-game position, frequency of used items, accuracy, and skill set, allow game designers to know more about the players, their behavior, proficiency, and experience. Games User Research [47] is a growing research area that focuses on the specific methodologies used to measure and interpret player behavior [42,49].

In short, Player Modeling is a research area that uses players’ in-game behavior to produce models that represent players’ behavior [80] with the primary goal of explaining, predicting, or mimicking player play style, and using it either to enhance gameplay [4,70] or to improve business models of games [43,69].

From Player Modeling to Player Profiling

For some years already, there has been an effort to develop player profiles using player models, i.e., instead of creating models that represent the players’ playing style, these models aim to represent players’ personalities. Past examples of personality profiling in this area include the Big-Five Personality Traits [38,39,81], the six dimensional HEXACO [78], and a person’s moral alignments [22]. [19] present a comprehensive design framework for mapping in game behaviors to personality traits; this has not yet been validated empirically. Although there already is a body of work investigating personality traits and their links to gameplay, there are still numerous salient traits that have not been considered.

1 <http://www.square-enix.com>

Games as a Personality Profiling Tool

Game environments offer ample opportunity for players to exhibit behavior within a controlled environment, while making it straightforward to record their actions. Several arguments can be made in favor of using games as a suitable platform for implicit psychologic measurements.

These are:

- Players can concentrate on measurement tasks more regularly and for longer periods, provided these are engaging and fun.
- The reactivity phenomenon, i.e., the fact that participants alter their behavior because they are aware that they are being observed, can be contained as players focus on the game play as a primary activity and can even be unaware of being observed (assuming this is ethically justifiable).
- Games provide the opportunity and flexibility to combine implicit and explicit measures.

Using a computer game as a profiling tool can have multiple applications such as a diagnostic tool with the explicit purpose of screening players, or a support tool for remote interventions and preventive care. Such a screening self-assessment tool is a realistic prospect given the high penetration of games [2]. We find that games can have an important role in preventive healthcare, where regular screening becomes particularly important for diagnosing and managing chronic diseases.

Future studies can also explore the temporal dimension of preventive care. We strongly believe that diagnosing through games is a cost-effective tool for healthcare organizations to use to target risk groups, engage patients in understanding the disease, and keep a communication channel open with practitioners while monitoring their health. Although these possibilities remain a long way ahead, our contribution takes the first steps in that direction by showing the feasibility of profiling personality traits through games. Future studies need to expand the evidence on the reliability of this measurement and its application in such screening and intervention contexts.

Below, we combine and extend the presentation of two studies originally reported in [58] and [54], which demonstrate how games and game mechanics can be purposefully designed and developed to profile specific personality traits, which here are Need for Cognition and Self-Esteem. Our presentation aims to show that game behavior can help predict these personality traits. Also, we demonstrate how a game can be systematically and purposefully developed to infer a specific personality trait by exploring the induced behavioral change by the trait within the game context. Finally, we discuss how specific player personalities are manifested in games by describing consequences, implications, and considerations in terms of game design.

Modelling Player's Need for Cognition: Study 1

Need for Cognition (NfC) is a simple and stable personality trait, and it has been applied across multiple situations, but it is best known for its use in marketing to predict the effectiveness of persuasion strategies, e.g., it has been linked to how individuals perceive information and the influence of contextualizing and framing effects [65].

Definition

NfC is a stable personality trait associated with the extent to which an individual is inclined to engage in cognitive activities [13]. High NfC defines individuals who are more inclined to mental problem solving and assessment of situations with a higher degree of elaboration. On the other hand, people with lower NfC are prone to less elaboration and follow more heuristic and empirical strategies [33].

Knowing an individual's NfC can be useful for a variety of purposes beyond marketing, most notably for providing tailored health communications to persuade individuals to engage in healthier behaviors. In the study presented in [67], it discusses how NfC influences users acceptance and compliance on smoking-cessation messages. It was found that individuals with lower NfC had greater intention to quit after reading a gain-framed message (i.e., what an individual would gain if they quit smoking), but a similar framing didn't affect people with higher NfC. Another study [76] involving 602 women evaluated the impact of tailoring persuasive health communications in favor of screening mammography; their results suggest that messages tailored to an individual's NfC are better at motivating them to participate in mammography preventive screening.

In the sections that follow, we report on a study that examines whether or not the score of such a traditional measure can be predicted by game metrics. Our study, similarly to the two above mentioned studies use the 18 item questionnaire [13].

The original NfC scale questionnaire [12] uses 34 items, but, subsequent research has proposed a shorter 18-item questionnaire [13] which is become the norm for measuring NfC scale of individuals. Although other researchers have investigated NfC relation to other social-personality measures (such as self-esteem, openness to experience) [8], or its inter-individual variations [14], or even investigated how to better administer the questionnaire itself [82], since 1982 those two questionnaires [12,13] have the only way to measure NfC scale. Thus, this would be the first attempt to have an implicit test of the construct, and it would be the first attempt to do that within a game context.

Design of a Game to Model NfC

We set out to design and develop a game that would allow differences in behavior between people with high and low NfC to manifest themselves naturally during game play. A key design decision in this process was that the game should not favor players with high or low need for cognition; were this not the case, players would quickly adopt the behavior that helps them win the game. Rather, the game was intended to provide a venue where the player's disposition (or relative lack thereof) towards mental elaboration would be possible to observe in terms of their spontaneous choice behavior exhibited during game play.

During our initial game design process, we focused on understanding the behavior and attitude alterations caused by NfC. Using an informal, small focus group (6 persons) with different NfC scales, we pretested some behavioral alteration theories, namely acceptance of information contextualization, framing effects, and reduction in cognition effort. The last mechanic showed consistent outcomes, allowing us to develop a paper prototype with some promising results, which matured into the design for a full game. In the end, we develop a two-player, turn-based game, where players control the movement of a set of units (Figure 2-1).



Figure 2-1: Left - Nanobots is a game designed to study players' Need for Cognition. The screenshot shows a typical player turn. Right - Need for Cognition scale questionnaire embedded into the Nanobots game is displayed before a game starts

Each player has a team and starts with three units. The overall objective is to eliminate the units of the opposing team. In a typical turn, the player can only move one out of the three available units. Each unit can only move up to two tiles and has three main properties: shield, energy (health), and attack power (displayed in the upper-right corner). Small vertical progress bars at the right of each unit represent energy. The key game mechanic is the unit movement; with every turn, each player has to decide which unit to move and where to move it. The game becomes challenging since, on average, a player can choose from 50 possible moves every turn, and, for every possible action, a player needs to consider their opponent's counter moves.

A secondary mechanic was purposely introduced to shortcut/reduce the cognitive load by introducing a hint system that supports the player by suggesting a reasonably good move. Hints are displayed directly on the board by lighting up tiles, indicating which unit should move and where to (yellow tiles in the left image in Figure 2-1). Hints can be requested explicitly by pressing the light-bulb button or implicitly if a timer times out. The timer is displayed via a progress bar. The time was fixed based on the average time a player takes to play; during the first testing phase, we found that a player took approximately 6 seconds, on average, so we set the timer to 4 seconds.

Hints were purposefully shown under the average play time, since we theorize that some players could be influenced by the hints, depending on their NfC scale, because they provide a cognitive load reduction during the decision process. It is also important to note that the hint mechanic suggests a beneficial movement, but not the optimal one; if the optimal movement were offered, playing the game would be contrived since hint followers would always win.

The game is strategic by nature, and any wrong move may cause the player to lose the game. Although players' performances may depend upon their intelligence, this should not be reflected in their NfC, since all information and cognition required is simple and clearly visible on the board.

Explanation of the Design of the Game Mechanics

Some NfC literature [13,64] clarifies how an individual's actions tend to differ depending on their NfC scale. The core behavioral biases that we focused on were the tendency of low NfC individual be more prone to accepting information without questioning it, and that high-NfC personalities tend to pursue understanding and consider thinking fun.

In that sense, our game mechanic requirements were a systematic and recurring mechanic that would facilitate good, but sub-optimal, mental shortcuts, which we believed could influence low- NfC players to make the suggested moves, while high-NfC players would pursue alternatives more frequently. After surveying different mechanics present in games, we found that the hint system fitted the requirements. We believe that other game mechanics that provide mental shortcuts could provide similar results.

Technical Implementation

We used the Unity game engine to develop our game with the support of a game company, and published it on the Google Play Store. Using the Unity Web Plugin system, the game was also made available on a website. The game included a telemetry library that allowed us to remotely collect all gameplay metrics explained below. Each player was solely identified by a unique ID, and all other data that could have identified the player in real life was discarded.

User Study

A quasi-experimental study was carried out to validate the main assumption underlying the game, namely that NfC can be predicted by player behavior. Remote users could download and play the game in their own environment as their game behavior and relevant game metrics were logged. They were asked to play several rounds of the game against a fair and well-balanced Artificial Intelligence (AI) adversary. Our analysis aimed to establish the extent to which NfC can be predicted by game metrics.

Before each game round, players were asked to answer three questions from the NfC questionnaire [13], requiring participants to play multiple rounds to fully complete the questionnaire, and we compared the results of those questionnaires with the purposely designed game mechanics.

Process.

A typical participant followed the following steps:

1. Received an invitation to play a web-based game or at the Google Play Store;
2. Start the game;
3. Filled in the NfC Scale Questionnaire;
4. Started playing the first round (starting with a tutorial);
5. Completed the first round;
6. Played the next round;
7. Closed the game.

Participants

All recruited players were explicitly informed that by participating they would be consenting to the collection of psychometric data for research purposes and assured that all logged data was anonymous. The purpose of the data collection was not fully revealed, suggesting that the data collection was for improving the AI adversary.

We uploaded the game online and on the Google Play Store as a free-to-play game. Participants were partly recruited through convenience sampling (mail campaign) and partly through crowdsourcing in order to recruit sufficient participants. A total of 188 participants took part in the study, but data from only 97 participants was retained, ensuring that all players included in the analysis had played at least 6 games (rounds) and had fully answered the NfC questionnaire [13].

Measures

The following game metrics were tracked anonymously across all games played and all players:

- In-game Metric 1 (M1) – Number of Wins. The metric recorded how many rounds a player won or lost out of all played games.
- In-game Metric 2 (M2) – Game Time. This metric saved the time each player took to complete a game.
- In-game Metric 3 (M3) – Number of Turns. As a last performance metric of player behavior, we also measured how many in-game actions were taken during a game round.

We did not expect to notice any difference in these three metrics between players with high and low NfC scores.

- In-game Metric 4 (M4) – Hints Pressed. This game metric measured the total number of hints a player explicitly requested during the six initial games. Hints provided a natural extension to the basic game mechanic as a support for less experienced players. At the same time, whether or not the player will rely on the hint could arguably relate to the player's NfC, as explained previously. We hypothesized that players with lower NfC would take more shortcuts during the decision-making process by requesting more hints. Since there was no downside to requesting or using hints, this metric served the purpose of measuring the explicit intent of a player requesting a reduction in the cognitive process.
- In-game Metric 5 (M5) – Hints Followed. As explained above, hints were displayed, even if the player did not explicitly request them, after 4 seconds. The intention was to influence players in their decision-making process by providing a suggestion, and, therefore, attempted to persuade the player to make a specific move. Note that, compared with the previous metric (M2), hints followed are a posterior metric and only measure if the player is influenced by hints. We expected that players with lower NfC would be more influenced by the play that is suggested because it reduced the cognitive load. This game metric logged the frequency of a player following a suggested move.

Hypotheses

There are two main hypotheses that we can summarize as:

First, the game is not biased towards Need for Cognition (H1). The main game mechanic is based on strategic thinking and not directly related to NfC. All the information is clearly presented to the player and should, therefore, not present any data correlations between NfC and M1, M2 and M3.

Second, players with higher NfC should use hints less frequently (H2). At least one of the hint mechanics, M4 (the explicit request of hints) or M5 (hints followed,) should demonstrate a negative correlation with players' NfC, i.e., the higher the NfC of the player, the less hints they should use.

Results

We classified the participants according to their NfC scores and analyzed the results of the rounds 3 to 6, excluding round 1 and 2 from the analysis to avoid confounding learning effects. Rounds 1 and 2 were included in the initial summary, merely for completion and transparency.

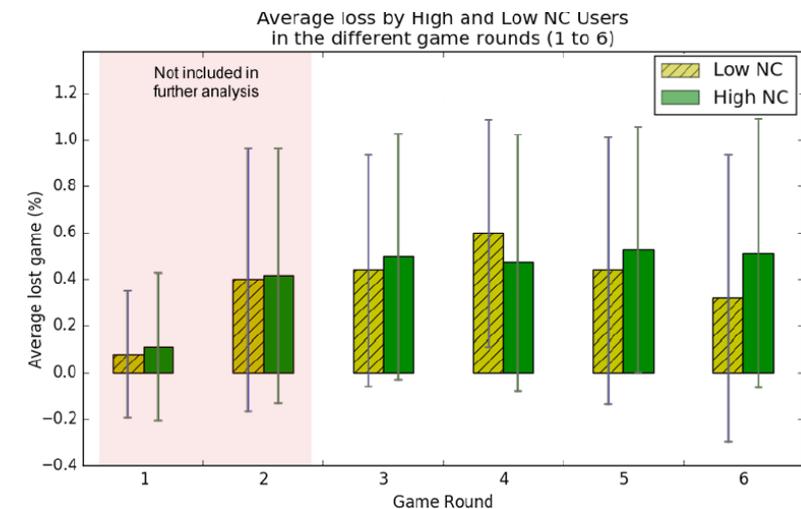


Figure 2-2: Average game losses for high- and low-NfC players, including the standard deviation error.

Figure 2-2 shows the average game scores of low- and high-NfC players, and, excluding game 1 (tutorial); there is no noticeable difference between high- and low-NfC players.

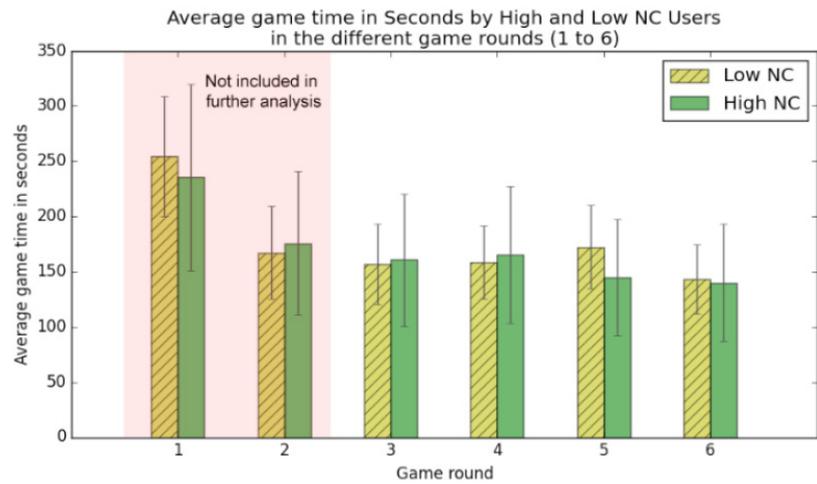


Figure 2-3: Average game time for high- and low-NfC players, including the standard deviation error.

The average game time is depicted in Figure 2-3, and it shows a substantial decrease in time spent playing a round after the first round for both high- and low-NfC players.

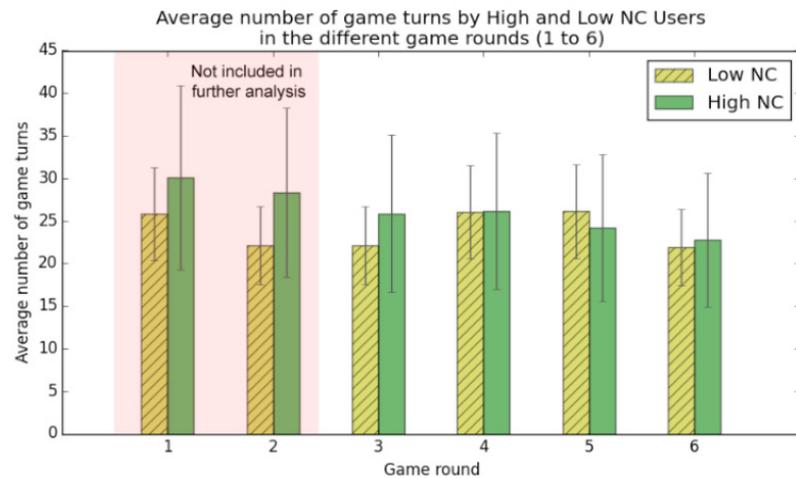


Figure 2-4: Average number of turns for high- and low-NfC players, including the standard deviation error.

We also examined the average number of turns for both high and low NfC (see Figure 2-4), and across all games no significant difference was found. Results of all three metrics seemed to remain relatively stable in games 3 to 6, supporting that the game was sufficiently understood and no confounding variables seem to be perceptible.

A similar examination was performed for both hints explicitly requested and

followed, with the results shown in Figure 2-5 and Figure 2-6. An immediately visible result (Figure 2-6) is that a large number of players did not explicitly request hints; the average number of explicitly requested hints is very low. Nonetheless, it seems that, on average, low-NfC players request at least one hint, while high-NfC players request approximately zero. With such low numbers, it is hard to draw conclusions with regards to explicit hint requests.

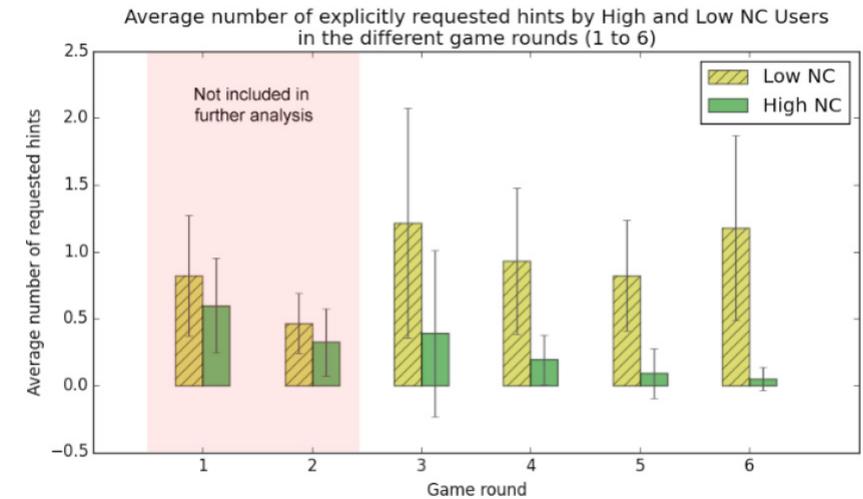


Figure 2-5: Average explicitly requested hints by high and low NfC players, including the standard deviation error.

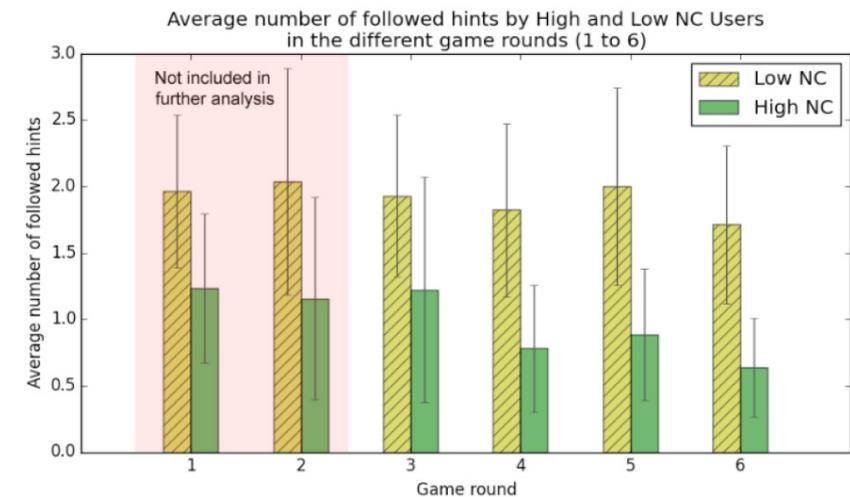


Figure 2-6: Average number hints followed by high and low NfC players, including the standard deviation error.

On the other hand, hints followed (Figure 2-6) are clearly manifested in the game round with significant differences between high and low NfC. As suspected, it seems that players with high NfC use fewer hints than those with low NfC.

Correlation Analysis

The results of Spearman's correlation analysis, as depicted in Table 2-1, shows similar results: all game metrics related to game bias (M1, M2, and M3) did not show any correlation, as expected, and M5 shows a moderate correlation between player NfC and the number of hints followed, confirming what we suspected. The only variable that did not confirm a hypothesis was M4. It was clear in Figure 2-5 that the average number of explicitly requested hints was very low for all players. This suggests that there may be a game design problem.

NfC & In-Game Metrics	(N)	(r)	(p)
(M1) Win/Lose Ratio	97	0.028	0.787
(M2) Game Round Time	97	-0.016	0.878
(M3) Number of Turns	97	0.056	0.588
(M4) Hints Explicit Requested	97	-0.149	0.115
(M5) Hints Followed ^a	97	-0.223	0.028 ^a

^a Significant result

Table 2-1: Spearman's correlation results between NfC and "Nanobots" in-game variables

Regression Analysis

To conclude the data analysis, we performed simple linear regressions with NfC as the dependent variable and M1-M5 as the independent variables. The results are presented in Table 2-2 and show that the linear regression results in a moderate correlation and significant prediction of NfC by both M4 and M5, with a reasonable error margin. As expected M1, M2, and M3 were unable to predict NfC.

NfC & In-Game Metrics	(N)	(m)	(b)	(r)	(p)	(stderr)
(M1) Win/Lose Ratio	97	0.003	1.935	0.033	0.748	0.008
(M2) Game Round Time	97	-0.049	617.557	-0.003	0.976	1.597
(M3) Number of Turns	97	0.068	98.393	0.037	0.718	0.187
(M4) Hints Explicit Requested ^a	97	-0.080	2.015	-0.220	0.030 ^a	0.037
(M5) Hints Followed ^a	97	-0.127	5.902	-0.341	0.001 ^a	0.036

^a Significant result

Table 2-2: Linear regression results between NfC and in-game variables.

Discussion

We confirmed both of our hypotheses. In all performance-related game metrics (M1, M2, and M3), we didn't find any relation with NfC, making the developed game unbiased towards NfC scale, confirming H1.

In addition, we showed that NfC influences the use of hints, although that was not manifested in the correlation of NfC with the explicit request for hints (M4). We conjecture that the design of always showing the hint after 4 seconds may have conditioned the player to simply wait for the suggestion to appear. This was not revealed in our initial design tests and, therefore, influenced the Spearman's correlation, but both the game-round analysis and linear regression show the expected differences.

This complication in the hint mechanic introduces an ambiguity in our results, as we cannot distinguish instances where the player wished for a hint and got it automatically from instances where the player did not wish to have a hint. Explicitly requesting hints addresses the player's need for advice or support, whereas following hints or not addresses whether or not the player is actually influenced by the hint. We are planning to investigate this difference in future research.

Nevertheless, the H2 is confirmed by the negative correlation between followed hints (M5) and NfC, supporting an alternative and unobtrusive method of measuring NfC – a general and salient user trait – with a specific and commonly used game mechanic. The approach presented here offers several advantages with regards to ensuring a pleasurable and unobtrusive profiling of users.

Such profiling can improve players' experiences in several ways. For example, we would recommend that in strategy games, such as our Nanobots game, high-NfC players should receive additional (visually and conceptually related) cognitive tasks – e.g., puzzles – to solve before going to a next level, while low-NfC players should not. In other game types, such as adventure games, high-NfC players should get more riddles to solve compared to low-NfC players; or, in exploration games (e.g., RPGs or puzzle-platforms), low-NfC players should receive an influential figure (e.g., priest or political leader) to impact their engagement and exploration with the game.

Further, when considering other game aspects, such as content purchases, high-NfC players should receive in-depth information and low visual complexity as this would be favorable for them; whereas, low-NfC players should be given more visually complex stimuli. Naturally, the aforementioned recommendations need to be researched in terms of whether or not they would actually contribute to players' experiences.

Modelling player's Self-esteem: Study 2

Self-esteem reflects an individual's opinion and evaluation of his or her personal worth [7]. It is a common personality trait used to support the diagnosis of several psychological medical conditions such as depression [66], eating disorders [60], and narcissistic personality [74], among others.

Definition

Self-esteem, or, more precisely, Global Self-Esteem, is often defined as “the affective evaluation of one's worth, value or importance” [7]. It can be seen as the attitude towards an object, primarily focusing on the positive or negative disposition of this object and the associated emotional reactions, where the object and the holder of the attitude are the same [41].

It is important to emphasize that self-esteem is an opinion about the self; a person with high self-esteem does not have to be arrogant, conceited, or excessively prideful. High self-esteem does not automatically imply superiority, nor does low self-esteem automatically imply inferiority. However, these feelings do occur more often amongst people with high or low self-esteem, respectively.

Though self-esteem is the most common term used to refer to the described concept, many others have been used, including, but not limited to, self-concept, ego, self-worth, and self-image. Some terms are used interchangeably with self-esteem, while others denote a more specific or slightly different concept. Furthermore, self-esteem is a subjective individual trait and is not linked with someone's abilities or proficiency while performing a certain activity [7].

Self-esteem is a stable personality trait used and influences transitory states of mind, such as mood or anxiety. Stable personality traits are strong influences over transient mind states, but the opposite is not true [45,71].

Severe low self-esteem is one of the most common mental health problems that can affect women during the perinatal and postnatal period, with a point prevalence of up to 19% [29]. Known as perinatal and postnatal depression, this disorder directly accounts for the majority of costs attributed to perinatal mental disorders. In the UK alone, the average cost of each case of perinatal depression has been estimated at €105.000,00 [6].

Differences in self-esteem also demonstrated impacts in multiple other areas, like organizational psychology when looking at the relation to job satisfaction [24], and its impact on educational achievement [51].

Motivation

Traditionally, practitioners have used questionnaires to measure self-esteem [57]. The emergence of serious games as an area of investigation has opened up the possibility of using games to profile players. Stable personality traits can be

recognized as these manifest themselves in players' in-game behavior [3,27]. One major potential advantage of such an approach is that games can act as an unobtrusive instrument that is already embedded in everyday activities.

Arguably, profiling through observing behavior can potentially be more reliable than self-reporting through questionnaires, which can be subject to social desirability biases. Finally, going beyond measurement, games have the potential of being utilized for treatment purposes. However, there has been yet no earlier attempt to measure self-esteem by analyzing in game behavior.

Explicitly Measuring self-esteem

Evaluating self-esteem is almost exclusively done with self-test methods. Self-test methods usually measure explicit self-esteem, i.e., the conscious evaluation of the self. Because of the conscious evaluation, explicit self-esteem measurements may be influenced by the person being evaluated. Filters include social-desirability, self-deception, and personality, among others. The Rosenberg Self-Esteem Scale (RSES) is the most-widely used scale for measuring self-esteem in the social sciences [57].

The subconscious evaluation of self-esteem is known as “implicit self-esteem,” which has been defined as “an automatic, overlearned, and nonconscious evaluation of the self that guides spontaneous reactions to self-relevant stimuli” [9]. Due to it being a ‘natural’ response, implicit self-esteem is best measured by keeping the participant unaware of what is being evaluated or how this is being done.

Both explicit and implicit aspects are useful in their own right, but, by examining the differences between the two, information can be gained about a person's filters, which neither the explicit nor the implicit aspect provides on its own.

Implicit Measurement of Self-Esteem

There are no well-established methods to measure implicit self-esteem. The most frequently cited measures are:

- The Self-Esteem Implicit Association Test [28], which uses an individual's automatic association between mental representations of concepts in memory, and
- The Name-Letter Effect [50], which uses an individual's tendency to favor the letters in their name over the other letters of the alphabet.

However, these measures have recently received substantial criticism, which go so far as to state that they do not measure self-esteem at all since there is no strong or consistent support for the validity of either of those measures [11]. Further research is therefore needed to establish a method of testing implicit self-esteem.

Measuring of Self-Esteem Using a Game

We could not find previous work that attempted to establish a link between the personality trait in question and in-game behavior. For the particular case of self-esteem, the growing interest in games [2], offers an opportunity to investigate alternative ways for its measurement. The measurement details of self-esteem go beyond the scope of this paper. For further information on the type of measurements of this construct we would refer the reader to [56].

Study Research Plan

With this study, we contribute to the literature by investigating how games and in-game behavior relate to a player's self-esteem. More specifically, our research questions are:

- How does a player's self-esteem influence game performance?
- How does a player's self-esteem relate to in-game behavior?
- Which game mechanics are suitable for inferring a player's self-esteem?

We discuss how our findings may influence today's game design practices and how our work opens the possibility for game developers to introduce games that explore a player's self-esteem.

To answer the research question "How does a player's self-esteem relate to in-game behavior?", we conducted a study in which specific game mechanics and gameplay metrics were compared to the results of the RSES. We performed the study in two phases: the first phase with 98 participants, and the second with 85 participants.

According to the literature [45,71], since self-esteem is a stable personality trait it is not expected to be influenced by other transit states of mind, such as mood or anxiety. This reduces the influence of the conditions the subjects perform the experiment in and makes online administration viable.

Participants

In the first phase of the experiment, we tried to reach a large number of participants by using social media and personal contacts, which resulted in a total of 98 participants. Due to the nature of our social media network, most participants were rather young, with an average age of 22 years old ($SD= 3.31$), and very familiar with computers and games.

In the second phase, we wanted to reach a more diverse group, so we used a crowdsourcing service (Prolific.ac) to recruit participants and provided a small monetary reward (€5,00) for completing the experiment. The average age this time was somewhat higher ($M=26$, $SD= 5.74$), and the participants were less familiar with games.

In both phases, participants were informed and agreed on the fact that data was being collected for research purposes, and all logged data was anonymous, although the purpose of the study was not disclosed.

Procedure

A typical participant followed the following steps:

1. Received an invitation to play an online game;
2. Opened a URL to start the game;
3. Filled in the RSES (Figure 2-7);
4. Completed a profile questionnaire (age, gender, computer and game experience);
5. Started playing the first level (starting with a tutorial);
6. Completed the first level (multiple attempts might have been required);
7. Evaluated own level performance;
8. Played the next level and so on until all levels were completed;
9. Closed the game.

Statement	Strongly agree	Agree	Agree somewhat	Undecided	Disagree somewhat	Disagree	Strongly disagree
I feel that I am a person of worth, at least on an equal plane with others.			Selected				
I feel that I have a number of good qualities.					Selected		
All in all, I am inclined to feel that I am a failure.						Selected	
I am able to do things as well as most other people.							Selected
I feel I do not have much to be proud of.							Selected
I take a positive attitude toward myself.						Selected	
On the whole, I am satisfied with myself.							Selected
I wish I could have more respect for myself.						Selected	
I certainly feel useless at times.							Selected
At times I think I am no good at all.							Selected

Figure 2-7: The Rosenberg Self-Esteem Scale, which was embedded within our game.

Game Design Elements

Similar to the first study, we developed a purpose-specific game, titled 'Runner.' Runner is a barren, 2D platformer with a negligible narrative. The simple look and feel was an explicit choice that was made to avoid strong influences like plots, character self-representation, or empathy, keeping the game abstract.

For the same reason, we purposely did not add an avatar to the game. The player was represented by a centered red rectangle (Figure 2-8 & Figure 2-9).

The goal of the game is to reach the end of a level as fast as possible. The only actions the player can take are moving left, right, jumping, and grabbing (also known as wall jumping). The game is controlled with the keyboard keys W, A, and D, or the arrow keys and Space, which are common for platformer games. Due to the game genre, it is expected that a regular player will need multiple attempts to complete a certain level.

The game has two tutorial levels in which the controls are explained, and the player can practice basic game mechanics. After the tutorials, the game has eight levels of increasing difficulty. All data collected from the tutorial levels were ignored in our analysis to limit spurious effects observed during learning.

In addition to the minimalist look, we intentionally designed a set of mechanics, which we believe might relate with self-esteem and describe next.

In-game Metric 1 (M1) – Path Choices

In each level, the player has to choose one of two different paths. These paths are different routes to the end of the level and normally appear once each level. The paths are explicitly marked as “Easy” and “Hard” (logged as paths 0 and 1, respectively), but there are no penalties imposed for choosing either path presented (see Figure 2-8).

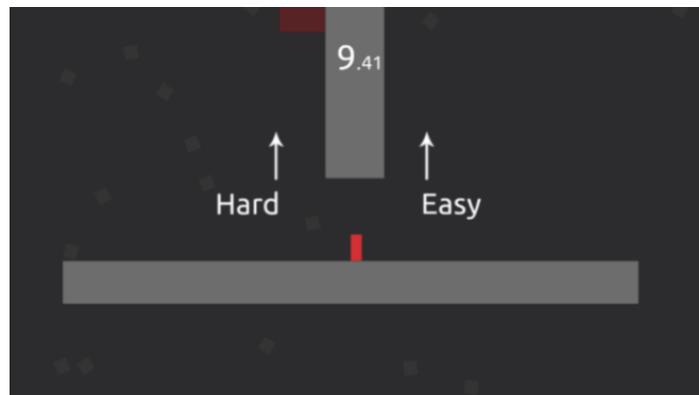


Figure 2-8: Example of a path choice (M1).

Our hypothesis was that players with lower self-esteem would pick the easy paths more often than the players with higher self-esteem would, i.e., players with lower self-esteem would use the less complex paths more frequently, while personalities that are more self-confident would attempt the harder paths more often.

For this mechanic, the following information was logged per level:

- M1.1 Frequency – Frequency of choosing the hard path, considering all attempts.
- M1.2 1st Attempt – The path that the player took the first time a level is played.
- M1.3 Complete – The path the player took when he/she successfully completed a level.

In-game Metric 2 (M2) – Player Pace

Levels include timeout zones: yellow areas (depicted in Figure 2-9) in which the timer (score) is paused, allowing players to have a “non-stress” zone.

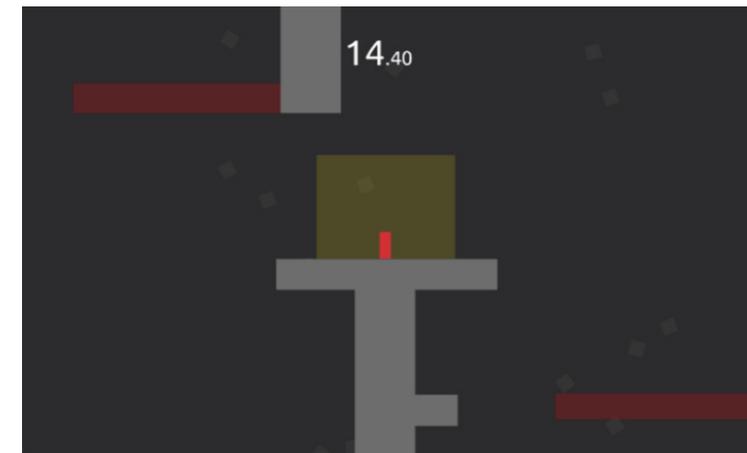


Figure 2-9: Example of a timeout zone (M2).

Our hypothesis is that players with higher self-esteem would remain in the timeout zones for shorter periods than players with lower self-esteem. Once again, we expected that less self-confident participants would take longer to re-evaluate and plan ahead and, therefore, spend, on average, more time in the timeout zones. In this mechanic, we logged per level:

- M2.1 Total Time – Total time spent by a player in the timeout zones.
- M2.2 1st Attempt – Time spent in the timeout zone in the first attempt of a level.

In-game Metric 3 (M3) – Performance Self-Evaluation

When players successfully complete a level, they are asked to rate and compare themselves to others. It is important to highlight that, at this point, no score information has been provided to the player.

Nevertheless, while playing a level, a timer could be seen at the top center of the screen (Figure 2-8 and Figure 2-9). The timer displays how many seconds have elapsed since the beginning of that attempt.

While obtained results are expected to depend on the player's performance, we hypothesized that the self-evaluation would be affected by the player's self-esteem. We chose three questions that we expected might be influenced by players' self-esteem (see Figure 2-9):

- M3.1 Relative – “Do you think your time was above or below average?” – Player self-evaluation comparing performance to the concept of an average player; the answer was purposely designed to be bipolar.
- M3.2 Absolute – “How would you rate your performance in this level?” – Player self-evaluation of performance on a ten-point scale.
- M3.3 Reported errors – “Please mark places where you feel you made mistakes.” – A mini-map version of the level is presented to the players, and they are requested to explicitly indicate areas where they thought they made mistakes. In that way, we could measure the number of mistakes, while at the same time have more detailed information on the places within the game where they thought they made errors.

Explanation of the Design of the Game Mechanics

Different literature pointed to different behavioral influences based on self-esteem:

First, we found the work of [30], which expanded on how self-esteem influences decision-making and risk assessment. Hence, we decided that we should include a game mechanic for which the requirement was to allow the player to evaluate two or more choices with different perceptions of risk and skill requirements.

Second, other works explain that self-esteem influences rhythm and pace, as measured in the Implicit Association Test [18]. Based on this literature, we defined the requirements for our second game metric, reviewing player stride and exploring player pace variations.

Third, the more traditional definition of self-esteem [7] is purely based on an individual reflective measurement. This meant that our third metric should review evaluative and reflective player behavior.

Summarizing, based on listed self-esteem behavior biases, we defined the requirements, then designed, developed, and integrated them within the game mechanics to assess if they would manifest behavior differences between high and low self-esteem.

Technical Overview

The game was developed in a popular game engine (Unity), exported to HTML5 format, and made available online. The game includes a telemetry library, which logged the in-game variables identified above.

The individual player data was sent to and stored in our self-developed online analytics platform. We assigned each player a unique identifier. This unique identifier served two main purposes: firstly, it identified the player, and, secondly, it ensured anonymity, since no other identification was requested or stored in our database.

The in-game analytical platform was developed to serve multiple projects. Its main goal was to log and centralize all the information in a relational database for posterior analysis. The collected data was then analyzed with statistical software, and the results are presented in the next section.

Research Hypotheses

To answer the research question, we tested the following hypotheses:

Player Self-Esteem Does Not Influence Performance (H1)

The literature explains that self-esteem is not linked to a person's individual ability [7], and we wanted to test this with the game we developed. We evaluated H1 by calculating the correlation between the measured game scores and the results of the RSES. The performance indicators that were measured were:

- P1 – Score of the level, i.e., time taken on the first successful attempt to reach the end of the level.
- P2 – Total time, i.e., the total time a player spent in a level, including all failed attempts.
- P3 – Number of attempts, i.e., the number of attempts required to complete the level.

Due the fact that levels have different lengths, degrees of difficulty, and challenges, it is necessary to calculate a Uniform Performance measure which can be normalized and, thus, be comparable across multiple levels. We defined this measure as:

$$UP_L(p) = -\frac{P_L(p) - \overline{P_L}}{stdev(P_L)} \quad (1)$$

where p represents the player, and PL represents one of the performance indicators for all players in a specific level. The result is negated because of the way the score works, i.e., the longer a player takes, the worse the score; scores are inverted so that positive values represent above-average performances.

For example, if a specific player p0 took 56 seconds to complete level l, and the average time players took for the level was 81 seconds with a standard deviation of 43.8 seconds, then the performance score for player p0 is 0.57. If, for the same level, another player p1 took 93 seconds, this would net the player a score of -0.27.

Player Self-Esteem Influenced In-Game Behavior (H2)

We developed two in-game mechanics (M1 and M2) to test if the players' pace and decisions are influenced by self-esteem. Those mechanics are directly associated with the players' belief in self-capabilities and are testing whether or not the player's in-game behavior is affected by self-esteem.

Spearman's rank correlation is used to test whether or not self-reported RSES values correlate with logged players' data regarding their path choices (M1) and/or timeout zones (M2).

We consider self-evaluation (M3) to be a game mechanic since it is related to the game and the player's performance, but it does not measure players' in-game behavior. Therefore, it was not evaluated for this particular hypothesis. The game mechanic was to keep it simple and direct, but there are ways to increase integration with the game, for example, by linking it with reward systems.

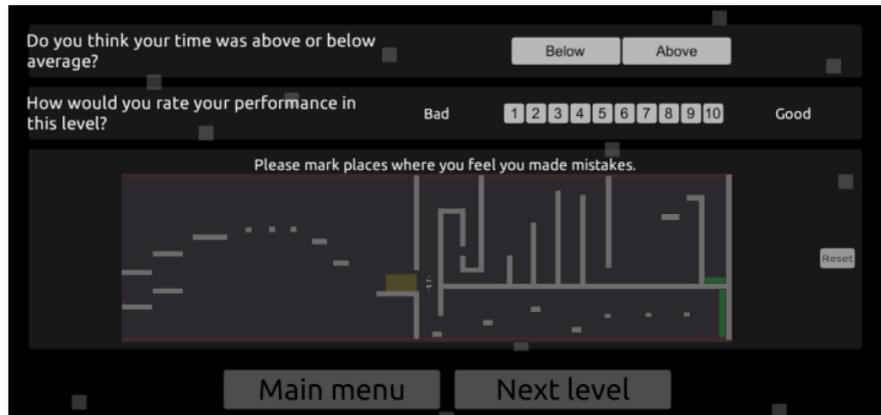


Figure 2-10: Player self-evaluation performance form presented after each level (M3 – first iteration).

Player Self-Esteem Can Be Inferred By Game Mechanics (H3)

Similar to the previous hypothesis, if strong correlations can be found between player self-esteem and in-game decision (M1) and/or pace (M2), then it will be possible to confirm this hypothesis. In addition, we will also test the correlation between M3 and RSES for this hypothesis since the player's self-evaluation could be affected by the player's self-esteem.

While each question (M3.1, M3.2 and M3.3) will be tested for correlation with self-esteem, we do not expect to find correlations because the performance directly influences the self-evaluation proposed by M3 since that is the focus of the questions.

Therefore, we propose the following equation for calculating the Self-Evaluation Bias:

$$SEB_L(p) = UP_L(p) - 2 * |M3.X_L(p)| - 1 \quad (2)$$

Self-Evaluation Bias is a scale that explains how much bias a player has in relation to the actual score. Since the range of Uniform Performance is between -1 and 1, where 0 represents an average player, we converted the M3.X answers into the same scale and subtracted both, leaving us with the difference between the actual performance and the player's self-evaluation. This Self-Evaluation Bias translates into (Table 2-3):

$SEB_L(p) < 0$	$SEB_L(p) \simeq 0$	$SEB_L(p) > 0$
Modest evaluation	Fair evaluation	Exaggerated evaluation

Table 2-3: Interpretation of the values that are obtained by the Self-Evaluation Bias value. Results

In this section we analyze the collected data in the first phase of the experiment, justify the alterations for the second phase, and then report the results of the second phase.

First Phase – Performance (H1).

In line with our expectations, based on the literature, we could not find a correlation between self-esteem and player performance (see Table 2-4). We calculated the correlation based on data from all completed levels (except for the tutorial) by all players and the respective player's RSES.

Performance & RSES	(N)	(r)	(p)
P1. Level Score	172	- 0.117	0.393
P2. Total Time	172	- 0.158	0.250
P3. Attempts	172	- 0.139	0.313

Table 2-4: Spearman's correlation results for the 1st phase between RSES and Player Performance.

We also could not find any statistical correlation between any of the three performance metrics we measured and the player's self-esteem. One interpretation of these results is that this specific game is not biased towards self-esteem. Prior literature suggests that self-esteem is unrelated to a person's individual ability [7].

First Phase – In-Game Behavior (H2)

As shown in Table 2-5, we could not find any relationship between self-esteem and the player's choice of path for all of the variables we logged (M1.1, M1.2 and M1.3). It is interesting to observe that self-esteem seems not to influence this specific behavior.

Although not one of our original hypotheses, it is interesting that we found a correlation between self-reported computer experience and the first path players picked in the levels (N= 172, r=0.266, p=0.05). It could be that players with more experience using computers (obtained by a self-report using a ten-point scale) are more likely to pick hard paths the first time they play a level. This result would need to be tested and confirmed in future experiments.

Mechanics & RSES	(N)	(r)	(p)
M1.1 Frequency	172	0.099	0.194
M1.2 1st Attempt	172	0.017	0.820
M1.3 Complete	139	0.130	0.127
M2.1 Total time ^a	150	0.228	0.005 ^a
M2.2 1st Attempt	150	0.133	0.104

^a Significant result

Table 2-5: Spearman's correlation results for the 1st phase between RSES and in-game behavior Mechanics (M1 & M2).

With regards to M2.1, the total time players spent in the timeout zone positively correlated with self-esteem (Table 2-5). Thus, this is a promising direction to highlight, i.e., that the player's pace in the game relates to self-esteem.

Mechanics & Age	(N)	(r)	(p)
M2.1 Total ^a	150	- 0.311	0.0001 ^a
M2.2 1st Attempt ^a	150	- 0.237	0.0034 ^a

^a Significant result

Table 2-6: Spearman's correlation results for the 1st phase between Age and Timeout Mechanic (M2).

Although a statistical correlation exists, we cannot confirm that self-esteem influences a player's pace since we also found a strong correlation between M2 and Age (Table 2-6). Therefore, there might be other indirect relationships, but we would require more data, and perhaps also using a more generic target group, to confirm this result.

Self-Evaluation Bias & RSES	(N)	(r)	(p)
Using performance P1 – Level Score			
M3.1 Relative	246	0.001	0.986
M3.2 Absolute ^a	246	0.147	0.021 ^a
M3.3 Reported error	246	0.031	0.634
Using performance P2 – Total Time			
M3.1 Relative	246	- 0.037	0.559
M3.2 Absolute ^a	246	0.170	0.007 ^a
M3.3 Reported error	246	0.013	0.833
Using performance P3 – Number of Attempts			
M3.1 Relative	246	0.009	0.891
M3.2 Absolute ^a	246	0.135	0.036 ^a
M3.3 Reported error	246	0.059	0.360

^a Significant result

Table 2-7: Spearman's correlation results of the 1st phase between RSES and Self-evaluation bias (M3).

First Phase – Self-Evaluation Bias (H3)

The results presented in (Table 2-7) show a correlation between self-evaluation (M3.2) and self-esteem in all performance indicators. When having a closer look at the results we observed discrepancies, specifically regarding M3.1 and M3.3.

First, when we asked players to compare themselves with other players (M3.1) by using a dichotomous scale ("below" and "above") we created a rather crude scale, and we realized that a more refined scale was necessary. Second, we noticed a mistake with the self-reported number of errors (M3.3). The form we provided could not distinguish between the players that were slothful (and would not report errors) and the players that purposefully wanted to report zero errors, which resulted in invalidating the results we wanted to measure. Thus, both M3.1 and M3.3 could not support the associated hypothesis.

First Phase Summary Analysis

Although, we found promising results in the first phase, namely that M2.1 (total time a player spends in a timeout zone) and M3.2 (absolute self-evaluation bias) showed statistical correlations with RSES, we had some reservations with regards to some results. Hence, we decided to run a second phase of the study to validate the results.

Based on the findings of the first phase, we planned a second phase, maintaining the same hypotheses, gameplay and mechanics, but altering the following:

- We noticed that harder levels seemed to provide stronger evidence than easier levels, so we increased the complexity of two levels by making them longer.
- We targeted a more diverse, generic user group to have a more representative sample.
- We changed M3.1 (the comparison scale) to a three-point scale (i.e., Below, Equal, Above) in order to have a slightly more refined scale while still keeping it simple to answer.
- We changed M3.3 by harmonizing the effort of reporting the errors and required players to explicitly report zero errors if that was the case.

Second Phase - Performance (H1)

The results were consistent with the first phase, showing that the player's game performance does not correlate with RSES (Table 2-8).

Performance & RSES	(N)	(r)	(p)
P1. Level Score	181	0.112	0.133
P2. Total Time	170	0.097	0.206
P3. Attempts	181	0.010	0.892

Table 2-8: Spearman's correlation results for the 2nd phase between RSES and Player Performance.

Second Phase - In-Game Behavior (H2)

When analyzing the in-game behavior with regards to player choices (M1), we found similar results, i.e., we could not find any correlation between player self-esteem and the path he/she picked (Table 2-9).

Mechanics & RSES	(N)	(r)	(p)
M1.1 Frequency	214	0.003	0.963
M1.2 1st Attempt	214	0.021	0.756
M1.3 Complete	181	0.025	0.742
M2.1 Total Time	236	-0.070	0.280
M2.2 1st Attempt ^a	236	-0.094	0.073 ^a

^a Significant result

Table 2-9: Spearman's correlation results for the 2nd phase between RSES and in-game behavior (M1 & M2).

In the data collected in this phase, it appears that there is a tendency for players to spend more time in the timeout zone in the first attempt to complete a level (M2.2) if they have low self-esteem. However, the correlation was very low (-0.094). Also, we did not find any correlation with M2.1.

Mechanics & Age	(N)	(r)	(p)
M2.1 Total	236	-0.097	0.137
M2.2 1st Attempt	236	-0.074	0.256

Table 2-10: Spearman's correlation results for the 2nd phase between Age and Timeout Mechanic (M2).

Although in the first phase we found a strong correlation between the timeout zones (M2) and players' age, this study could not replicate that result (Table 2-10).

Second Phase - Self-Evaluation Bias (H3)

Similar to the previous iteration, we found moderate correlations with statistical significance between player's comparison to others and self-rating of performance. The third item we introduced in the scale (equal) allowed for more precision and confirmed the result we got from the first phase (Table 2-11).

Even though we modified M3.3, the self-reported indication of errors to not allow ambiguity caused by slothful players and zero mistakes, we did not find any significant correlation.

Self-Evaluation Bias & RSES	(N)	(r)	(p)
Using performance P1 – Level Score			
M3.1 Relative ^a	181	0.186	0.012 ^a
M3.2 Absolute ^a	181	0.169	0.022 ^a
M3.3 Reported error	181	0.085	0.255
Using performance P2 – Total Time			
M3.1 Relative ^a	181	0.262	0.0004 ^a
M3.2 Absolute ^a	181	0.246	0.0001 ^a
M3.3 Reported error	181	-0.043	0.563
Using performance P3 – Number of Attempts			
M3.1 Relative ^a	181	0.195	0.009 ^a
M3.2 Absolute ^a	181	0.197	0.008 ^a
M3.3 Reported error	181	-0.085	0.909

^a Significant result

Table 2-11: Spearman's correlation results for the 2nd phase between RSES and self-evaluation bias (M3).

Second Phase – Regression Analysis

In this phase of the study, we confirmed one of the results found in the first iteration; namely, the self-evaluation bias. In addition, there is an indication that pace might be affected by players' self-esteem.

As in the previous study, we proceeded with a linear regression analysis of the in-game metrics based on the last results obtained. The results in Table 2-12 show that it is possible to obtain simple linear regressions with moderate correlations and significant predictions, although with moderate error margin.

Hypothesis Review

Based on our study results, we can confirm that there is no correlation between RSES and any of the three performance indicators we measured.

With regards to our second hypothesis, we could not confirm that self-esteem influences player behavior. Of the two game mechanics that were purposely designed to test self-esteem influence, the first M1 (Path Choice) did not show any indication of correlation with self-esteem, and M2 (Player Pace) had mixed results.

RSES & Bias Linear Regression	(N)	(m)	(b)	(r)	(p)	(stderr)
Using performance P1 – Level Score						
M3.1 Relative ^a	181	0.05	-1.916	0.165	0.02772 ^a	0.023
M3.2 Absolute	181	0.03	-1.598	0.142	0.05682	0.020
M3.3 Reported error	181	0.02	-0.813	0.067	0.37197	0.022
Using performance P2 – Total Time						
M3.1 Relative ^a	181	0.08	-2.888	0.304	0.00003 ^a	0.018
M3.2 Absolute ^a	181	0.07	-2.570	0.282	0.00012 ^a	0.017
M3.3 Reported error	181	0.05	-1.785	0.138	0.06438	0.025
Using performance P3 – Number of Attempts						
M3.1 Relative ^a	181	0.06	-2.401	0.234	0.00153 ^a	0.020
M3.2 Absolute ^a	181	0.05	-2.082	0.214	0.00383 ^a	0.018
M3.3 Reported error	181	0.03	-1.298	0.096	0.19676	0.026

^a Significant result

Table 2-12: Simple linear regression analysis for the results of the 2nd phase between RSES and self-evaluation bias (M3).

When we take a closer look on our findings, we believe that the player's path choice (M1) targets his/her skill level, but is not a conscious self-reflection on self-worth. In other words, while the player is engaged in playing the game he/she is acting and not reflecting. After further consultation, we found another study demonstrating a similar result; this specific study used electrophysiological recordings (electroencephalogram) to investigate the decisions and risk assessments of blackjack players. They found some differences in emotional signals between low and high self-esteem, but, interestingly, not in the behavior, i.e., there is no difference during risky decision-making [79].

This has relevant implications on the design of serious games that need to consider self-esteem. It appears that in-game choices are not affected by self-esteem, but, interestingly enough, a player's self-evaluation of the performance does correlate with their self-esteem.

The player's pace (M2) also gave us interesting results. We found a statistically significant correlation with self-esteem, but the results were inconsistent in the two phases of the study. It appears that the player's pace might be influenced by self-esteem. Nonetheless, the mechanic we designed does not sufficiently capture it.

The Self-Esteem Implicit Association Test [9] also uses timing and pace to measure self-esteem. This seems to be in agreement with the finding that the player's pace may be influenced by self-esteem. However, based on our results the relation between the two is not entirely clear.

For our third hypothesis, we demonstrated that it is possible to infer player self-esteem by using the concept of Self-Evaluation Bias. When asking players to rate their own performance, their evaluation is susceptible to reflection on their self-esteem. Moreover, it is possible to predict self-esteem based on self-evaluation bias using a simple linear regression model with a moderate error margin. Based on this result, we posit that self-evaluation is a promising game mechanic to be addressed in future games.

Game designers might find more creative and playful ways to let players to evaluate their own game performance than the one we designed, for example request a player to review their activities to "teach" in-game characters based on their performance. By doing so, it is even possible to increase the frequency and the duration of the player measurement, which we suspect that it will increase the accuracy of the predictions.

Study Limitations

The RSES is an explicit personality self-test; it requires respondents to consciously rate their own self-esteem, which may be difficult or bias results. We have not used implicit self-esteem tests, heeding relevant criticisms reported in the literature [11]. Nevertheless, we conjecture that the in-game mechanics that we designed might relate more strongly with the results of implicit self-esteem tests. Future research could consolidate the results found by comparing self-esteem as estimated using self-appraisal of game performance with implicit tests.

Finally, although the study involved two different target groups, a replication study with a larger sample that applies randomized sampling procedures is needed to enable generalizing claims regarding the measurement of self-esteem through games.

Discussion

There have been a few related studies on profiling players [38,39,78,81] which propose modeling techniques and methods through a post-game analysis. In these studies, the game metrics computed were decided prior to, and independently of, the specific traits or skills to be assessed. In contrast, this article has presented two games that were designed and developed specifically to enable profiling a specific trait with game metrics, i.e., our studies show that games can be purposefully designed to profile specific personality traits.

A systematic and methodological approach for designing games able to profile personality traits is fundamental since the chance of failure is high if an ad-hoc approach is used. Hoping to be able to profile specific traits with an ad-hoc approach is not advised mainly because the probability of leaving out essential game metrics for the profiling is high.

We were able to infer NfC by designing a simple strategy game that tested whether or not players' Need for Cognition correlates with players' behavior regarding hints. As hypothesized, we found a negative correlation between following hints and NfC. Further analysis also showed the ability to predict NfC using simple linear regression with moderate error. Similarly, in a second platformer game we were able to demonstrate that a player's self-evaluation of his or her performance could be used as a novel game mechanic that correlates with the player's self-esteem.

Systematic Method to Profile Player Traits

Understanding Game Mechanics as rule-based systems which control how players interact with the game world are a core element in any game design [17,61]. There are generic, known mechanics, which can be found in every game, such as, for example, tutorials, progression charts, levels, scores, rewards, achievements, and so on. Mechanics provide designers with an instrument that addresses a suitable level of abstraction and allow the implementation of in-game metrics that do not compromise the game.

The studies reported in this paper are offered as examples for game designers of how to enhance the value of games while adding an intentional profiling layer. We have presented two game mechanics that potentially lower cognitive requirements in measuring the bias players might exhibit during gameplay.

We recognize that, in order to advance the research topic addressed in this paper, future researchers require a systematic method. We foresee that a systematic method is necessary for at least three aspects: 1) to list existing game mechanics and their links to personality traits, 2) to identify existing game mechanics and their relevance to personality traits, and 3) to design new game mechanics in order to address untapped potential in mapping to personality traits.

For all of the aforementioned aspects, we believe that a platform that will support and advance the collaboration between different stakeholders is necessary. To that end, we have already conceptualized and developed a platform that supports three key stakeholders: game designers, researchers, and players [59]. The GURaaS platform supports game designers by allowing them to expose part of their games for researchers' needs. Researchers can then embed their research instruments into games. Players can consent and provide valuable data by utilizing those research instruments while playing a game.

We envision that our platform will reveal a list of relevant game mechanics and their match to research findings regarding personality traits for both game

designers and researchers alike. In that way, both groups of stakeholders can be informed and able to utilize the first two aspects we listed in the beginning of this section. Furthermore, our platform could, in the future, support the necessary collaborative features of the third aspect, i.e., support a conversation between researchers and game designers to conceptualize and develop new game mechanics for the purpose of personality traits research.

Our contribution defines game mechanics that relate player personality traits to in-game behavior. At a higher level, we contribute to the broader recognition of games as implicit measurement tools for players' psychological traits.

Player Profiles in the Games Industry

The motivation for developing user profiles is well established, at least for productivity-oriented software [35,77]. However, games differ from software that focuses exclusively on productivity, games focus on the user journey and experience as well. Game designers are often portrayed as experience designers, i.e., their role during development of the game is to define, develop, and tune the aspects of the player experience, such as flow [68], engagement [55] and usability [23].

The insights gained from player modelling may lead directly to designing better games, creating sophisticated audience models, and enabling the development of adaptive gameplay. Adaptive Artificial Intelligence [16] is a generic term used in game development that specifies ways and means to change dynamically games based on in-game behavior or trait profiling, personalizing the player experience. For example, adjusting the game to player skill to reduce frustration, increasing playing time by removing repetition, and increasing control and adapting the content to suit different target audiences.

In addition, with recent developments in game telemetry, the games industry is becoming increasingly motivated to collect, study, and develop player models since they lead to real commercial benefits, such as increasing the player base, improving development and design techniques, and developing new monetization models [20].

Profiling can directly support the design process by providing an understanding of how game mechanics can affect and trigger specific behaviors. By closely exploring specific personality traits, we can understand in-game behaviors and explain effects of specific game mechanics, including how they influence motivation and preference [31,53].

Relevance of Profiling Need For Cognition For the Games Industry

The hinting mechanic that we depict in our first study captures how players accept information, which is psychologically relevant and may be directly explored by game designers to improve players' experiences. By using procedural level generation, levels can be personalized for each player. For example,

the levels could present more frequent mental challenges to high-NfC players and, in theory, improve her/his enjoyment.

Furthermore, this information can be used to create customer profiles and support targeted advertisement for in-game content purchases [44]. It is known, for example, that low-NfC players respond better to visually complex stimuli.

Relevance of Profiling Self-Esteem For the Games Industry

In a clear and interesting outcome from the game design point of view, risk evaluation is not linked with self-esteem, i.e., if a target group study demonstrates that players have the tendency to have lower self-esteem, that should not influence the designer to make less challenging or risk-rewarding decisions. Our study demonstrated that players' self-esteem does not influence risk-assessment decisions.

On the other hand, self-esteem does influence the reflective aspect of players. Hence, if reflective or evaluative metrics are performed, they need to be carefully handled because they incur of self-esteem bias. In other words, if a designer is, for example, evaluating the player's experience by questioning the player's results, this might be influenced by player's self-esteem.

Furthermore, profiling players according to their self-esteem can further support the industry by providing an understanding of closely related problems of addiction, aggressive behavior, and anxiety, since they are connected with this trait [37,46].

Player Profiles For Serious Applications

Player profiling for a serious purpose, i.e., for purposes outside the game domain, can potentially offer substantial societal benefits. For example, assessing self-esteem through games can be applied in healthcare and preventive care to support the diagnosis and measurement of symptoms pertaining to mental pathologies at a large scale, e.g., the large prevalence of depression in pregnant women mentioned above [29], suggests a plausible application area for player profiling.

The possibility emerges to design games that could aid patients to recover from specific pathologies, like depression or narcissistic disorders. However, this can only be achieved if experience/game designers have an accurate understanding on how specific mechanics trigger, train, reinforce and explain specific behaviors.

Player profiling can be used in the education domain to provide richer information to educators for explaining behaviors and achievements of students. In a business context player profiling could be applied to assess personality of candidates during intake procedures or of employees. Unstructured selection processes are subject to social influence [5]. Even using structured questionnaires for assessing personality is subject to self-presentation biases that limit the validity of these tests; to reduce such situations [48] suggests the need to

avoid published personality inventories in job selection and the need to find alternatives to self-report personality measures.

Directions

The work in this line of research offers several other interesting directions to explore. We highlight a few in this section.

Games User Research Platform

We believe that integrating player profiling into games, as illustrated with the two cases reported above, can be approached more systematically and even extended to include existing popular games.

A drawback of the two case studies we presented is that a significant amount of resources need to be invested to design and develop such games. Furthermore, for such games to become popular, substantial investment would be needed in marketing and promotion. Thus, the investment in developing such games typically ends with the research studies themselves. Instead of developing games from scratch to investigate certain traits, we envision that a data collection platform specifically designed for researchers to embed their research instruments into existing popular games will be a very substantial addition. For example, such a platform could allow researchers to embed customized questionnaires into a specifically defined game context (e.g., in a certain part of a level) to collect players' opinions in combination with in-game data. While there is a clear benefit for researchers, there might also be added value for both game developers and even players.

Transferability

A direction for future research is to investigate the transferability of the profiling information in configuring and adapting other systems beyond the game itself. For example, in the context of crowdsourcing, which requires crowd workers to complete paid tasks, such systems could filter the type of tasks that are offered in terms of the potential crowd-workers' needs for cognition. More specifically, tasks that require visual recognition might be more suitable for users with low need for cognition. It is also possible to imagine this type of information being carried over into e-learning systems. For example, if an e-learning system already knows that a certain user has low self-esteem, at the very least, it can inform the instructor so that he/she can take that into account, or the system itself can provide more supportive messages to encourage such a user.

Ethics of Player Profiling

Player profiling raises salient questions about the ethics of such a practice, especially if it is used for commercial gain. On the one hand, we recognize the potential and value of player profiling for socially desirable purposes. On the other hand, we can also imagine misuses of implicit player profiling. Further work needs to examine in depth the ethics of player modeling and to develop appropriate guidelines, and eventually policies, to safeguard the public.

Conclusion

We present two studies that investigate the role of specific, yet potentially widely applicable, game mechanics for correlating to and predicting players' personality traits. More specifically, in the first study, we presented *Nanobots*, a strategy game to investigate whether or not players' Need for Cognition (NfC), a general and stable personality trait, correlates with players' behavior regarding hints. As expected, we found a negative correlation between following the displayed hints and NfC. In the second study we presented *Runner*, a platformer game, to investigate which game mechanics are likely to infer self-esteem. The results of this study suggest that player performance is not influenced by self-esteem. Remarkably, players with low self-esteem chose harder paths as often as players with high self-esteem, indicating that self-esteem does not influence the decision process when actively engaging and participating in such activity. We found some influences related to a player's pace in the game, but our data has not been conclusive. However, we were able to demonstrate that a player's self-evaluation of his or her performance could be used as a novel game mechanic that is biased towards self-esteem. Further statistical analysis in both studies showed that the aforementioned game mechanics could predict the respected personality traits.

Our studies imply that game designers, developers, and researchers can directly benefit from the unobtrusive detection of stable personality traits such as NfC and self-esteem. For example, system behavior could be adapted according to detected player's traits. Extending the principle demonstrated in this paper offers significant promise by defining further links between game mechanics and players' traits to not only support the development of better games, but also to enhance the value and utility of such games.

References

- [1] Icek Ajzen. 2005. Attitudes, Personality and Behavior. <http://doi.org/10.1037/e418632008-001>
- [2] Entertainment Software Association. 2015. Essential Facts About The Computer And Video Game Industry. Retrieved from <http://www.theesa.com/wp-content/uploads/2015/04/ESA-Essential-Facts-2015.pdf>
- [3] Sander C.J. Bakkes, Pieter H.M. Spronck, and Giel van Lankveld. 2012. Player behavioural modelling for video games. *Entertainment Computing* 3, 3: 71–79. <http://doi.org/10.1016/j.entcom.2011.12.001>
- [4] Sander C J Bakkes, Pieter H M Spronck, and H. Jaap van den Herik. 2009. Opponent modelling for case-based adaptive game AI. *Entertainment Computing* 1, 1: 27–37.
- [5] Murray R. Barrick, Jonathan A. Shaffer, and Sandra W. DeGrassi. 2009. What you see may not be what you get: Relationships among self-presentation tactics and ratings of interview and job performance. *Journal of Applied Psychology* 94, 6: 1394–1411. <http://doi.org/10.1037/a0016532>
- [6] Annette Bauer, Michael Parsonage, Martin Knapp, Valentina Iemmi, Bayo Adelaja, and Sally Hogg. 2014. The costs of perinatal mental health problems. Centre for Mental Health.
- [7] J. Blascovich and J. Tomaka. 1992. Measures of Personality and Social Psychological Attitudes: Measures of Self-Esteem. *Behavioral & Social Sciences Librarian* 11, 2: 115–160. http://doi.org/10.1300/J103v11n02_05
- [8] Douglas A. Bors, François Vigneau, and Francis Lalonde. 2006. Measuring the need for cognition: Item polarity, dimensionality, and the relation with ability. *Personality and Individual Differences* 40, 4: 819–828. <http://doi.org/10.1016/j.paid.2005.09.007>
- [9] Jennifer K. Bosson, William B. Swann, and James W. Pennebaker. 2000. Stalking the perfect measure of implicit self-esteem: The blind men and the elephant revisited? *Journal of Personality and Social Psychology* 79, 4: 631–643.
- [10] John H. Bradley and Frederic J. Hebert. 1997. The effect of personality type on team performance. *Journal of Management Development* 16, 5: 337–353. <http://doi.org/10.1108/02621719710174525>
- [11] Michael D Buhrmester, Hart Blanton, and William B Swann. 2011. Implicit Self-Esteem: Nature, Measurement, and a New Way Forward. *Journal of Personality and Social Psychology* 100, 2: 365–385.
- [12] J T Cacioppo and R E Petty. 1982. The need for cognition. *Journal of personality and social psychology* 42, 1: 116–131. <http://doi.org/10.1037/0022-3514.42.1.116>
- [13] J T Cacioppo, R E Petty, and C F Kao. 1984. The efficient assessment of need for cognition. *Journal of personality assessment* 48, 3: 306–307. http://doi.org/10.1207/s15327752jpa4803_13

- [14] John T. Cacioppo, Richard E. Petty, Jeffrey A. Feinstein, and W. Blair G. Jarvis. 1996. Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin* 119, 2: 197–253. <http://doi.org/10.1037/0033-2909.119.2.197>
- [15] Fabio Celli, Elia Bruni, and Bruno Lepri. 2014. Automatic Personality and Interaction Style Recognition from Facebook Profile Pictures. In *Proceedings of the ACM International Conference on Multimedia - MM '14*, 1101–1104. <http://doi.org/10.1145/2647868.2654977>
- [16] D Charles, a Kerr, and M McNeill. 2005. Player-centred game design: Player modelling and adaptive digital games. *of the Digital Games* 285, 6: 285–298. <http://doi.org/10.1.1.97.735>
- [17] Daniel Cook. 2006. What are game mechanics? *Lostgarden*, 1–11. Retrieved from <http://www.lostgarden.com/2006/10/what-are-game-mechanics.html>
- [18] R. R. Costa, & McCrae. 1992. Neo PI-R professional manual. *Psychological Assessment Resources* 4, 1: 5–13. <http://doi.org/10.1037/0003-066X.52.5.509>
- [19] Benjamin Cowley and Darryl Charles. 2016. Behavlets: a method for practical player modelling using psychology-based player traits and domain specific features. *User Modelling and User-Adapted Interaction* 26, 2–3: 257–306. <http://doi.org/10.1007/s11257-016-9170-1>
- [20] Anders Drachen and Alessandro Canossa. 2009. Towards gameplay analysis via gameplay metrics. In *Proceedings of the 13th International MindTrek Conference: Everyday Life in the Ubiquitous Era on - MindTrek '09*, 202–209. <http://doi.org/10.1145/1621841.1621878>
- [21] Magy Seif El-Nasr, Anders Drachen, and Alessandro Canossa. 2013. Game Analytics: Maximizing the Value of Player Data. <http://doi.org/10.1007/978-1-4471-4769-5>
- [22] Patrick J. Ewell, Rosanna E. Guadagno, Matthew Jones, and Robert Andrew Dunn. 2016. Good Person or Bad Character? Personality Predictors of Morality and Ethics in Avatar Selection for Video Game Play. *Cyberpsychology, Behavior, and Social Networking* 19, 7: 435–440.
- [23] Alessandro Febretti and Franca Garzotto. 2009. Usability, playability, and long-term engagement in computer games. In *Conference on Human Factors in Computing Systems - Proceedings*, 4063–4068. <http://doi.org/10.1145/1520340.1520618>
- [24] Donald G. Gardner and Jon L. Pierce. 1998. Self-Esteem and Self-Efficacy within the Organizational Context. *Group & Organization Management* 23, 1: 48–70. <http://doi.org/10.1177/1059601198231004>
- [25] Jennifer Golbeck, Cristina Robles, Michon Edmondson, and Karen Turner. 2011. Predicting personality from twitter. In *Proceedings - 2011 IEEE International Conference on Privacy, Security, Risk and Trust and IEEE International Conference on Social Computing, PASSAT/SocialCom 2011*, 149–156. <http://doi.org/10.1109/PASSAT/SocialCom.2011.33>
- [26] Samuel D. Gosling, Peter J. Rentfrow, and William B. Swann. 2003. A very brief measure of the Big-Five personality domains. *Journal of Research in Personality* 37, 6: 504–528. [http://doi.org/10.1016/S0092-6566\(03\)00046-1](http://doi.org/10.1016/S0092-6566(03)00046-1)

- [27] Lindsay T. Graham and Samuel D. Gosling. 2013. Personality Profiles Associated with Different Motivations for Playing World of Warcraft. *CyberPsychology, Behavior & Social Networking* 16, 3: 189–193. <http://doi.org/10.1089/cyber.2012.0090>
- [28] a G Greenwald and S D Farnham. 2000. Using the implicit association test to measure self-esteem and self-concept. *Journal of personality and social psychology* 79, 6: 1022–1038. <http://doi.org/10.1037//0022-3514.79.6.1022>
- [29] Catherine Habel, Nancy Feeley, Barbara Hayton, Linda Bell, and Phyllis Zelkowitz. 2015. Causes of women's postpartum depression symptoms: Men's and women's perceptions. *Midwifery* 31, 7: 728–734. <http://doi.org/10.1016/j.midw.2015.03.007>
- [30] Ramón Alzate Sáez De Heredia, Francisco Laca Arocena, and José Valencia Gárate. 2004. Decision-making patterns, conflict styles, and self-esteem. *Psicothema* 16, 110–116.
- [31] Shih-Ping Jeng and Ching-I Teng. 2008. Personality and Motivations for Playing Online Games. *Social Behavior and Personality: an international journal* 36, 8: 1053–1060. <http://doi.org/10.2224/sbp.2008.36.8.1053>
- [32] Maurits Kaptein, Panos Markopoulos, Boris De Ruyter, and Emile Aarts. 2015. Personalizing persuasive technologies: Explicit and implicit personalization using persuasion profiles. *International Journal of Human Computer Studies* 77: 38–51. <http://doi.org/10.1016/j.ijhcs.2015.01.004>
- [33] Maurits Kaptein, Panos Markopoulos, Boris De Ruyter, and Emile Aarts. 2015. Personalizing persuasive technologies: Explicit and implicit personalization using persuasion profiles. *International Journal of Human Computer Studies* 77: 38–51.
- [34] A Karpinski and J L Hilton. 2001. Attitudes and the Implicit Association Test. *Journal of personality and social psychology* 81, 5: 774–788. <http://doi.org/10.1037//0022-3514.81.5.774>
- [35] Judy Kay. 1995. The um toolkit for cooperative user modelling. *User Modeling and User-Adapted Interaction*. <http://doi.org/10.1007/BF01100243>
- [36] Alan E Kazdin. 2000. *Encyclopedia of psychology*, Vol. 7. (2000) Washington: US; New York, NY, US: American Psychological Assoc.
- [37] Eun Joo Kim, Kee Namkoong, Taeyun Ku, and Se Joo Kim. 2008. The relationship between online game addiction and aggression, self-control and narcissistic personality traits. *European Psychiatry* 23, 3: 212–218. <http://doi.org/10.1016/j.eurpsy.2007.10.010>
- [38] Giel Van Lankveld, Sonny Schreurs, and Pieter Spronck. 2009. Psychologically verified player modelling. In *GAMEON*, 12–19.
- [39] Giel van Lankveld, Pieter Spronck, Jaap van den Herik, and Arnoud Arntz. 2011. Games as personality profiling tools. 2011 IEEE Conference on Computational Intelligence and Games (CIG'11): 197–202. <http://doi.org/10.1109/CIG.2011.6032007>
- [40] Falk Leichsenring and Eric Leibling. 2003. The effectiveness of psychodynamic therapy and cognitive behavior therapy in the treatment of personality disorders: A meta-analysis. *American Journal of Psychiatry* 160, 1223–1232. <http://doi.org/10.1176/appi.ajp.160.7.1223>

- [41] James W. Loewen, Thomas J. Cottle, Morris Rosenberg, and Roberta G. Simmons. 1975. Black and White Self-Esteem: The Urban School Child. *Contemporary Sociology* 4, 281. <http://doi.org/10.2307/2063225>
- [42] Andrés Lucero and Juha Arrasvuori. 2010. PLEX Cards: A Source of Inspiration When Designing for Playfulness. *Fun and Games* 15, 17: 28–37.
- [43] Tobias Mahlmann, Anders Drachen, Julian Togelius, Alessandro Canossa, and Georgios N. Yannakakis. 2010. Predicting player behavior in Tomb Raider: Underworld. In *Proceedings of the 2010 IEEE Conference on Computational Intelligence and Games, CIG2010*, 178–185.
- [44] Brett A.S. Martin, Michael J. Sherrard, and Daniel Wentzel. 2005. The role of sensation seeking and need for cognition on Web-site evaluations: A resource- matching perspective. *Psychology and Marketing* 22, 109–126. <http://doi.org/10.1002/mar.20050>
- [45] Gerald Matthews, Ian J. Deary, and Martha C. Whiteman. 2003. Personality Traits. <http://doi.org/10.1017/CBO9780511812736>
- [46] Mehwash Mehroof and Mark D. Griffiths. 2010. Online Gaming Addiction: The Role of Sensation Seeking, Self-Control, Neuroticism, Aggression, State Anxiety, and Trait Anxiety. *Cyberpsychology, Behavior, and Social Networking* 13, 3: 313–316. <http://doi.org/10.1089/cyber.2009.0229>
- [47] Pejman Mirza-Babaei, Veronica Zammito, Jörg Niesenhaus, Mirweis Sangin, and Lennart Nacke. 2013. Games user research. CHI '13 Extended Abstracts on Human Factors in Computing Systems on - CHI EA '13: 3219. <http://doi.org/10.1145/2468356.2479651>
- [48] Frederick P. Morgeson, Michael A. Champion, Robert L. Dipboye, John R. Hollenbeck, Kevin Murphy, and Neal Schmitt. 2007. Reconsidering the use of personality tests in personnel selection contexts. *Personnel Psychology* 60, 683–729. <http://doi.org/10.1111/j.1744-6570.2007.00089.x>
- [49] Lennart E Nacke, Anders Drachen, Kai Kuikkaniemi, and Yvonne a W De Kort. 2009. Playability and Player Experience Research. *Proceedings of the IEEE*: 1–11.
- [50] Jozef M. Nuttin. 1985. Narcissism beyond Gestalt and awareness: The name letter effect. *European Journal of Social Psychology* 15, 3: 353–361.
- [51] Patrick M O'Malley and Jerald G Bachman. 1979. Self-esteem and education: Sex and cohort comparisons among high school seniors. *Journal of Personality and Social Psychology* 37, 7: 1153–1159. <http://doi.org/10.1037/0022-3514.37.7.1153>
- [52] Gaby Odekerken-Schröder, Kristof De Wulf, and Patrick Schumacher. 2003. Strengthening outcomes of retailer-consumer relationships. The dual impact of relationship marketing tactics and consumer personality. *Journal of Business Research* 56, 177–190. [http://doi.org/10.1016/S0148-2963\(01\)00219-3](http://doi.org/10.1016/S0148-2963(01)00219-3)
- [53] Jowon Park, Yosep Song, and Ching-I Teng. 2011. Exploring the Links Between Personality Traits and Motivations to Play Online Games. *Cyberpsychology, Behavior, and Social Networking* 14, 12: 747–751. <http://doi.org/10.1089/cyber.2010.0502>

- [54] Carlos Pereira Santos, Kevin Hutchinson, Vassilis-Javed Khan, and Panos Markopoulos. 2017. Measuring Self-Esteem with Games. In ACM Intelligent User Interfaces. <http://doi.org/http://dx.doi.org/10.1145/3025171.3025196>
- [55] Andrew K. Przybylski, C. Scott Rigby, and Richard M. Ryan. 2010. A motivational model of video game engagement. *Review of General Psychology* 14, 2: 154–166. <http://doi.org/10.1037/a0019440>
- [56] Richard W. Robins, Holly M. Hendin, and Kali H. Trzesniewski. 2001. Measuring Global Self-Esteem: Construct Validation of a Single-Item Measure and the Rosenberg Self-Esteem Scale. *Personality and Social Psychology Bulletin* 27, 2: 151–161. <http://doi.org/10.1177/0146167201272002>
- [57] M. Rosenberg. 1965. Rosenberg self-esteem scale (RSE). *Acceptance and commitment therapy., Measures package* 61: 52. <http://doi.org/S0034-98872009000600009>
- [58] Carlos Pereira Santos. 2016. Inferring A Player's Need For Cognition From Hints. 3–6.
- [59] Carlos Pereira Santos, Jeroen van de Haterd, Kevin Hutchinson, Vassilis Javed Khan, and Panos Markopoulos. 2017. GURaaS: An end-user platform for embedding research instruments into games. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 34–48. http://doi.org/10.1007/978-3-319-58735-6_3
- [60] Maureen E. Shea and Mary E. Pritchard. 2007. Is self-esteem the primary predictor of disordered eating? *Personality and Individual Differences* 42, 8: 1527–1537.
- [61] Miguel Sicart. 2008. Defining game mechanics. *Game Studies* 8, 2.
- [62] Andrew E. Skodol. 2012. Personality Disorders in DSM-5. *Annual Review of Clinical Psychology* 8, 1: 317–344. <http://doi.org/10.1146/annurev-clinpsy-032511-143131>
- [63] Adam M. Smith, Chris Lewis, and Kenneth Hullett. 2011. An inclusive taxonomy of player modeling.
- [64] SM Stephen M. Smith and Irwin P. IP Levin. 1998. Need for Cognition and Choice Framing Effects. *Journal of Behavioral Decision Making* 9, 4: 283–290. [http://doi.org/10.1002/\(SICI\)1099-0771\(199612\)9:4<3.0.CO;2-7](http://doi.org/10.1002/(SICI)1099-0771(199612)9:4<3.0.CO;2-7)
- [65] Stephen M. Smith and Irwin P. Levin. 1996. Need for cognition and choice framing effects. *Journal of Behavioral Decision Making* 9, 4: 283–290.
- [66] Julia Friederike Sowislo and Ulrich Orth. 2012. Does Low Self-Esteem Predict Depression and Anxiety? A Meta-Analysis of Longitudinal Studies. *Psychological bulletin* 139, 1: 213–240.
- [67] Wayne T. Steward, Tamera R. Schneider, Judith Pizarro, and Peter Salovey. 2003. Need for Cognition Moderates Responses to Framed Smoking-Cessation Messages1. *Journal of Applied Social Psychology* 33, 12: 2439–2464. <http://doi.org/10.1111/j.1559-1816.2003.tb02775.x>
- [68] Penelope Sweetser and Peta Wyeth. 2005. GameFlow: a model for evaluating player enjoyment in games. *Computers in Entertainment (CIE) - Theoretical and Practical Computer Applications in Entertainment* 3, 3: 1–24. <http://doi.org/10.1145/1077246.1077253>

- [69] Christian Thurau and Christian Bauckhage. 2010. Analyzing the Evolution of Social Groups in World of Warcraft®. *Proceedings of the 2010 IEEE Conference on Computational Intelligence and Games*: 170–177. <http://doi.org/10.1109/ITW.2010.5593358>
- [70] Julian Togelius, R De Nardi, and Sm Lucas. 2006. Making racing fun through player modeling and track evolution. *Optimizing Player Satisfaction in Computer and Physical Games*: 61.
- [71] Kali H. Trzesniewski, M. Brent Donnellan, and Richard W. Robins. 2003. Stability of self-esteem across the life span. *Journal of Personality and Social Psychology* 84, 1: 205–220. <http://doi.org/10.1037/0022-3514.84.1.205>
- [72] Anders Tychsen. 2008. Crafting user experience via game metrics analysis. *Goals and Strategies for Studying User Experience*, July: 1–5.
- [73] Annette Vincent and Dianne Ross. 2001. Personalize training: determine learning styles, personality types and multiple intelligences online. *The Learning Organization* 8, 1: 36–43. <http://doi.org/10.1108/09696470110366525>
- [74] P. J. Watson, Tracy Little, Steve M. Sawrie, and Michael D. Bideman. 1992. Measures of the Narcissistic Personality: Complexity of Relationships With Self-Esteem and Empathy. *Journal of Personality Disorders* 6, 4: 434–449.
- [75] James B. Weaver. 1991. Exploring the links between personality and media preferences. *Personality and Individual Differences* 12, 12: 1293–1299. [http://doi.org/10.1016/0191-8869\(91\)90203-N](http://doi.org/10.1016/0191-8869(91)90203-N)
- [76] Pamela Williams-Piehot, Tamera R. Schneider, Judith Pizarro, Linda Mowad, and Peter Salovey. 2003. Matching Health Messages to Information-Processing Styles: Need for Cognition and Mammography Utilization. *Health Communication* 15, 4: 375–392. http://doi.org/10.1207/S15327027HC1504_01
- [77] Wolfgang Wörndl, Alexander Manhardt, and Vivian Prinz. 2010. A Framework for Mobile User Activity Logging. In *Mining Ubiquitous and Social Environments MUSE 2010 Workshop 21st European Conf on Machine Learning 14th European Conf on Principles and Practice of Knowledge Discovery in Databases ECMLPKDD 2010*.
- [78] Narnia C. Worth and Angela S. Book. 2014. Personality and behavior in a massively multiplayer online role-playing game. *Computers in Human Behavior* 38: 322–330.
- [79] Juan Yang, Katarina Dedovic, and Qinglin Zhang. 2010. Self-esteem and risky decision-making: An ERP study. *Neurocase* 16, 6: 512–519. <http://doi.org/10.1080/13554791003785893>
- [80] Georgios N. Yannakakis, Pieter Spronck, Daniele Loiacono, and Elisabeth André. 2013. Player Modeling. *Dagstuhl Follow-Ups* 6, 59. <http://doi.org/10.4230/DFU.Vol6.12191.45>
- [81] Nick Yee and Nicolas Ducheneaut. 2011. Introverted elves & conscientious gnomes: the expression of personality in world of warcraft. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*: 753–762. <http://doi.org/10.1145/1978942.1979052>
- [82] Xijuan Zhang, Ramsha Noor, and Victoria Savalei. 2016. Examining the effect of reverse worded items on the factor structure of the need for cognition scale. *PLoS ONE* 11, 6. <http://doi.org/10.1371/journal.pone.015779>

Chapter 3

Profiling Ethics Orientation through Play

Abstract

Research studies and recruitment processes often rely on psychometric instruments to profile respondents with regards to their ethical orientation. Completing such questionnaires can be tedious and is prone to self-presentation bias. Noting how video games often expose players to complex plots, filled with dilemmas and morally dubious options, the opportunity emerges to evaluate player's moral orientation by analysing their in-game behaviour. In order to explore the feasibility of such an approach we examine how users' moral judgment correlates with choices they make in non-linear narratives, frequently present in video games. An interactive narrative presenting several moral dilemmas was created. An initial user study (N=80) revealed only weak correlations between the users' choices and their ethical inclinations in all ethical scales. However, by training a genetic algorithm on this data set to quantify the influence of each branch on recognizing moral inclination we found a strong positive correlation between choice behaviour and self-reported ethical inclinations on a second independent group of participants (N=20). The contribution of this work is to demonstrate how genetic algorithms can be applied in interactive stories to profile users' ethical stance.

Summary:

Chapter 3 details a case study that uses the methodology presented in the previous chapter and applies it to design a game to profile a multi-dimensional trait, namely ethical norms.

Published in:

Journal Behaviour & Information Technology
Taylor & Francis, Volume 37 - Issue 9, September 2018

Authors:

Carlos Pereira Santos, Vassilis-Javed Khan, Panos Markopoulos

Introduction

Following the financial crisis of 2008 there has been a growing concern surrounding news reports regarding unethical practices in business. Unethical behaviours like stealing, lying or fraud are affecting adversely the economy but also society in general [7,9,22,42]. In a recent analysis of 2,410 cases of occupational fraud reported between 2014 and 2015 it is argued that the studied organizations lost about 5% of their annual revenue as a result of fraud while the total losses resulting from unethical behaviours exceeded \$6.3 billion, with an average loss per case of \$2.7 million [1]. Growing public awareness of unethical behaviour in business has prompted calls for explicit actions to instil ethical values in corporations [6], which makes methods for assessing the moral judgement of employees very relevant.

Moral judgment is the thinking process that determines if an idea, entity or action is right or wrong; but also how it fares in terms of other related concepts such as good, bad, unconventional, bizarre, ethical or irrational, to name a few. The process an individual takes to determine "right" or "wrong" is directly dependent upon the individual's cultural setting, context and the perceived consequences an action might have [27]. People often find themselves in a situation where a difficult choice has to be made between two or more undesirable alternatives, otherwise known as a dilemma. People face ethical dilemmas in their professional capacity, e.g., end of life judgments for medical doctors [25], in major choices pertaining to family and social life, but even in mundane daily activities where people break social norms or laws; e.g., cutting in line to make it for lunch, or speeding while driving to catch a flight.

Different ethical theories provide guidance for resolving moral judgments based on diverse and sometimes conflicting values. Representative theories include: Moral equity, Contractualism, Relativism, Egoism and Utilitarianism [27]. The conflict between these values underlies classic dilemmas (e.g., Heinz's dilemma or the trolley problem, [21]), and is manifested in numerous entertainment products like movies, music, books, and recently even games.

Games as profiling tools

Games are a medium promoting very extensive and diverse player interactions, which encourages game companies to use telemetry to monitor player's in-game behaviour. Analysing such data can provide clear and detailed information about what players do within the game, which can serve marketing purposes or the development of more engaging games. In the last few years this practice has rapidly expanded and matured across the gaming industry, which is increasingly recognizing the value of understanding player behaviour [10].

The term "Game Metric" is defined as data-driven information that is collected with games. Game metrics are interpretable measures of in-game interactions

by the players [46], e.g., at what times do people play a game, how much time or money they spend. Game metrics often serve predictive business analyses pertaining to topics such as the expected sales figures or the expected number of players. Other examples of game metrics concern in-game player behaviour that can be valuable during game design, authoring and content creation, e.g., in-game position, frequency of used items, accuracy and skill set, which allow game designers to know more about the players, their behaviour, proficiency and experience.

Player Modelling [46] pertains to developing means to measure, interpret and represent player behaviour [29]. Recently, there has been an effort to assess personality profiles based on in-game behaviour, like the Big Five Personality Traits [23,24,47], HEXACO [45], Need for Cognition [32] and Self-esteem [31]. In these studies videogames are seen as a sandbox that encourages role-play and that allows players to make choices and take actions while avoiding their real-life consequences, in a way that reflects their individual personality.

This research aims to contribute to Player Modelling literature by showing how in-game choices relate to a player's moral judgment. Specifically, this study addresses the following research questions:

- Is players' moral judgment related to their in-game behaviour?
- How can an interactive narrative be designed to allow implicit profiling of players' ethical orientation?

The contribution of this paper is twofold. First, it pertains to application of machine learning algorithms to relate users' choices in interactive stories with their moral judgment. We show that choices players make during interactive narratives can serve as an implicit measure of one's ethical judgment. This can be useful as an alternative to psychometric questionnaires, with some advantages over those, such as being more fun and engaging, and potentially leaving less room for self-presentation bias when compared to explicit self-report.

Multi-Dimensional Ethics Scale

Most profiling ethics methods require participants to self-report opinions of specific situations. Perhaps the most used psychometric instrument for understanding ethical norms is the Multidimensional Ethics Scale (MES) which was originally developed by Reidenbach and Robin [34] and then further refined by Reidenbach et al. [36].

MES represents the evaluative criteria that individuals use in making a moral judgment based on five independent ethical norms: moral equity, deontology, relativism, egoism and utilitarianism. Scores along these scales represent a 'moral compass' of respondents. Since its publication, numerous studies have used MES to evaluate all five or a subset of the five ethical norms, demonstrating a strong degree of validity and a high degree of reliability. MES has been applied

in a wide range of contexts, including evaluating marketing ethics [37], business ethics [15] and questionable Internet practices [44].

In the MES questionnaire a participant evaluates a set of vignettes according to several semantic scales, i.e. a short story is presented to the participant that explains the setting of a dilemma or an ethical issue and followed by an action explaining how the situation was resolved. The participant is then requested to evaluate the action by classifying (in a Likert scale) multiple semantic statements that are associated with the different ethical norms, like just-unjust, fair-unfair, does-doesn't violate an unspoken promise, etc. [34] (see Figure 3-1).

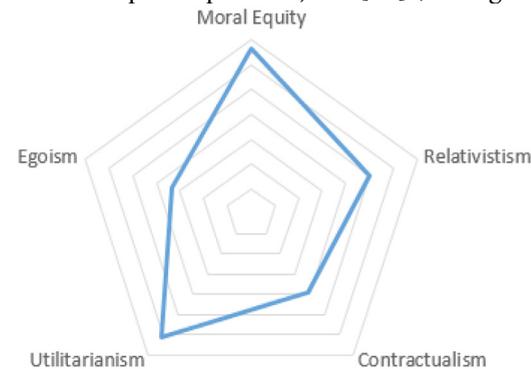


Figure 3-1: Graphical example of the output of MES questionnaire for all five norms. The graph indicates moral judgment being influenced largely by Moral Equity and Utilitarianism and much less by Egoism.

Ethical Norms

A complete overview of all ethical norms associated with Moral Philosophy is outside the scope of this paper; readers can find more information in relevant textbooks [27]. It is quite common in research to address only a subset of specific and distinct ethical theories rather than the complete MES inventory [4,12,18,19]. Accordingly and to keep the size of the study manageable, we demonstrate the feasibility of our approach by considering only three out of the five ethical norms as including all five norms would increase considerably the writing effort and the logistics of experiments. Specifically, we addressed the following norms: Contractualism, Utilitarianism and Moral Equity. A brief discussion of all ethical norms is provided as background information for this study.

Moral Equity

Much of the most influential concepts of moral equity (also known as Justice) derive from the writings of Aristotle. The Principle of Formal Justice states: "Treat similar cases similarly and different cases differently in the proportion to their difference(s)" [5]. It is considered formal because it specifies the form that matters in justice but still does not specify what counts as alike and unlike,

leaving that for a substantive principle of justice [5]. The most important concept in this theory is related to the equal distribution of value and to the develop rules or procedures that result in fair or just outcomes. Moral equity can be subdivided into procedural and distributive justice.

Procedural justice relates to the fairness in processes, and it is divided into: *i)* pure, if the rules guarantee an objectively just outcome in every occurrence, *ii)* perfect, if the rules provide an individual with a fair result in every case, and *iii)* imperfect procedural justice, if the rules represent the best attempt to produce fair results, but sometimes the outcomes are unjust.

Distributive justice is a contemporary ethical theory focused on the uniform distribution of values and/or goods [35]. An action is ethical if it provides equal wealth, rewards, opportunities, obligations and/or burdens by all involved members, unless there is a morally significant and justifiable difference between them.

Contractualism

Contractualism also known as Deontology, stems from the Greek word *deon* (δέον), which means "duty", pertaining directly to an individual's duty towards others. In other words, it is the duty of an individual to act according to genuine and legitimate claims of others, even if not enforced by any authority [38]. This ethical theory defines, for example, the sense of responsibility to pay our debts or care for family. Actions are the main drivers of duty ethics, which teaches that some acts are right or wrong, and individuals have a duty to act accordingly, regardless of the good or bad consequences that may be produced.

Emmanuel Kant is the father of the "Categorical Imperative", which uses the notion of maxims to define moral acts. Maxims are universal rules that define what ethical actions are. Simple examples are: it is wrong to kill innocent people, it is wrong to steal, or it is right to keep promises [20]. As an extreme though illustrative example, Kant argued that it would be wrong to tell a lie in order to save a friend from a murderer. This ethical theory is linked with the individual's rights, by enforcing duties and codes of honour for individuals to follow. A general understanding of maxims comes through political, social or religious grounds like constitutions, military codes or sacred texts.

Utilitarianism

Utilitarianism states that individuals should act to produce the greatest possible good for all of society. It requires the decision maker to consider all possible outcomes of their actions/inactions and weigh which results in the best ratio of good versus bad outcomes for society [28]. Effectiveness and productivity are also very much related to this theory, since a less efficient action is likely to produce less utility and would therefore be qualified as less ethical. Also, concepts like democracy where the opinion of the majority rules, demand and supply or capitalism are linked to utilitarianism theories [28,38].

An important notion to understand is the fact that according to utilitarianism, actions that may cause harm to individuals or minorities can be defined as ethical if they can be compensated with a minor gain to a majority. An extreme example would be to refuse medical care to someone during rush hour because it will create a traffic jam for thousands of people. In addition, the definition of greater good can be broken down into multiple parts like: social, economic and ecological. Historically those aspects often clash, and it is in practice difficult to define and quantify ethical action based on imperfect information and the inability to measure or compare outcomes [38].

Ethical Relativism

Relativists believe that ethical values are dependent on one's culture, feelings, or religion. Thus a decision may be deemed ethical in one's own culture, while the same behaviour under the same circumstances may be judged unethical in another culture. An individual has to be careful and distinguish what people believe is right or wrong, and there can be no objective basis for a moral judgment as every culture devises its own set of moral standards, and therefore, no universal ethical rules exist that apply to all [30].

Anthropologists have reported diverse and contradictory ethical values in different cultures, which is the main evidence and the justification for relativism. There is some criticism over relativism, pertaining to it not serving the main task of ethics: the preservation of conditions that allow people to pursue a happy and stable life [17]. Several managers cite relativism as a justification for alleged unethical behaviour, especially in multi-cultural and international situations [17].

Egoism

Egoism is a teleological (consequentialist) theory that states decisions should reflect solely on one's best long-term interests, i.e., an ethical action should be performed if it produces a greater ratio of good versus evil for the individual in the long run. In short, it means actions that result in the greatest good for the individual. Some variations of this theory focus on a short-term hedonism, stating that everyone is psychologically programmed to behave only in their own self-interest [33].

Modern philosophers consider egoism seriously since the personal long-term interests create a suitable context, which motivates individuals to help others, as well as help to define and follow the rules of society. There are some critical positions against egoism, namely that it disregards dogmatic principles, like what most people commonly agree is right and wrong. In addition, and similar to relativism it does not serve the purpose of ethical philosophy [26].

Games and Ethics

Earlier work looking directly into games and ethics is sparse. Zagal [48] reviews several video games notable for their moral dilemmas and discusses ethical settings that are presented to players that challenge ethical norms. Zagal states that although games portray ethical codes similar to reality, it is possible that a game might define its own cultural settings and encode a new ethical system which the player is required to learn and embrace in order to succeed. Zagal distinguishes the following categories of ethical perspectives on games: Artefact (e.g. narrative), Business (e.g. development and production), Play (e.g. sportsmanship and cheating) and Gameplay (e.g. in-game behaviour actions). Zagal concludes that moral judgment can be used to elicit and create emotionally meaningful experiences but the full potential for ethical reasoning and reflection in games has not yet been explored.

Several scholars reflect on moral disengagement in video games [13,16,40]. While such critique and concerns are well founded, it is also important to note that a deeper understanding on how a player's moral judgment relates to in-game choices is still missing, while it can precisely help measure the influence of games on players' ethics.

An attempt to profile players is reported in [11], where a specific game mechanic (avatar building) is used to predict moral deviation. This work suggests that avatars represent not only their players' psychological traits, but also correlate with the moral disengagement scale which assesses the propensity to overlook or ignore one's own rules of moral conduct. Compared to that work, we are particularly interested in measuring players' alignment with ethical norms rather than just their tendency to apply them in their in-game behaviour.

Method

We set out to answer two research questions: *i)* "Is players' moral judgment related to their in-game behaviour?" and, *ii)* "How can an interactive narrative be designed to allow the implicit profiling of players' ethical orientation?" For this, we adapted a branching narrative and examined players' choices within the story and how their choices compare with their responses to the Multidimensional Ethics Scale (MES).

We compared the MES scores for 80 participants with the choices they made within the branching narrative by calculating the correlation between the two scales.

We purposely used just a branching narrative, stripped down of any other known game mechanics. With this setup we could prevent strong influences like character self-representation or rewards that might affect the players' drive towards a certain branch of the narrative and in that way bias the study.

Branching narratives are considered as non-linear gameplay, which emphasizes on the player's choices as a central part of the design of the game. Using Crawford's game definition and taxonomy [8], the branching narrative is classified as an entertainment interactive challenge or puzzle (see Figure 3-2).

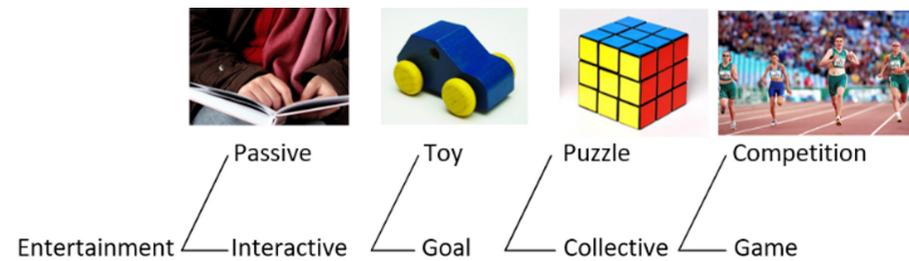


Figure 3-2: Illustration of Crawford classification of entertainment products. Non-linear narratives are interactive, goal oriented but not collective, therefore considered a puzzle.

Branching Narrative

The branching narrative that we developed was inspired by the movie *Interstellar* (2014). Our story is based on a sci-fi setting and places the reader in command of a spaceship where he or she must manage with limited resources and a small crew in a last-ditch attempt to find a solution to a plagued and threatened Earth. The story is written in English, capturing the main plot of the movie and featuring similar characters. The narrative is around 10.000 words (including all branches), and our data shows that native English speakers take 30.5 minutes on average to complete a single read through (SD=9.1). The story was split in multiple small segments (from 10 to 250 words each) to be presented separately to readers and which provide distinct narrative sections. At the end of a section a way forward is presented with one or more links to following sections (see Figure 3-3).

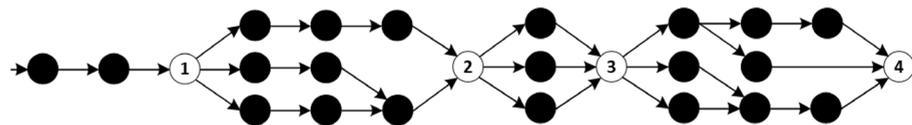


Figure 3-3: Visual illustration of part of the structure of the branching narrative. Every node represents a text section, and the arrow the choices available. The numbered nodes are sections where the choices are linked with ethical behaviors.

Singular links in a section provide a linear progression, while multiple links provide readers with branching choices in the narrative. A reader is asked to commit to a certain choice and is not allowed to revise his or her decision, i.e., when given multiple choices, a reader can only choose one specific path, and

there is no option to go back and modify a choice once it has been made. Across the narrative there are 15 moments that all readers go through independently of their previous choices and where they need to make a choice (illustrated as numbered nodes in Figure 3-3). In these specific 15 moments there are always three choices presented to them, each based on one of three aforementioned ethical norms.

To support the data collection a web-based platform has been developed, which allowed participants to read the story online and make their choices. The order of choices available to players is randomly presented in order to avoid placement bias. In-game metrics record all players' decisions. For the purposes of this study only the choices in these 15 moments matter. An example of a page of the narrative on one of those 15 moments with three options is depicted in Figure 3-4.

Figure 3-4 shows an example of a branching narrative page. The page displays a text block with a decision point. The text reads: "Now that I brought you here, our work just started". Below this, it says "You stated. We need to:" followed by three bullet points: "- Establish a energy link between the between the aircraft and Dr. Wilson base.", "- Inform Earth about the mission status and Dr. Wilson's findings.", and "- Give Dr. Wilson some of the best meals we have, he deserves it." Below this, it says "Our first priority should be:". Then, under "Decision:", there are three radio button options: "Dr. Wilson base is running on fumes and we might lose crucial information", "The mission is crucial, lets report back to the command center.", and "It is only fair that Dr. Wilson regains some strength.". At the bottom, there is a "Choose" button.

Figure 3-4: Example of a branching narrative page, where the reader has to make a choice of the path to take. We varied randomly the order of possible choices for each participant.

Participants

To reach a more diverse participant group than the typical university participant pools [2], a crowdsourcing service (prolific.ac) was used and a small monetary reward (£2,50) was provided for completing the study. Detailed participant demographic profiles were provided by the crowdsourcing service, and all participants were informed of the fact that data was being collected for research purposes and that all logged data was anonymous.

All our participants were native English speakers, and the average age of our participants was 35.6 (SD=11.4), with 54% being female, almost all were Caucasian, but the sample showed good diversity on religion and political

affiliations. Even though most of the participants belong to the Occidental culture, it is not expected that this study would be influenced by cultural differences; the plot reflects a futuristic environment and a global scale problem, avoiding references to countries or strong cultural ties. In addition, the collected data is measured against a validated model (MES), which has been used in multiple target audiences with diverse religions and cultural backgrounds.

Process

A typical participant followed the following steps:

1. Receive an invite to play an online game
2. Open a URL to start the branching narrative
3. Complete a demographic questionnaire
4. Complete the MES questionnaire
5. Read the branching narrative, and chose an option for each of the branches (see Figure 3-4)
6. Receive the results of the MES questionnaire and their choices in the branching narrative (Figure 5).

To avoid learning effects and self-presentation bias, we only considered participants' first attempt at the game. Furthermore, we excluded participants that had previously seen the movie that our story was based on; a specific question was included in our demographic questionnaire for that purpose. Our last exclusion parameter related to how long people took to read the story, all outliers were also excluded, i.e., players that read the story under 12 minutes and above 49 minutes.

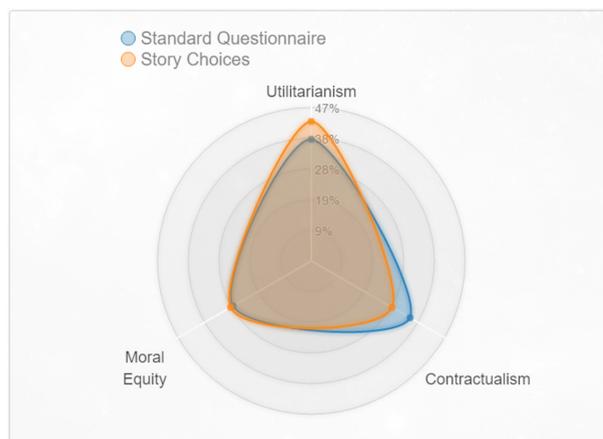


Figure 3-5: Results showing the ethical profile to the user after completing a read through; (blue) MES Questionnaire, (orange) story choices.

Results

After applying the aforementioned exclusion criteria, we retained 55 participants (out of 80). For the 55, the data was more or less evenly distributed among the different ethical theories in both the MES scores and the followed branches in the narrative (Figure 3-6). We note a slight preference of our sample for utilitarianism and contractualism over moral equity; this pattern is noticeable both in the MES scores and in the branching narrative choices.

In the following paragraphs we explore the relation between the choices made in the branching narrative and the players' MES scores.

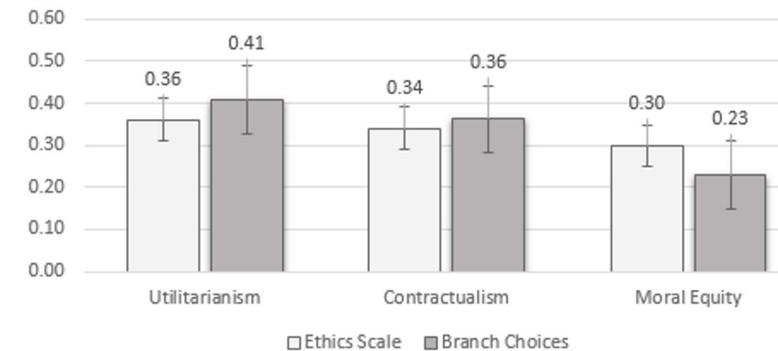


Figure 3-6: Overall comparison between the MES score and the narrative branching choices. The graph shows spread of the different choices performed by players for all branches ($N=825$) and the standard error.

Raw scores

As a first approach we correlated all branch choices to the MES scores. As an example to clarify the calculations; player p starts by filling out the MES questionnaire, from which we obtain an independent score for each of the different ethical norms $MES_{p,e}$ where e is the ethical norm (i.e. Moral Equity, Contractualism and Utilitarianism). $MES_{p,e}$ ranges from 0 to 48; it is calculated by adding a score ranging from 0 to 6 for two different attributes per norm for each of the four vignettes.

While the player reads we record all choices p made throughout the 15 branches (Figure 3). $B_{p,n=e}$ represents player p 's ethical norm choice on branch n . After p completes reading the story, we calculate the branch score $BS_{p,e}$ (1), which indicates how frequently e has been selected.

$$BS_{p,e} = \frac{\sum_{n=1}^{15} \begin{cases} B_{p,n} = e & 1 \\ \text{else} & 0 \end{cases}}{15} \quad (1)$$

In this analysis we calculate Spearman's correlation coefficient independently for every ethical norm e by comparing all the pairs ($MES_{p,e}$, $BS_{p,e}$) for all players. This analysis does not provide any evidence of a correlation between the choices made in the story and the MES scores (Table 3-13).

Story vs. MES (Raw)	N	R	p
Moral Equity	55	-0.004	0.971
Contractualism	55	0.039	0.776
Utilitarianism	55	0.241	0.075

Table 3-13: Direct correlations between story choices and MES

Machine Learning

Since a direct correlation was not obvious we decided to apply a machine learning approach [49]. The approach we followed uses a genetic algorithm commonly used in feature construction and optimization algorithms [3,14] to identify how much each branch contributes to the scores for each of the ethical norms. Using the branch choices from the 80 participants we calculated a weight $W_{n,e}$ for each branch n which optimized the inference of the ethical orientation e . With these weights we wanted to see if they could be used to correlate with other participants' ethical orientation.

The genetic algorithm rules are straightforward: the weight range is between 0.0 and 2.0, and the sum of all the weights needs to be 15.0 (1.0 per branch). The algorithm uses elitist selection (best solutions are stored from multiple generations), and the fitness function is the Spearman's statistical correlation. We aim to maximize the correlation (r) while minimizing the significance (p).

The W_e (set of $W_{n,e}$ for all branches) were generated by: *i*) randomly generating the full set of weights, *ii*) mutating a single weight by assigning a new random value for a single branch and *iii*) by cross-over where new sets are generated by cross-breeding existing sets. The algorithm computed multiple generations of W_e , and convergence was achieved after 8 to 10 generations.

$$BS_{p,e} = \frac{\sum_{n=1}^{15} \begin{cases} B_{p,n} = e & W_{n,e} \\ else & 0 \end{cases}}{15} \quad (2)$$

Validation with a new data set

We recruited a new cohort of participants ($N=20$) from Prolific.ac. These 20 participants showed similarities with our original group; their average age was 36.9 ($SD=11.2$) and 50% were female. They followed the same process and similarly we excluded three participants applying the same exclusion criteria as

before. The difference with this group pertained solely to how we compute the branch score, by applying the $W_{n,e}$ scores that were calculated with the data collected from the first group of participants.

The results (see Table 3-14) show strong positive correlations between the new set of readers' narrative choices and their MES scores, allowing us to conclude that there is a relation between players' in-game choices and their moral judgment, when appropriate weights are applied.

Story vs. MES (MLearn)	N	R	p
Moral Equity	17	0.511	0.036 ^a
Contractualism	17	0.613	0.009 ^a
Utilitarianism	17	0.631	0.007 ^a

^a Significant result

Table 3-14: Strong positive correlations were found between the weighted branch choices and MES. This means that tracking user choices in interactive stories could help predict ethical inclinations.

Discussion

A strong narrative intertwined with gameplay often helps create compelling ethical dilemmas. Game designers put a lot of effort in creating an authentic interactive story and an engaging experience. On the other hand, it is also common for designers to use dilemmas within the narrative to create emotional bonds with the player's decisions and give players the feeling that they are actually in the starring role with meaningful consequences. The use of ethical theories within game narrative design is not entirely new, as some designers actively study them to apply them within game development [40]. However, our study has taken a different approach to examine if and how the players' moral judgment is manifested in their gameplay. By uncovering the correlation between in-game player choices with their moral judgment we show *i*) that players' moral judgment is related to their in-game behaviour, and *ii*) how to design an interactive narrative support the implicit profiling of players' ethical orientation.

Profiling moral judgment through play

Our results support the hypothesis that player in-game choices are influenced by their moral judgments. Initial analysis of the raw data from all recorded choice behaviours did not reveal any correlation to their moral judgment scores obtained through MES. However, using a machine learning approach to weigh

the contribution of each node of the interactive narrative, we could obtain a score that describes the player's observed moral choice behaviour which strongly correlates to the MES score. With this work we contribute to the literature by showing how an interactive story can provide a measure that correlates to self-reported ethical judgments and how the application of a machine learning approach which designers of interactive narratives can apply to profile their players' moral judgments.

While this result was obtained in the context of an interactive narrative, we conjecture that it could be possible to compute a similar correlate to players' ethical judgments with any game (also of a different genre) that places the player in a position which he or she has to choose between two or more alternatives that correspond to different ethical choices as a way to profile players' moral judgment, like: survival, role-playing, construction, management, stealth, life simulation or war games. To do so successfully it is important to map actions and choices that are available to the player on ethical norms. Our results also show that not all branches are effective identifiers for ethical norms. In the next section we discuss in more detail these differences.

Designing effective Choice/Actions identifiers

A careful look at the narrative of high-weighted nodes shows that unambiguous contexts, referring to known terms that are associated with norms are better able to evoke choice behaviour that relates to players' ethical norms. In the example depicted in Figure 4 (B13), it is clear that the first option is linked with utilitarianism and represents consequences ("might lose"), while the second option is linked with contractualism ("mission"), and finally the third is linked with moral equity ("fair").

On the other hand, a complex introduction or less clear choice statements can lead to different moral interpretations. For example, in B5 readers have to decide in which planet to land first and a long story plot is revealed with multiple side effects. In addition the choices presented to the user may reflect different norms: *i)* 'What if he needs our help?', *ii)* 'We should conserve fuel' and *iii)* 'We will do what the majority wants'. These could all be taken to represent a more utilitarian view rather than alignment to different ethical norms.

Therefore, simple and clear text should be used exposing the situation and lending clarity to options and consequences. Furthermore, the design challenge is in that the presented choices need to be directly associated with the ethical norm while at the same time they should provide a meaningful outcome for the story.

Designing Ethical Games

Games, like other entertainment products, use ethical dilemmas and break moral codes as a way to enhance user experience. Due to interactivity and role play, players are confronted by dilemmas embedded within the games, which exposes players to situations like corruption, theft, violence or gender discrimination just to name a few morally debatable game experiences, but, can also serve as a canvas to explore loyalty, companionship and compassion [16,39,40,48].

Rather than just a medium for conveying negative influences, a game can have a positive contribution to the ethical development of individuals and companies. In this sense, our research also opens up the opportunity to use moral judgment as a game metric. By extension, game developers can use player models [46] to measure players' moral judgment in real time and alter the game to increase player engagement and enhance their play experience. This research directly contributes to player profiling by uncovering a mechanic that can be easily replicated in other interactive stories.

Opportunities

There is a clear opportunity for applying player profiling techniques to other applied fields. As described in the introduction, there are multiple domains in society, business, and healthcare whether measuring individual ethical orientation be useful. For example, there are specific studies in the financial and business sectors which show that directly ethical attitude/orientation is important to maintain ethical behaviour [41] and professional commitment [43], and recommend to include the ethical profiling as part of the Human Resource Management.

Despite the study being very promising and our conscious effort to recruit participants with diverse demographics (including gender, age and political affiliation) the generalizability of our results for different populations and cultural contexts would still need to be established by a replication study that involves a larger and more varied sample of participants. In addition, MES has two dimensions which have not been covered in the present study (Egoism and Relativism). For completeness future studies should address those two norms as well.

Although it is out of the scope of this study, working with game telemetry and ethics raises itself salient ethical questions about profiling players, especially for commercial purposes. There can be great societal value derived from profiling ethical orientation, which in turn calls for deliberating on ethics of player modelling as such, and for taking measures to inform and safeguard the public (e.g., to make people aware of profiling taking place and its purpose, to decide when it is morally justifiable to carry out such profiling with or without player cooperation). Guidelines for the ethical use of player modelling, informed consent procedures, and safeguards for the privacy of players and compliance with the evolving privacy legislation still need to be developed.

Conclusions

In this work, we have examined whether one's choices in an interactive story are correlated with one's moral orientation; distinguishing between Moral Equity, Contractualism and Utilitarianism. For this purpose we developed a branching interactive story, inspired by a science fiction movie, which presents players with moral dilemmas. We compared player choice behaviour within the interactive story with player's ethical orientation as measured by a questionnaire that gauges one's ethical inclination (the Multidimensional Ethics Scale). Based on that data we then applied a genetic algorithm to calculate weights to the different story's branches. Our analysis of experimental data resulted in a weighted model of player choices which is correlated with their ethical inclination and found promising results. We can therefore conclude that the identifying one's ethical inclinations through the choices they make in interactive stories is a promising direction as an alternative to questionnaires. This work adds to a growing body of research in using games to profile individuals through observing their in-game behaviour, contributing towards the broader recognition of games as an implicit measurement tool for people's attitudes and traits.

References

- [1] Association of Certified Fraud Examiners. 2016. Report To the Nations on Occupational Fraud and Abuse. *Global Fraud Study*: 1–92.
- [2] Tara S Behrend, David J Sharek, Adam W Meade, and Eric N Wiebe. 2011. The viability of crowdsourcing for survey research. *Behavior research methods* 43, October 2015: 800–813.
- [3] Robert R. Bies, Matthew F. Muldoon, Bruce G. Pollock, Steven Manuck, Gwenn Smith, and Mark E. Sale. 2006. A genetic algorithm-based, hybrid machine learning approach to model selection. *Journal of Pharmacokinetics and Pharmacodynamics* 33, 2: 195–221.
- [4] H Buchan. 2014. Reidenbach and Robin's Multidimensional Ethics Scale: Testing a Second-Order Factor Model. *Psychology Research* 4, 10: 823–834.
- [5] Craig L. Carr. 1981. The concept of formal justice. *Philosophical Studies* 39, 3: 211–226.
- [6] J. Owen Cherrington and David J. Cherrington. 1992. A menu of moral issues: One week in the life of The Wall Street Journal. *Journal of Business Ethics* 11, 4: 255–265.
- [7] Jeffrey Cohen, Yuan Ding, Cedric Lesage, and H. Stolowy. 2010. Corporate Fraud and Managers' Behavior: Evidence from the Press. *Journal of Business Ethics* 95, SUPPL. 2: 271–315.
- [8] Chris Crawford. 2008. The Art Game Design. *Epilepsy Research* 49, Clinical Trial PT-Journal Article PT-Multicenter Study PT: 512.
- [9] Nature Editorial. 2006. Ethics and fraud. *Nature* 439, 7073: 117–118.

- [10] Magy Seif El-Nasr, Anders Drachen, and Alessandro Canossa. 2013. *Game Analytics: Maximizing the Value of Player Data*.
- [11] Patrick J. Ewell, Rosanna E. Guadagno, Matthew Jones, and Robert Andrew Dunn. 2016. Good Person or Bad Character? Personality Predictors of Morality and Ethics in Avatar Selection for Video Game Play. *Cyberpsychology, Behavior, and Social Networking* 19, 7: 435–440.
- [12] Steven M Flory, Thomas J Phillips, R. Eric Reidenbach, and Donald P. Robin. 1992. A multidimensional analysis of selected ethical issues in accounting. *The accounting review* 67, 2: 284–302.
- [13] A. Gabbiadini, P. Riva, L. Andrighetto, C. Volpato, and Brad J. Bushman. 2014. Interactive effect of moral disengagement and violent video games on self-control, cheating, and aggression. *Social Psychological and Personality Science* 5, 4: 451–458.
- [14] David Edward Goldberg. 1989. *Genetic Algorithms in Search, Optimization and Machine Learning*.
- [15] Seema Gupta. 2010. A Multidimensional Ethics Scale for Indian Managers' Moral Decision Making. *Electronic Journal of Business Ethics and Organization Studies* 15, 1: 5–14.
- [16] Tilo Hartmann and Peter Vorderer. 2010. It's okay to shoot a character: Moral disengagement in violent video games. *Journal of Communication* 60, 1: 94–119.
- [17] W. Michael Hoffman, Robert E. Frederick, and Mark S. Schwartz. 2014. *Business Ethics. Readings and cases in corporate morality*. Wiley-Blackwell.
- [18] Neil Humphreys, Donald P Robin, R Eric Reidenbach, and Donald L Moak. 1993. the Ethical Decision Making Process of Small Business Owner/Managers and Their Customers. *Journal of Small Business Management* 31, 3: 9–22.
- [19] Insung Jung. 2009. Ethical judgments and behaviors: Applying a multidimensional ethics scale to measuring ICT ethics of college students. *Computers and Education* 53, 3: 940–949.
- [20] Immanuel Kant. 2000. *Foundations of the Metaphysics of Morals*. *World Philosophers & Their Works JN - World Philosophers & Their Works*: 1–5.
- [21] L. Kohlberg. 1981. *Essays in Moral Development*. In *the Philosophy of Moral Development*.
- [22] KPMG. 2013. *KPMG Malaysia Fraud, Bribery and Corruption Survey 2013*. KPMG forensic: 1–72.
- [23] Giel Van Lankveld, Sonny Schreurs, and Pieter Spronck. 2009. Psychologically verified player modelling. In *GAMEON*, 12–19.
- [24] Giel van Lankveld, Pieter Spronck, Jaap van den Herik, and Arnoud Arntz. 2011. Games as personality profiling tools. *2011 IEEE Conference on Computational Intelligence and Games (CIG'11)*: 197–202.
- [25] Liz Lloyd. 2004. Mortality and morality: ageing and the ethics of care. *Ageing and Society*.

- [26] Justin G. Longenecker, Joseph A. McKinney, and Carlos W. Moore. 1988. Egoism and independence: Entrepreneurial ethics. *Organizational Dynamics* 16, 3: 64–72.
- [27] Roselie McDevitt, Catherine Giapponi, and Cheryl Tromley. 2007. A model of ethical decision making: The integration of process and content. *Journal of Business Ethics* 73, 2: 219–229.
- [28] Tim Mulgan. 2007. *Understanding Utilitarianism*.
- [29] Lennart E Nacke, Anders Drachen, Kai Kuikkaniemi, and Yvonne a W De Kort. 2009. Playability and Player Experience Research. *Proceedings of the IEEE*: 1–11.
- [30] Edward Newman and M.C. McFarland. 1990. Comments, with reply, on “Urgency of ethical standards intensifies in computer community” (Standards). *Computer* 23, 6: 6.
- [31] Carlos Pereira Santos, Kevin Hutchinson, Vassilis-Javed Khan, and Panos Markopoulos. 2017. Measuring Self-Esteem with Games. In *ACM Intelligent User Interfaces*.
- [32] Carlos Pereira Santos, Vassilis-Javed Khan, and Panos Markopoulos. 2016. Inferring A Player’s Need For Cognition From Hints. In *Proceedings of the 21st International Conference on Intelligent User Interfaces (IUI ’16)*, 76–79.
- [33] Ayn Rand. 1964. *The Virtue of Selfishness: A New Concept of Egoism*. New American Library.
- [34] D. M. Randall and A. M. Gibson. 1990. Methodology in business ethics research: A review and critical assessment. *Journal of Business Ethics* 9, 6: 457–471.
- [35] John Rawls. 1999. *A theory of justice*. Harvard University Press - Belknap.
- [36] R. E. Reidenbach and D. P. Robin. 1990. Toward the development of a multidimensional scale for improving evaluations of Business Ethics. *Journal of Business Ethics* 9, 8: 639–653.
- [37] R. Eric Reidenbach, Donald P. Robin, and Lyndon Dawson. 1991. An application and extension of a multidimensional ethics scale to selected marketing practices and marketing groups. *Journal of the Academy of Marketing Science* 19, 2: 83–92.
- [38] Thomas M Scanlon. 1982. Contractualism and Utilitarianism. In *Utilitarianism and Beyond*, Amartya Kumar Sen and Bernard Arthur Owen Williams (eds.). Cambridge University Press, 103–110.
- [39] Marcus Schulzke. 2009. Moral decision making in fallout. *Game Studies* 9, 2.
- [40] Miguel Sicart. 2009. The banality of simulated evil: Designing ethical gameplay. *Ethics and Information Technology* 11, 3: 191–202.
- [41] Viruli Silva, H Opatha, and Aruna Gamage. 2018. Does Ethical Orientation of HRM Impact on Employee Ethical Attitude and Does Ethical Orientation of HRM Impact on Employee Ethical Attitude and Behavior? Evidence from Sri Lankan Commercial Banks. *International Business Research* 11, 1.

- [42] Bahram Soltani. 2014. The Anatomy of Corporate Fraud: A Comparative Analysis of High Profile American and European Corporate Scandals. *Journal of Business Ethics* 120, 2: 251–274.
- [43] Metin Uyar and Gökhan Özer. 2011. The ethical orientation and professional commitment : An empirical examination on Turkish accountants. 5, 23: 10023–10037.
- [44] Stan Williamson, Kenneth E. Clow, Bruce C. Walker, and T. Selwyn Ellis. 2011. Ethical Issues in the Age of the Internet: A Study of Students’ Perceptions Using the Multidimensional Ethics Scale. *Journal of Internet Commerce* 10, February 2015: 128–143.
- [45] Narnia C. Worth and Angela S. Book. 2014. Personality and behavior in a massively multiplayer online role-playing game. *Computers in Human Behavior* 38: 322–330.
- [46] Georgios N. Yannakakis, Pieter Spronck, Daniele Loiacono, and Elisabeth André. 2013. Player Modeling. *Dagstuhl Follow-Ups* 6, 59.
- [47] Nick Yee, Nicolas Ducheneaut, Les Nelson, and Peter Likarish. 2011. Introverted Elves & Conscientious Gnomes: The Expression of Personality in World of Warcraft. *Proceedings of the Annual Conference on Human Factors in Computing Systems - CHI ’11*: 753–762.
- [48] José P Zagal. 2009. Ethically Notable Videogames: Moral Dilemmas and Gameplay. *Proceedings of DiGRA 2009*: 9.
- [49] Tom M. Mitchell, 1997, *Machine Learning*, McGraw-Hill Science/Engineering/Math, ISBN: 0070428077

Chapter 4

GURaaS: An End-User Platform for Embedding Research Instruments into Games

Abstract

In this paper, we detail a software platform that enables game developers to expose aspects of their games to researchers who are not necessarily familiar with game development, providing them the possibility to customize game content for behavioral user research, and more specifically to embed survey items in a game context. With this platform we introduce the concept of Games User Research as a Service (GURaaS). This article describes the process we followed to design GURaaS, its high level architecture and its application in a case study. We envision that GURaaS will assist researchers and organizations by helping them expand their reach in finding participants and in collecting survey data reducing the tedium for survey participants.

Summary:

The previous chapters establish that personality traits reflect on players' behavior. Still, a question remains on how to operationalize a platform which would allow to support researchers better understanding players. Chapter 4 examines the design process of an in-game behaviour data acquisition platform while providing an alternative business model for game developers.

Published in:

6th International Symposium on End User Development (IS-EUD 2017), the Netherlands

Authors:

Carlos Pereira Santos, Jeroen van de Haterd, Kevin Hutchinson, Vassilis-Javed Khan, Panos Markopoulos

Introduction

There is a growing number of research studies in utilizing games for a research purpose; well-known examples include quantifying malaria parasites [13], creating accurate protein structure models [10], or tagging images to support the semantic web [1].

As in all aforementioned cases, the general approach researchers follow is to design a purpose-specific game, which are point-solutions that are not aimed to address broader research needs. Although such an approach can be very effective, it does require a considerable investment by behavioral researchers in game development which is typically neither their expertise nor their interest. In addition, this process does not easily scale up: first, because only well-resourced projects and ventures will be able to develop such games, and second, this approach fails to make good use of existing games that could be adapted for research purposes.

In practice, such an approach is restrictive in terms of resource requirements: a research team would either need to have a game development background or be able to outsource this task. It is obvious that not every research group would have the people or the budget required in order to develop such a game.

In addition, designing a game that would be fun to play and become popular is an extra challenge. We argue that for a variety of purposes scalability could be achieved if researchers are allowed to add their content (research instruments) into existing games. To realize that vision, a platform that interfaces existing games with researchers is required, that enables them to re-use game software and facilitates them in adapting it for their research purpose.

Although there is a clear benefit for the researchers, one may question the added value for game developers of such a platform. Game developers are already using game telemetry techniques to improve their own development procedures, like remote debugging, target audience analysis and revenue tracking. However, the rapid development cycles of casual games combined with low development budgets create a complex development context, such that game developers are always looking for easy integration tools to support their work.

An easy to use, well documented platform which allows developers to automate the remote logging and to conduct proper Games User Research [16] to assess user experiences would be an added value for game developers [18]. In addition, it can be used as an alternative revenue stream for game developers by allowing their games to become a research platform, in a fashion similar to how Apple ResearchKit^[1] is used and the way it exposes mobile data to researchers.

Our vision is to provide Games User Research as a Service (GURaaS). Specifically we aim to support this vision through a cloud-based platform that

¹ <http://www.apple.com/lae/researchkit/>

uses games as a dissemination tool, and collects large amounts of data by logging in-game player behavior, and which empowers researchers (including game researchers) to integrate and conduct studies among the player population. The GURaaS platform was designed and developed based on a stakeholder-centered analysis of the needs and requirements of each stakeholder (players, researchers and game developers). In this paper, we present in detail the findings of our process; the multi-tier high-level architecture; and how the online platform works. We conclude by explaining the setup of a case study and the next steps for future research.

GURaaS Design

At a conceptual level there are already several frameworks for developing serious games [29]. Such frameworks identify on an abstract level which components are important in a serious game. However, what is currently lacking is a practical implementation of such frameworks in the form of a platform that helps to match games with research needs. The platform have two main features: (F1) an in-game recording system and, (F2) a set of customizable player research activities.

F1: In-game recording system

The platform contains a game and player behavior data recording system, which allows developers and researchers to collect and analyze detailed player data. This feature is common in today's game development processes to support game designers to use data-driven design and researchers to use quantitative and statistical research methods to analyze players' in-game behavior.

This data is also valuable for the player; a Player Dossier is a data-driven reporting tool which allow players to track, analyze and share their own in-game performance [14]. Player Dossiers are frequently available and are normally available on supporting online platforms linked to a specific game.

F2: Player research activities

This feature allows researchers to dynamically add the option to embed in the games explicit activity requests to players. In detail, a research activity is defined as a task that researchers can request to collect evidence from players that is relevant for their study. Such a request can be as simple as answering a Yes/No question.

Although, researchers could potentially have access to detailed player data (F1), it is difficult to understand the player's experience based solely on the player in-game data, and so a more explicit protocol would be required to obtain more information from the players.

Feature Analysis.

Both features F1 and F2 are valuable for both researchers and developers. The In-game recording (F1) is common practice in today's game development practices, but the Player research activities is pioneering by allowing researchers and practitioners to customize the collection of evidence. The combination of both features can help obtain complementary enriched data sets, which allows developers and researchers to gain a deeper understanding of the player. In the sections below we report the findings of our design approach, where we analyze in detail the system features by studying each actor's main motivation and needs, namely: *i) Players; ii) Developers; and iii) Researchers.*

Players

Although players don't have an active role on deciding the game content, they have an active role in selecting the games they are playing, so it is essential to understand player motivations and how those may clash with the platform features (F1 and F2).

The in-game recordings (F1) has the essential role of facilitating the logging of player interaction, but for players this is completely transparent and therefore there is no need for a detailed analysis. On the other hand, the players are not used to engaging in research activities (F2), therefore we need to consider whether are player motivations are affected.

There is substantial literature that documents the video-game industry's focus in the players' engagement with their products; engagement is what addresses players' motivations and translates into notable user experiences which keep players involved with the product, brand or sequels [9,23,26].

A lesson to be drawn from the past, especially from serious games which do not solely focus on entertaining players and which may consequently fail to do so sufficiently, is to maintain player motivation. Charsky [5] clearly describes and justifies that the failure to address player motivation is the main factor that differentiates between Serious Games, Simulations and Edutainment; if players do not have an extrinsic motivation to engage with simulations and edutainment products, then those products are fated to fail. When the simulations and edutainment provide intrinsic motivations they become serious games.

For the purpose of our platform, it is essential to understanding player motivations and to assure the motivation is not deflected by the introduction of the games user research elements.

Games

Games reflect a fast rule based simulation, easy to achieve, which is a simplification of life. There are multiple theories that may explain why people play games e.g. the Maslow's Need Hierarchy Theory [8]. Although entertainment and games are definitely linked, there is a difference between

both; Crawford [6] defines games by subdividing entertainment into multiple categories (illustrated in Figure 4-1).

Entertainment can be an inert activity, like reading a book or watching TV, but also may involve interactivity. If an interactive entertainment product does not have a goal it is considered a toy, otherwise it is considered a challenge. Interactive challenges can be defined as individual (puzzles) or collective. A collective interactive challenge that does not allow people to react/confront each other is considered a competition (like a race), while an interactive entertainment challenge that allows conflict between participants defines a game.

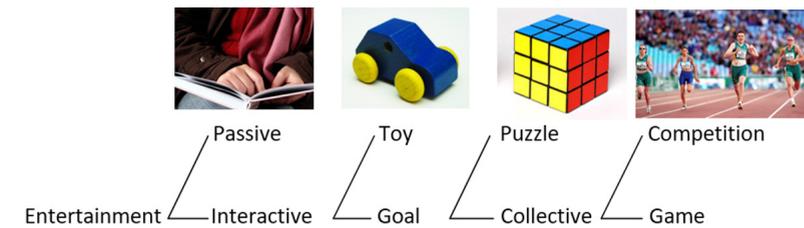


Figure 4-1: Illustration of Crawford classification of entertainment products.

Video-games can be clearly defined as interactive entertainment products which use game mechanics to challenge players and provide immediate feedback on their performance, fitting Crawford taxonomy [6]. Today's video games are quite adept in keeping the player's focus in small tangible tasks, and provide rewards when those tasks are completed successfully (e.g. new level, experience points, in-game currency, etc.). Such game mechanics support different types of fun to keep the players motivated [25].

Fun

Understanding what attracts players and supplying them the right keys/types of fun is extremely relevant. Lazzaro [12] describes the four keys to fun, which are able to identify game genres but also, to identify player motivation:

- Easy fun: curiosity from exploration, role play, and creativity;
- Hard fun: challenge, winning, achieving of a difficult goal;
- People fun: amusement from social interaction, competition and cooperation;
- Serious fun: excitement from educating or changing the player and their world.

Normally game designers, determine their target audience and identify their main motivations. By utilizing those fun keys while designing a game, they encourage increased player engagement.

Thoughtlessly adding research activities into game environments may disrupt the gameplay, meaning that using and disrupting in-game player behavior may

confront the player and reduce the entertaining value (fun). Still, some games provide intermittent external stimuli to players by introducing advertisements in their game context, in the form of interactive audio-video and graphic representations, which require the player's action to either engage or dismiss the advertisement (see Figure 4-2).

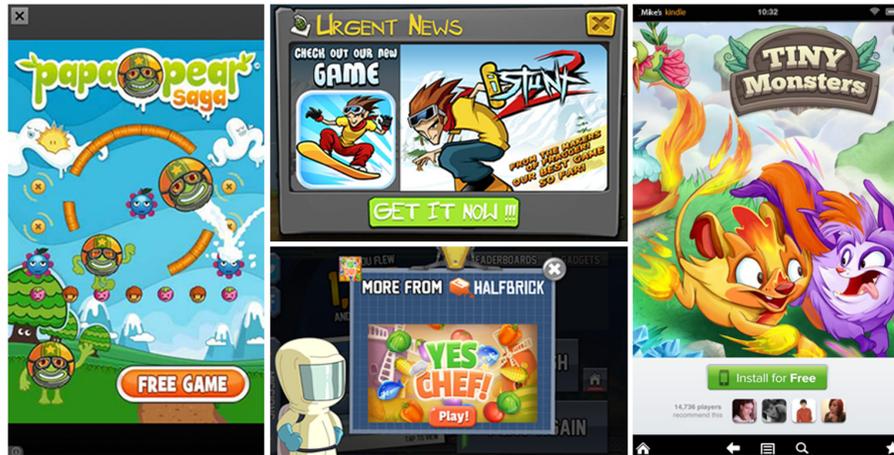


Figure 4-2: Examples of in-game advertisements.

Free to play games

A study with over two thousand developers and players took a closer look into in-game advertisements [30]. The data showed that in-game advertisements in mobile games are popular with today's developer community. The reason is simple: the number of players who are actually willing to pay for the games is rather low. Thus it leaves game studios with a single monetizing option: that of advertisements. The same study makes clear that players understand the need of in-game advertisement. In fact, 54% of the players selected rewarded videos as their preferred way to pay for mobile in-game content. Rewarded videos are advertisements in video format that reward the player after viewing them, by providing some sort of virtual goods (e.g. virtual coins).

Player rewards

Known game design patterns use rewards as a positive feedback for players [4]. Most player rewards are purely virtual game goods, like access to new in-game content, but other rewards like prizes, merchandising, services or even the players own data (player dossier), have been used to reward players.

Furthermore, player rewards have been used to stimulate and/or persuade player behavior in real life; for example Berkovsky [3] uses in-game virtual goods to stimulate real physical player activity. Another example was described above, where the acceptance and satisfaction of player towards in-game advertisement can be improved by providing in-game rewards [30].

Rewarding Research Activities

Presenting unsolicited research activities to players, might cause some distrust and annoyance. In addition, the nature of research activities requires researchers to request specific and sometimes personal questions like personality traits, and performance metrics that might create privacy concerns. To increase players' compliance it is important to (optionally) offer rewards like in-game content (e.g. currency, new levels, outfits, perks). Note that a balance between the activity (time, effort and interference) and the reward is important; this might mean the need to phase the rewards for example through a point system.

Game Developers

For the context of this report we define game developers as institutions or single developers that contribute to the development and publication of a game.

In today's video-game context there are three major target platforms (PC, Console and Mobile) and all have multiple well known online stores that allow game developers to self-publish their games. This led to a shift in the industry that allows more independent groups and companies to self-production. Although the market is large it is dominated by the online stores and their ranking. The large majority of the published games have to fight for market share in order for the game to be profitable [20].

In terms of tools, developers already make use of a large and diverse set: game engines, modeling software, bug tracking, source control systems, management, documentation, analytics and/or publishing. Most of those decentralized online systems, which allow different people to take the responsibility over different aspects of the development process.

Additionally, developers are accustomed to integrating external libraries in their products like analytic tools, in-game advertisements or links to social network systems [15].

Game Developer Motivations

For professional developers games are a business which needs to be profitable. The investment in the integration of a platform in the game needs to be compensated by incentivizing the developer. Therefore for developers, they need to see benefits of using it which can be measured in multiple ways, like: development effort or time reduction, increased user base, better product or larger profits.

A tangible benefit for game developers is In-game recording (F1) which allows remote tracking, especially during the final stages of the development cycle where the pressure to launch the game increases and player experience needs to be measured accurately. Added value can be provided to developers if extra features, like a remote logging system, is present which directly supports the game development and which is similar in scope and implementation procedures.

Creating a space for using the same platform that is used by the game developers and allowing external parties to conduct research using the player base, might open new revenue streams which is really important in the current market setting.

Games User Research

Game User Research (GUR) is a research area that focuses on methodologies to measure and interpret player behavior. Game companies are increasingly using Games User Research in their development process since it has been shown to be extremely valuable to measure user experiences [16,17]. Although there are already known methodologies that are applied regularly, each company has their own GUR data collection and analysis tools. For new companies that want to start this process, it is a large investment to learn these research techniques, develop an internal tool and only then be able to apply them [18]. It is clear that at this point there is a need for a proper player behavior collection (F1) and user experience measurement tool (F2).

Researcher

Many researchers around the world have difficulties gathering research participants to perform all sorts of experiments and pilots for studies that require human subjects. Crowdsourcing platforms have recently emerged as attractive alternative to recruit participants for the purposes of collecting data and they have proven to be as good or better than other recruiting methods like sampling university students [2]. In addition, crowdsourcing can provide a more ethnically diverse and more work-experienced sample.

Although crowdsourcing platforms seem to be efficient data collection tool, they are viewed as labor portals which are not clear in terms of ethical guidelines and labor regulations in some regions [7]. For example, collecting data from children using crowdsourcing may be unethical or even illegal. In other cases, attaching a reward to participation in the study may conflict with the research goals and be incompatible with the sampling strategy of the researcher.

This work does not propose to replace Crowdsourcing platforms but it rather provides an alternative which allows researchers to reach specific target groups, for example by focusing on groups that play specific game genres, but also are able to have more control over the tasks of the participants, by allowing them to directly embed studies within the games and gather data to perform statistical analyses.

Research activities (F2) are requested to be performed by a set of selected and curated players depending on their profile, and past research activities reducing problems with bias and quality of the result, they also can be flagged for untrusted wordy results and being avoided in future studies.

Researcher's Goal

We believe a researcher's motivation for using a tool like this is to have access to research participants. Games allow them to reach specific target populations and have control over the task the player is required to perform. This can be complemented by the ability to access in-game data (through F1), but also, query players about their experiences and motivations (through F2).

Moreover, we believe that this platform will create a good opportunity for games to be used as personality measuring tools. Player Modeling is a research field that analyses a player's in-game behavior and attempt to produce models that explain or predict the player's behavior.

The main goal is to use the player models to improve gameplay or develop business models of games [28]. As of late, player modeling as also been used to profile player personalities like the Big Five Personality Traits [11], HEXACO [27], Need for Cognition [22] and Self-Esteem [21]. Through games personality traits tests can be made without potential biases, such as awareness of the study objectives, because players are having fun.

Enabling Research through Games

The platform allows researchers to embed research instruments directly within games and in that way be able to collect self-report or behavioral data form players. These two features will allow researchers to target specific target audiences, and complement the results with in-game player behavior data gathered by the system.

Architecture

In this section we describe the architecture that focuses on gathering in-game player behavior in large scale.

To consider a platform that is able to intake a constant influx of in-game log data, for multiple games and each might have thousands of simultaneous players, we have to consider a highly available and highly scalable cloud-based system.

We choose to prioritize service availability and resilience, in detriment to data loss and consistency. It is most important to not impact the game's performance, even if a few logs are lost, and note that, the selection of the data that is sent is directly dependent on the bindings the game developers add to the game.

Bellow (see Figure 4-3) we describe in detail the main components of the platform:

- Game User Research Kit;
- Game User Research Server; and
- Game User Research Portal.

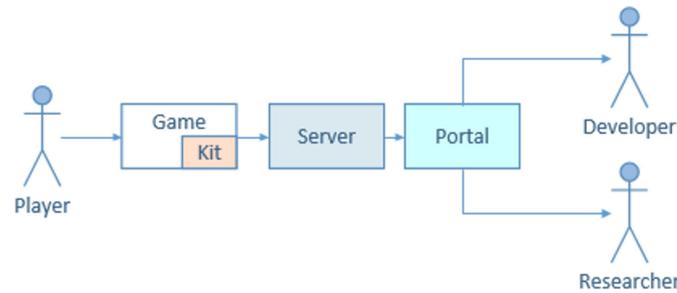


Figure 4-3: Main architecture components depicting a software kit which will be embedded in the game, and which sends game information to a server, which is then visualized through a portal for both developers and researchers alike.

Game User Research Kit

The Game User Research Kit aims to be a compact extension library that will be incorporated into the main game built. This would work as a plugin system directly available from the game engines' asset store.

The Game User Research Kit is focused on the following principles:

- **Performance:** such a library needs to have minimal impact on the game performance, and used resources that may interfere with the game experience (e.g. bandwidth, memory, CPU).
- **Adjustable:** most games try to innovate; the player's in-game behavior logs need to be flexible and tied to custom data schemas.
- **Availability:** ready to be used within most game engines, use off-the-shelf or be easy to develop, and also supports software debugging and testing processes.
- **Well-documented:** support the developers' learning curve, by providing simple and clear instructions with code samples, tutorials and/or online support.

Embedding Game User Research Kit in a game

Like described above, developers have access to a programming library which provides access to most GURaaS required functionality.

For the in-game recording system (F1) it requires the direct involvement of game developers in the sense that each game it is unique and requires different elements to be tracked/logged; it is up to the game developer to explicitly identify the relevant elements that require to be recorded. The library takes the responsibility to competently and with minimum performance loss to store this information locally and push it through the online API.

Customizable player research activities (F2) are simplified by requiring from developers to only identify the space (canvas) where the research activities are

going to be displayed. GURaaS takes the responsibility to obtain, display, collect and store player responses for the research activities.

Since the Game User Research Kit is automatically integrated with the game engine environment, it is able to publish for several target platforms (PC, HTML5, Mobile and even Console). In addition, the server API is public and documented, allowing developers to build their own custom libraries if required, for example, create a library in JavaScript and make use of the open API which allows to register events directly from webpages, or even use iframes to directly display research activities that are made available by the Game User Research Server.

Game User Research Server

The system we are designing should potentially handle large amounts of data. To support a highly scalable and available system the micro-services design pattern was used to support a scalability [19]. We envision that the system should be able to grow and evolve by adding new services as user needs shift. On the other hand, we also need to understand the main complications of using micro-services: the system will have dozens of different services, therefore the information and overall system state is distributed, meaning it requires a set of tools to support its indexing and management of the resources.

All external APIs made available follow RESTful protocols over HTTP to encourage third party entities to create their own libraries and sub-systems to connect to it. Since there are no sensitive data –such as medical records– involved, security is not a priority, but, there are secure protocols to improve and keep safe sensitive data if required in the future. The most relevant services of the server are presented in Figure 4-4. The main data entry point is the Game Recording Gateway, and since it uses HTTP, the overall system availability can be maintained by the number of deployed instances and an effective load balancer.

The Game Recording Gateway service maintains a temporary copy of the delivered records and periodically pushes the records into a Message Bus. The Message Bus will allow an easy and scalable way to share records amongst specialized Profiling Services. When required to integrate new Profile Services, they will only require to be linked with the Message Bus, and without requiring to alter any of the existing systems.

Game and Player Profiling Services perform parallel and specialized actions over the same data, some examples would be: store player game sessions, analyze log issues, create indexed heat maps, or create player models. Although the Profiling Services may vary in nature, they essentially have a common goal: process in-game data and provide outcomes through an external API.

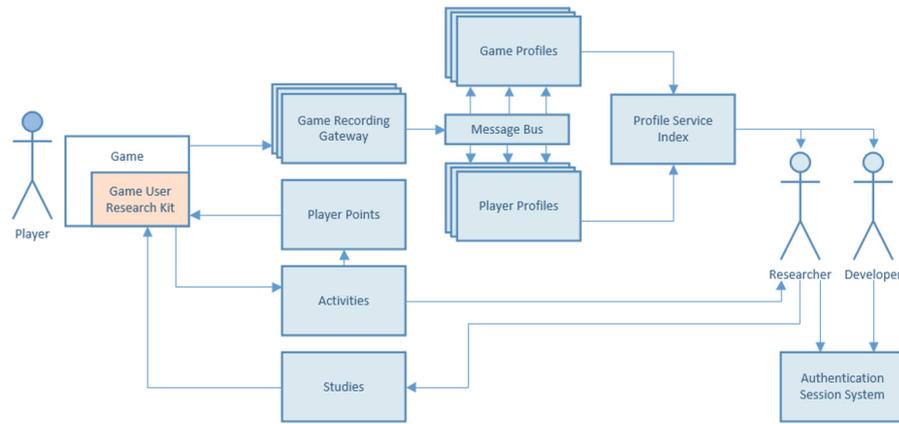


Figure 4-4: Detailed chart of the most relevant services and communication channels of the Game User Research Server.

We highlight that not all data need to be processed by all services. Services will selectively and through an accumulative process enrich their knowledge base and make information available to external entities. Moreover, services can filter and provide data information to other services, hence creating more compelling and complex information systems.

A simple example of a profiling service is a heat-map, which accumulatively populates 2D level information about specific game player events (movement, deaths, or missed jumps). The only in-game records that are processed are the ones containing relevant information for heat-maps, the others are simply discarded by the service.

The Game User Research Kit allows Developers to embed Research Activities directly in their games, like depicted in Figure 4-4. Those studies can optionally award a set of points/coins that the player can then trade for in-game content. The services Player Points, Activities and Studies are responsible for those features that are available to researchers through the Game User Research Portal.

Game User Research Portal

The portal (see Figure 4-5) acts as the interface to both Researchers and Developers to access the GUR profiled game content, like: most played games and levels, gameplay, user and target audience profiling.

Most of the data will be readily available to be imported into appropriate software such as SPSS or R. Even though a raw format is provided, specific profiling services might have an intermediary service to translate into a HTML or another readable format for most users.

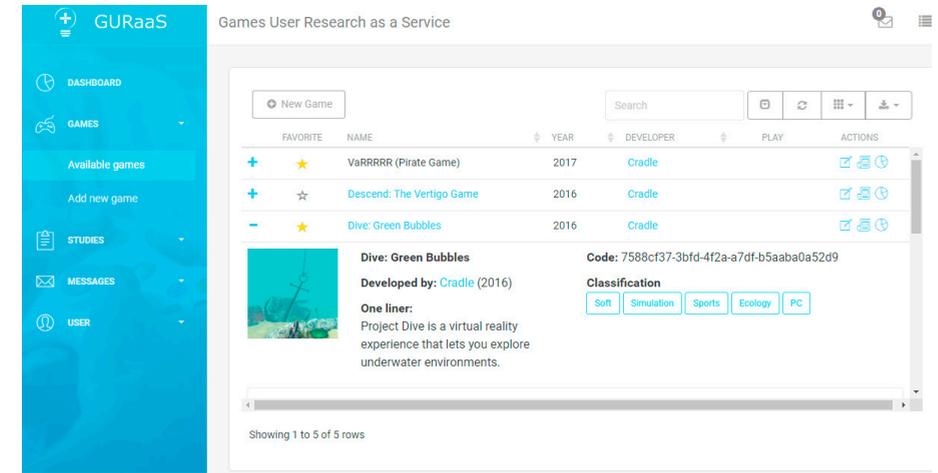


Figure 4-5: Game User Research portal, depicting the wizard screen where researchers can define a study. In this case, the Dive game is displayed. Researchers can choose this game to embed their research instruments (e.g. a questionnaire).

Case Study

Runner (Figure 4-6) is a game designed and developed to be a data-gathering tool that was and will be used for multiple research projects. Runner is a platformer game where the goal is to reach the end of the level as fast as possible. The only actions the player can take are moving left, right, jumping and grabbing (also known as wall jumping). The game is controlled with the keyboard keys: W, A and D, or the arrow keys and Space, which are common for platformer games.

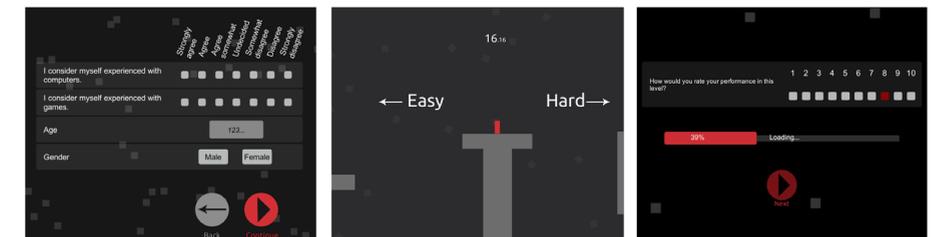


Figure 4-6: Screen captures of Runner, a platform game using the GURAAS platform which logs player behavior and includes dynamic questionnaires. (Left) Dynamic questionnaire which requests player demographics before the game. (Center) Game screen where the player is faced with a choice of paths. (Right) A post-game questionnaire is shown while the next level loads.

The game was purposely designed to be a barren 2D platformer with negligible narrative, which only exposes players to the core game mechanics.

This means aspects that could affect the player's interaction with these core mechanics, like plot, character-self representation and graphics, were avoided or kept to a minimum. An example of this abstraction is the player's avatar: the simple red rectangle that can be seen in the center image of Figure 4-6.

By avoiding the influence of these aspects the chance that results of studies performed on the game are specific to this game are minimized. Additionally, by not relying on specific narratives or graphics in the core gameplay these elements could also easily be replaced if required for a future experiment. However, it must be noted that by avoiding these influences certain other assumptions are made. An experimenter should therefore always look for the most suitable game and a game shall not be suitable for all types of experiments.

The game has two tutorial levels in which the controls are explained and the player can practice the basic game mechanics. After the tutorials the game has eight levels of increasing difficulty. Overall the game is tuned to have a high difficulty, as games that require dexterity and hand-eye coordination often do well catering to Lazarro's Hard Fun category [12].

The screenshot displays the GURaaS Portal interface for configuring a questionnaire. It is divided into three sections:

- Likert Scale:** A table with 7 items ranging from 'Strongly Agree' (Value: 1) to 'Strongly Disagree' (Value: 7). A 'New scale value' button is at the bottom.
- Open Question:** A question 'Age' with an 'Answer number of Lines' set to 1.
- Single Choice:** A question 'Gender' with two options: 'Male' and 'Female'. An 'Add new option' button is at the bottom.

Figure 4-7: GURaaS Portal demonstration of how a researcher can create or modify an in-game research task. In this case, it is demonstrating a section of the interface which is defines the in-game questionnaire which can be seen at Figure 4-6 (left).

Using GURaaS in Runner

Runner was developed in Unity and published in HTML5 to be easily available and accessible. The Game User Research Kit was embedded into the development environment allowing the developers to have access to methods that could benefit from GURaaS features.

Since the specific game was developed to have a high degree of difficulty, the developers mainly collected information about player key performance indicators, namely: *i)* time of the first successful attempt to reach the end of a level, *ii)* the total time a player spent in a level, including all the failed attempts, and *iii)* the number of attempts required to complete a level.

Other variables pertaining the player behavior were also recorded namely: *i)* player pace, and *ii)* path choices.

The developers also created space for researchers to introduced research activities, namely: *i)* when the game starts, *ii)* when the game ends, *iii)* before a level starts, and *iv)* after a level ends.

Using Runner to study Self-Esteem

The first study that was performed on Runner aimed to test whether self-esteem correlates with player behavior in games. Results showed a positive correlation between self-esteem and the players' post-level self-evaluation [21].

Researchers were able to define research activities to be performed by the players through the GURaaS portal. Specifically, when the game starts the players answer a simple demographic questionnaire followed by the Rosenberg Self-Esteem Scale [24]. In addition, after a player successfully completes a level, he/she answers also a Self-Evaluation questionnaire while the next level is loading (see Figure 4-7). Although developers defined two other spaces for research activities, they were not used in this study.

All player data related to in-game behavior, performance and answers to the research activities, were accessible through the GURaaS portal and easily downloadable and then analyzed by 3rd party statistic tools, where all recorded variables were studied, namely: questionnaire answers, performance indicators and player behavior.

A future study is planned that examines the correlation between game difficulty and user experience. Gameplay data will be used to determine a player's performance in a level and the results thereof can be compared to a questionnaire that evaluates user experience after the level. A similar analysis can be done for the overall game by looking at a player's overall performance and user experience.

Conclusion

This paper has introduced the concept of Game User Research as a Service, and demonstrated its feasibility by the development of the GURaaS platform which allows researchers to easily implement games for a research purpose. This paper has described the software architecture of GURaaS that we envision will align researchers' and game developers' agendas. Researchers wish to utilize games for their studies more and more often. Game developers look for new revenue streams and tools to improve the experience of their players. The proposed platform helps both parties by opening the wealth of player data and willingness to play already successful games to a much broader group.

The paper has described a case of using GURaaS to develop a game that measures player's self-esteem by logging their in-game behavior. This case was conducted with the involvement of the developers of GURaaS. Future studies, will examine how third parties may be able to develop games with a research purpose without the direct involvement of GURaaS developers. The platform is finalizing its development and it is in the testing phase, after which it will be released publicly.

Future studies will focus on improving and measuring the usability of the platform from the stakeholder's point of view. We hope this becomes an open platform that will enable both game and researchers to cooperate in knowledge development, but also to bring those two communities closer together.

References

- [1] Luis von Ahn and Laura Dabbish. 2004. Labeling images with a computer game. In ACM Conference on Human Factors in Computing Systems, 319–326.
- [2] Tara S Behrend, David J Sharek, Adam W Meade, and Eric N Wiebe. 2011. The viability of crowdsourcing for survey research. *Behavior research methods* 43, October 2015: 800–813.
- [3] Shlomo Berkovsky, Mac Coombe, Jill Freyne, Dipak Bhandari, and Nilufar Baghaei. 2010. Physical Activity Motivating Games : Virtual Rewards for Real Activity. *CHI 2010: Games and Players*: 243–252.
- [4] Staffan Björk and Jussi Holopainen. 2005. *Patterns in Game Design*.
- [5] D. Charsky. 2010. From Edutainment to Serious Games: A Change in the Use of Game Characteristics. *Games and Culture* 5, 177–198.
- [6] Chris Crawford. 2008. The Art Game Design. *Epilepsy Research* 49, Clinical Trial PT-Journal Article PT-Multicenter Study PT: 512.
- [7] Julie McDonough Dolmaya. 2011. The Ethics of Crowdsourcing. *Linguistica Antverpiensia*, 10: 97–110.
- [8] Abraham H. Maslow. 1970. *A Theory of Human Motivation*.

- [9] Bobby Hoffman and Louis Nadelson. 2010. Motivational engagement and video gaming: a mixed methods study. *Educational Technology Research and Development* 58, 245–270.
- [10] F. Khatib, S. Cooper, M. D. Tyka, K. Xu, I. Makedon, Z. Popovic, D. Baker, and F. Players. 2011. From the Cover: Algorithm discovery by protein folding game players. *Proceedings of the National Academy of Sciences* 108, 18949–18953.
- [11] Giel van Lankveld, Pieter Spronck, Jaap van den Herik, and Arnoud Arntz. 2011. Games as personality profiling tools. 2011 IEEE Conference on Computational Intelligence and Games (CIG'11): 197–202.
- [12] N. Lazzaro. 2004. *Why We Play Games: Four Keys to More Emotion Without Story*.
- [13] Miguel Angel Luengo-Oroz, Asier Arranz, and John Freen. 2012. Crowdsourcing malaria parasite quantification: an online game for analyzing images of infected thick blood smears. *Journal of medical Internet research* 14, 6: e167.
- [14] Ben Medler. 2011. Player dossiers: Analyzing gameplay data as a reward. *Game Studies* 11, 1.
- [15] Florian Mehm, Christian Reuter, Stefan Göbel, and Ralf Steinmetz. 2012. Future trends in game authoring tools. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 536–541.
- [16] Pejman Mirza-Babaei, Veronica Zammito, Jörg Niesenhaus, Mirweis Sangin, and Lennart Nacke. 2013. Games user research. *CHI '13 Extended Abstracts on Human Factors in Computing Systems on - CHI EA '13*: 3219.
- [17] Lennart E Nacke, Anders Drachen, Kai Kuikkaniemi, and Yvonne a W De Kort. 2009. Playability and Player Experience Research. *Proceedings of the IEEE*: 1–11.
- [18] Lennart Nacke, Christiane Moser, Anders Drachen, Pejman Mirza-Babaei, Andrea Abney, and Zhu (Cole) Zhenyu. 2016. Lightweight Games User Research for Indies and Non-Profit Organizations. In *CHI Extended Abstracts on Human Factors in Computing Systems*, 3597–3603.
- [19] Dmitry Namiot and Manfred Sneps-Sneppe. 2014. On Micro-services Architecture. *International Journal of Open Information Technologies* 2, 24–27.
- [20] Geraldus Galehantomo P.S. 2015. Platform Comparison Between Games Console, Mobile Games And PC Games. *SISFORMA* 2, 23–26.
- [21] Carlos Pereira Santos, Kevin Hutchinson, Vassilis-Javed Khan, and Panos Markopoulos. 2017. Measuring Self-Esteem with Games. In *ACM Intelligent User Interfaces*.
- [22] Carlos Pereira Santos, Vassilis-Javed Khan, and Panos Markopoulos. 2016. Inferring A Player's Need For Cognition From Hints. In *Proceedings of the 21st International Conference on Intelligent User Interfaces (IUI '16)*, 76–79.
- [23] Marc Prensky. 2001. Fun , Play and Games : What Makes Games Engaging. *Learning, Digital Game-based* 1: 1-05.

- [24] M. Rosenberg. 1965. Rosenberg self-esteem scale (RSE). Acceptance and commitment therapy., Measures package 61: 52.
- [25] Jesse Schell. 2008. The Art of Game Design: A Book of Lenses. Morgan Kaufmann.
- [26] Matthew J Sharritt and Michel A Sharritt. 2010. User-experience game research? In Games + Learning + Society (GLS) 6.0.
- [27] Narnia C. Worth and Angela S. Book. 2014. Personality and behavior in a massively multiplayer online role-playing game. Computers in Human Behavior 38: 322–330.
- [28] Georgios N. Yannakakis, Pieter Spronck, Daniele Loiacono, and Elisabeth André. 2013. Player Modeling. Dagstuhl Follow-Ups 6, 59.
- [29] Amri Yusoff, Richard Crowder, Lester Gilbert, and Gary Wills. 2009. A conceptual framework for serious games. In Proceedings - 2009 9th IEEE International Conference on Advanced Learning Technologies, ICALT 2009, 21–23.
- [30] Unity. 2016. Report: In-game advertising the right way: Monetize, engage, retain. Retrieved 01/19/2019 from <http://response.unity3d.com/in-game-advertising-the-right-way-monetize-engage-retain-whitepaper>

Chapter 5

Player Experience Evaluation in the Games Industry

Abstract

This study aims to bridge the gap between the academic field of Game User Research and the game development industry by looking directly into company needs and investigating the role of player experience evaluations within a game's life cycle. In the pursuit of this aim, qualitative, semi-structured interviews of eleven different game companies of different sizes and nationalities were conducted.

Our study reveals a discrepancy between the frequency of player experience evaluation studies carried out by large game companies and those conducted by small to mid-sized companies. We argue that smaller players can be supported, by adopting a more varied set of user experience evaluation methodologies, developments of more and complementary tools that can make the process of data collection and analysis more efficient. Further new approaches, such as in-game questionnaires, increase company cooperation and sharing of game data with researchers, are put forward as approaches for closing that gap.

Summary:

Chapter 5 reviews the current industry practices for player experience evaluation, and reviews if a tool like the one presented in the previous chapter would be useful within a game company development cycle.

Submitted to:

ACM Annual Symposium on Computer-Human Interaction in Play (CHIPLAY)

Authors:

Carlos Pereira Santos, Heleen van Geelen, Vassilis-Javed Khan, Panos Markopoulos

Introduction

In the past forty years, the video game industry has grown considerably at a global scale, which is evidenced by annual growth rates ranging from 9% to 15% over a 25 year period [1-2]. Such is the shift in the entertainment field that nowadays more people choose to play video games instead of going to the movies [3]. Recent studies forecast that, for 2018, the overall revenue on video games should reach \$139.9 billion, which represents a 13.3% increase over the previous year. It is stated that digital sales revenues will capture 91% of the global market at \$125.3 billion, of which 51% or \$70.3 billion will be generated by mobile games, and the remaining 49% will be the combined revenue from PC and console games [4].

The increasing importance of the video game industry has been accompanied by a growth in related research. A growing body of research investigates the design of games and the game experience. However, such research activity is still proportionally much smaller compared to the research effort invested in other entertainment industries, such as television, music or movies [5].

While there is a growing body of research focusing on game-related experience evaluation methods and tools [6-8], it is currently not clear how such methodology is taken up by industry, and an understanding of current practices within video game companies to evaluate the player experience is still missing.

Therefore, the aim of this study is to consult directly with companies within the game industry and identify how much of the current knowledge centring on the player experience is being applied by game developers, how that knowledge has been integrated into their development cycles, what tools and methodologies are being used, how company needs are identified, and their takes on data sharing.

Background

Video games are an entertainment media branch i.e., they require voluntary acts from the consumer side that are based on the amount of pleasure or fun gained from playing a game. Hence, it is important to provide an enjoyable playing session and provide users with the best gaming experiences possible. When discussing player experience, the focus is not on the creation, implementation or design of the video game, but on the experience generated by playing it, i.e., the interactions between the player and the game.

Typically, positive experiences are linked with concepts such as fun, flow, immersion, playability, engagement and presence, and, when studying experience constructs, these positive aspects are associated swiftly with the physiological, cognitive, physiological and emotional aspects of gameplay [9]. User Experience in Games or Game User Research (GUR) is a research field

concerned with the development of tools and techniques to measure and evaluate players' experiences [10-12].

Player Experience Evaluation

To gain insights into how companies evaluate their players' behaviours and gaming experiences, we need to understand how player experience can be evaluated. Therefore, in this section, we review broadly the current methods used in GUR in order to later compare and contrast these methods with what is currently being utilized in the video game industry.

Obviously, depending on the planned research goal, there are many GUR methods for observing, measuring, and analysing players. A combination of methods can be used to collect different types of data sequentially or even simultaneously, and all methods have their advantages and disadvantages. This is not an exhaustive survey, it is just meant to highlight the diversity and completeness of such methodologies. Details concerning the different methodologies can be found in [10-14].

Probably the most well-known method for evaluating the player experience is player testing, in which players are observed while playing a prototype or pre-release of a game. In order to collect more in-depth information, individual or focus groups interviews, heuristic evaluations or open questionnaires are effective qualitative methods that can be utilized. Diaries and probes in which players report experiences can be used to track longer term engagement. Questionnaires, video encoding, biometrical data and, as of late, data analytics of metrics (player behavioural data), are quantitative methods that are becoming increasingly relevant.

Typically within a player experience evaluation, multi-measure approaches are preferred as they provide more leverage in identifying factors that influence the player experience and how they are related to certain feedback compared to a single, isolated measure [15].

Game User Research-Focused Tools

In both industry and academia, the use of appropriate tools is fundamental for GUR. Not only are the datasets becoming bigger, but their depths and complexities are increasing. Moreover, the terminology and integration within the different individual roles, game development phases and processes around the evaluation of user experiences are closely intertwined [16].

Game analytics has recently emerged as a fresh and powerful methodology, one that allows companies to increase the quality of games by understanding players better by testing and proving theories. Still, specialized and adaptable tools that can be integrated with each other to analyse and interpret complex datasets that are generated during game play [17] are required.

Earlier studies have sought to chart and establish the state-of-the-art tools for an ever-evolving industry [18], and even some have looked into defining tools for independent game developers [19]. The present study aims to add to these works, by exploring this topic directly with the industry defining industry needs with respect to GUR methods and tools.

Method

To understand the needs of industry we conducted a set of semi-structured interviews with experts from industry. The interview was designed in line with the general interview guide by Gall, Gall, and Borg's [20], combined with a standardized open-question approach. Semi-structured interviews give some flexibility to delve deeper into specific questions based on what answers are given. Using such an approach also means that there may not be equal consistency within the interviews because the questions can be altered when needed, resulting in different answers [21].

Participant Pool

With the aim of procuring a representative population sample of game companies, we employed a criterion sampling approach combined with maximal variation [22].

The interviews were purely voluntary, no monetary compensation was offered to the participants, and research consent forms were signed by all participants. In addition, all participants were offered the chance to keep their names and that of their company anonymous. The intension of the interview was to explore how player experience evaluation are performed within the industry, i.e., we were not interested in targeting specific internal products or processes within companies, and offering anonymity allow us to overcome some company barriers.

The participation pool contained 11 different video game development companies in different gaming areas and targeted various platforms and genres. We also attempted to ensure participation and representation from small-sized (1 person) to large studios (up to 12,000 persons); a detailed demographic table is included bellow (Table 5-1 and Table 5-2).

The average age of our participants was 34.2 (SD=8.8), and all were male. They reported an average of 10.6 (SD=6.8) years of experience, and, together, they had accumulated more than 110 years of experience.

Ethical Considerations

In order to protect the quality and integrity of the study, ethical guidelines for the study were defined. Before data collection, an informed consent form and information sheet were created. It provided the participants with information

concerning how the study would be conducted, what their role was, what the ethical guideline were and contact information for the researchers, which provided them with the opportunity to contact the researchers in case (ethical) questions or complaints arose. In the consent form, it was stated clearly that: *i)* participants had the right to stop the interview or withdraw from the research at any time, *ii)* the interview would be recorded and transcribed by an external service and *iii)* participants had the option to anonymize the quotes.

After transcribing the interviews, the transcripts were shared with the participants to allow them to read and adjust them, if needed. The data gathered in this study is to be used solely for research purposes.

Interview Procedure

To create appropriate research questions for the interviews, McNamara's recommendations for research questions for interviews were followed [21], i.e. questions should be: *i)* open-ended, *ii)* neutral, *iii)* structured in multiple parts with their own subjects, *iv)* complete with explanations for specific critical terms to prevent preconceptions, and *v)* supportive, i.e., only use answers that were already provided on a certain topic.

To make sure that the set of questions was suited for the study would contain no flaws or limitations, a pilot study was done to allow us to alter the questions or add questions, if needed [24]. The pilot test was performed in a game studio under conditions similar to those to be used for the full set of interviews [25]. The interview questions were then reviewed by a third-party researcher before the official interviews were conducted. In addition, written procedures were compiled to assure consistency for all interviews.

Participants were invited to participate in the study via an email which contained a short description of the goals of the study and conditions for participation. Contacts were provided to answer any questions and arrange a date and time for the interview.

The interviews were done either in person, by phone or via a video call. They began with the interviewers (normally two) introducing themselves, followed by a revisiting of the purpose of the interview.

An information sheet was also given to the participants to explain in more depth how the study would be conducted and what their role would be in the study and contained contact details should they have further questions or encounter problems after the interview.

The consent form (sometimes provided in advance at the request of the participants) was given to the participants to read and sign. No time limit was given, allowing the participants to read, reflect and ask questions. After clarifying all questions about the interview or procedure and the consent was signed properly, the audio recording device was set up and the interviewed commenced.

Company Name	Years	Continent	N Persons	Platform/ Genre	Revenue model(s)
Company 01	6	Europe	4	PC/PS/ Xbox/ Mobile	Premium
Oven Games	3	Europe	1	PC/PS/Xbox/ Mobile	Premium & Freemium
Short Fuse	2	Asia	10	VR Arcades Shooters	Premium & Subscription
Twirlbound	5	Europe	7	PC/PS/Xbox	Premium & DLC's
GDM	6	Europe	3	Mobile	Advertisement Freemium
Split Polygon	5	Europe	8	PC Multiplayer Space Sandbox	Premium & DLC's
Company 07	30	International	12000	PC/PS/ Xbox/ Mobile	Premium, DLC's &Freemium
Company 08	6	International	40	PC Space Fighter	Freemium
Ronimo	11	Europe	18	PC/Mobile/PS/ Xbox/Nintendo	Premium & Freemium
Space Nation	2	Europe	50	Mobile NASA Space	Investments & Partnerships
Company 11	12	Asia and North America	1100	Mobile mainly RPG & Strategy	Freemium & Premium

Table 5-1: Detailed company demographics information.

Company Name	Participant Name	Age	Years of Experience	Title
Company 01	Anonymous 01	28	6	Creative Director
Oven Games	Bojan Endrovski	35	10	CEO (founder)
Short Fuse	Frank Bloemendal	30	5	Game Designer
Twirlbound	Mathijs van Laar	23	4	Creative Director
GDM	Jerry van Heuvel, Mikel Agricola	28, 32	6, 12	Both CEO (founders)
Split Polygon	Joost Meulenkamp, Paul Mertens	28, 33	5	Both CEO (founders)
Company 07	Anonymous 07	43	20	Senior Tools Programmer
Company 08	Anonymous 08	58	27	Senior Producer and Evangelist
Ronimo	Jasper Koning	37	11	Founder and Game Designer
Space Nation	Fábio Florencio	37	15	Studio Manager
Company 11	Anonymous 11	32	11	Business Dev Director & Senior Producer

Table 5-2: Detailed participant's demographics information.

The first part of the interview (5 to 10 minutes) was focused on demographic information concerning the participant and the company they worked for, which resulted in the summary provided in Table 5-1 and Table 5-2. Some additional information pertaining to the participants' educational backgrounds and work experiences was also requested, but not included in the table.

For the second part of the interview (25-35 minutes), the focus shifted to current practices. We quoted the definition of Player Experience Evaluation, multiple diverse examples were given, and clarified any questions the participants may had. This allowed us to minimize any different preconceptions of the term, and then we asked questions concerning its integration within the development pipeline, the toolset and methodologies used within the company.

For the third and final part of the interview (20-25 minutes), we focused on opinions and attitudes regarding future tools. A video describing the core features of a player experience evaluation platform – Games User Research as a Service (GURaaS) [26] – was shown to the participants. It was requested that the participants share their opinions and thoughts about the platform. This segment was then followed by judgement questions concerning broader topics, such as the future challenges related to player experience for the company. The interviews took normally between 60 to 80 minutes.

We used an external service which had signed a non-disclosure agreement to transcribe the audio, then we proceeded to anonymize the document. Once it was anonymized, the pertinent parts were sent back to the individual participants, and they were given a two-week period to provide any alterations.



Figure 5-1: Game User Research as a Service portal, depicting the the most active games information over time.

Games User Research as a Service

Since a video of GURaaS [26] we describe in this section more in depth the concept of the platform. GURaaS is a player experience evaluation cloud-based platform, able to collect large amounts of data and logging in-game player behavior, and which empowers game researchers to integrate and conduct studies among the player population (see Figure 5-1).

The video^[1] presented to the participants is about 55 seconds long, and shows, players playing games, some development scenes within a game engine, the data analytics portal, and a game with a questionnaire embedded, we also present the transcript of the video that was displayed to the participants in this section.

“Completely comprehending the player behaviour within games is extremely valuable to provide great experiences. Games User Research as a Service, or GURaaS for short, is a toolset that was specifically designed and developed to support game developers and researchers to capture, analyze and interpret game metrics. GURaaS allows from small to large game companies to integrate through an open source library behavioural telemetry in their games in a quick and simple way.

Through a web portal the gathered data can then be downloaded, filtered, plotted, or shared with others for deeper analysis. GURaaS also integrates a questionnaire tool, by allowing through the portal the design of qualitative questions and target a specific player base, to more completely understand player behaviour.”

Data Analysis Procedure

Considering that the data collection consists solely of interview material, qualitative data analysis methods were used to derive general themes inductively.

Elements of the grounded theory approach as this is described by Glaser [27] were adopted, i.e., data collection was focused on developing theories, concepts, and patterns. The data collection, analysis, and reflection undertaken during this study were meant to provide more insights into and depth to the concept of player experience evaluations conducted within video game companies. This approach was accomplished via the constant comparison method utilised in grounded theory, i.e., the interviews were thoroughly cross-cased to compare them with each other but also to compare different parts within one interview or theme [28]. An inductive approach was used to transfer raw text data into an overall summary format and to develop a model or theory about the processes and experiences based on the raw data.

1 <https://www.youtube.com/watch?v=i4geF2crnnE>

All transcribed interviews were subject to open-coding analysis, and axial coding (called labelling in this study) was applied. Open coding is done in order to read and see what patterns evolve throughout an interview objectively without linking the analysis to a theoretical framework or theory.

Labels for different types of data were extracted from the transcripts, i.e., the labels functioned as the overall themes for categories, actions, or statements that were discussed. Different keywords or sentences from the answers of all of the participants were then collected and put into one model for constant comparison to see if patterns could be discovered [29].

Results

This section describes the most important findings from the study of the 11 video game companies (see Table 5-1 and Table 5-2). From here onwards, these companies will be referred to as C.01 up to C.11, respectively.

Overall, all the video game companies describe player experience as being a very important factor. Nevertheless, when discussions came around to how that evaluation was being performed, there were clear differences between the systematic approach employed by large video game companies, and how closely it was integrated with their development cycle, and the approaches utilized by smaller companies, whose evaluations clearly evidenced a lack of organization and structure, when done at all.

When requested to describe how the evaluation of the player experience would impact their product development, they claimed that it could have major impacts on the products at multiple levels. Although the return on investment is difficult to quantify and measure, the interviewees reported on drastic changes and massive revisions of specific game features based on the outcomes of the player experience evaluation. They also mentioned making more informed decisions on how to proceed, including refactoring a specific mechanic, changing the ordering of the levels and major project restructurings, such as modifying full mechanics, changing monetization strategies and even restarting full projects.

Developers doing data collection after a product launch reported on the importance of cultural differences that can be revealed by GUR. Such differences on a global market are extremely relevant and seem to not have been fully taken advantage by some developers.

“ C.05: “I would say it’s the most important thing.”

“ C.05: “Most important thing publishers look at is retention. (...) For us, for instance, Mandala Garden was soft launch in the Philippines. The reason why is because it has an Asian type of culture.”

“ C.03: “I think that we had a quite a limited approach [to player experience evaluation] because we’re always working and dealing with deadlines (...) We were not focused enough on collecting data, I think that’s one of the reasons why our game didn’t sell that well.”

“ C.02: “We would make big, large, redesigns of what the game would play like based on those things. (...) The last game that we made, we scrapped the game four months into development. We had a contract towards Microsoft that we will deliver, we scrapped the game and started from zero because it was unexciting.”

“ C.11: “Where in our more recent test, we had about 100 people playing, we were making updates every day. I think, it’s a really important role when you’re talking about user experience is to be able to get it, try something, iterate, move forward. You need the infrastructure to do that.”

“ C.7: “I see that perhaps even more so in China and Asia in general, where people playing games much more on say like a mobile device.”

Referenced Methodologies

Most of the interviewed companies used a very wide variety of methods depending on the project stage, but the most frequently mentioned method was is holding playing sessions, with players recruited from within the company itself, from friends, or by recruiting anonymous participants.

“ C.08: “Yes, we play internally a lot.”

“ C.03: “Testing process we did every Friday, we got the company game demo and we would play the game with the whole company and from there on would try to find bugs and we would write it down in a back log where we’re working on the next sprint. Also, we create the testing days where we would invite players from friends to come into the studio and we would host gaming events basically where we would gather feedback from the players.”

Another frequent method used to collect player behaviours was to collect game metrics remotely and aggregate them to perform data analysis.

“ C.02: “For Google is, we’re using a lot of Google services. It’s a one log-in to kind of solve everything so you don’t have to do a lot of different websites, different services that you need to integrate.”

“ C.11: “My team, we’re using a little bit more data analytic side. We’re trying Tableau and working at that as a pipeline.”

“ C.07: “When we have so many millions of people playing the game, we have vast, vast quantities of data.”

“ C.09: “We integrated the software from the company called deltaDNA mostly because the front-end stuff that would take a lot of time to build ourselves. I mean, for a technical stand points, logging an event in the game to some server online isn’t that hard. It’s mostly the way it was able to filter and display all the data that was very hard to do internally. They have a very flexible setup for that.”

For some companies, showcasing their games at public events is very important: it serves as product advertisement and allows them to evaluate the experiences of anonymous people, while serving as motivation, time pressure or a deadline to deliver something. The term “events” was repeated by different companies, in reference to settings that allowed them to have direct contact with players, observe them play live, and ask them questions about it.

“ C.02: “What has really worked well as well previously for me has been events because an event will push you to really wrap up things (...). Then I observe how people are playing the game so I come there with a piece of paper, listen to everyone’s advice, you’re excited, your ears are waiting to hear other people, you can observe them playing.”

“ C.04: “In the company we’ve always said that we always try to do a public showcase as a deadline because that means that there’s no leeway. We have to show something and even though it’s not completely finished it will always teach you something about what you’ve built.”

Another frequent method mentioned was direct contact with players or communities through media channels such as Reddit^[2] and Discord^[3]. In addition, they pay close attention to social media platforms looking for reviews.

“ C.06: “We tend to have an open style of communicating with players. We have a Discord channel. We’re always replying on the forums. I’m always looking at reviews, looking at what they’re saying into this channel. I do invite them, “Hey, what do you think?” We’re doing streams as well. (...) We don’t really do it formally. When you look at the community, all the different things they say.”

² <https://www.reddit.com/>

³ <https://discordapp.com/>

“ C.08: “No it’s online, they get a Steam Key, but we also have a discord channel setup for interacting with the player base and we have a couple of people that had experience with moderating those forums, step up to help us out there and we get a lot of daily interactions with players there and they propose suggestions, ask questions, and we periodically let them know about something that is going”

Most of the developers seem to be conscious of the limitations of gathering player experience through such methods. One problem that was commonly mentioned was knowing that local persons might lack target audience representability. Discussing the game with the developer may also lead to participation bias, and community-based communications only reach persons with specific personalities. Other, more traditional, research methods, such as using surveys and interviews, were also referenced, but they were not mentioned frequently.

“ C.04: “We’ve come to realize that at conferences people are not that honest especially if you’re standing next to them. They won’t just blatantly flame your game or whatever. It’s actually really interesting to step back into the audience and out of conference and zip up your vest so that your shirt is invisible and then hear what people actually say when they walk by.”

“ C.08: “In my earlier industry experience, I got to observe a focus test once that Microsoft organized for Blood Wake (...) It showed me the value of focus test, doing, setting it up and running it professionally, you certainly can get a lot of good information. You also have to be careful that you’re not skewing the results to get to hear what you want to hear.”

“ C.09: “That means that you’re only getting word from the people who are mostly invested into your game. Well, it’s just limited. I mean, it’s reliable in a sense that you know what they want, but it’s not necessarily reliable to know as a token of what the whole community wants.”

While large companies are willing to invest in the resources to conduct a proper player experience analysis across all stages of the product lifecycle, smaller companies have more difficulty in doing so. When questioned about why player experience evaluations were not used more extensively, the responses were rather similar. There were clear references to a lack of expert skills or complexity to analyse data, but the most cited reasons were time and money.

It was also clear that small companies, after successfully publishing a game, struggled to define the best strategy to either keep performing maintenance in the just-launched game or to use the limited resources (human and funds) to start new projects and repeat the cycle.

“ C.01: “In the previous game, we didn’t. When we shipped the game, we were just running out of funds, so we couldn’t really add more features in there or these measuring systems (...) Even if we were to get all this very specific information, we probably wouldn’t use it for anything apart from, “Well, that’s interesting.””

“ C.04: “I just think it’s so important and that’s the irony in it, obviously, because if I have to say it’s more important than actually making the game, then I would say no. Then again, I know how important it is, so if we had a bit larger company, I would definitely want to spend more resources on testing player experience and the analysis of it, as well. Yes, I just hope it will have a bigger role in our company in the future.”

“ C.09: “After we made Awesomenauts, we made Sword & Soldiers while we’re supporting Awesomenauts.”

Tools for Player Experience Evaluation

With regards to collecting player experiences, a lot of companies use communication-based contacts, either through media channels or face-to-face to reach participants. Nevertheless, companies are also using data analytics tools. The usage of internal or external third-party tool services depends on company policy or preference independent of company size. There are companies that tend to prefer to use as many in-house development tools as possible, from engines to the data analytics tools, while others tend to use third-party software packages.

These third-party tools are considered to be valuable but also adds some challenges due the need to adapt them to a company’s needs, which leads sometimes to hurdles to properly used and integrating them in the development pipeline.

“ C.07: “We have a lot of proprietary technologies and tools and methodologies of processes.”

“ C.05: “What we’ve noticed is that if you bring in an external tool that you can download somewhere or buy for a few bucks, it always takes more time to actually properly integrate and get rid of the things you don’t need than it is to build something of your own that is very specifically tailored to what you need.”

“ C.01: “We use a system that comes with Unity itself. I think it’s called Unity Analytics. You can put hooks in your game code that sends information about what the player is doing, when they’re doing it, from where they’re doing it.”

It is not attempted here to draw general conclusions regarding the strengths and weaknesses of each of the various software packages mentioned by the companies, since answers are linked to company differences and individual preferences. However, the tools mentioned include (in no specific order): Google Analytics, Unity Analytics, Flurry, ElectronJS, Apache log4net, DataDog, DeltaDNA, SumoLogic, Tableau, Crashlytics and Periscope Data.

Attitudes Towards the GURaaS Platform

The overall reaction to the video on the GURaaS concept was positive, and the terms “interesting” and “valuable” were used broadly. Some participants were enthusiastic and eager to know more, while others were a bit sceptical since it was a short video and did not contain enough information for them.

The participants from larger companies had interesting reactions to the platform, while C.07 clearly recognized the added value for mid-size companies, but said categorically that they would not use it, with the main reason being that they have a tailor-made, in-house toolset developed for this purpose that has been in use for years. C.11 uses an off-the-shelf tool but became curious about adopting some of its differentiating features.

“ C.05: “It really needs to convince me that it’s better than the logger we have and to be worth it for me to spend time and money to implement it.”

“ C.07: “They have been doing it for 10, 20 years or something. (...) seems like it could be something that would be really helpful for mid-scale developers who have perhaps already released their game and they want to get that live data, that post launch data”

“ C.03: “We don’t have a lot of resources to basically focus on this. So, a tool like this will be very useful for smaller companies like us.”

“ C.10: “You can have everything in the same place. It’s already collecting data and providing you a result or a summary of what happens. (...) I don’t see any other [platform], thinking about any other holistic approach as this.”

“ C.11: “We have integrated our own tools, not in this kind of fashion, but in the same angle. I think, it’s a nice tool to wrap everything around.”

The GURaaS concept was seen as “powerful” and interesting, namely, because it focuses on in-game data collection, has the ability to target specific player audiences, and has the ability to insert questionnaires into games. However, this last feature was seen risky, since the interruption could affect the player experience severely, and there were some concerns related to the validity of the results.

- “ C.03: “I think that breaks the experience or the flow of players.”
- “ C.07: “As long as it doesn’t interrupt the player, doesn’t disrupt the player experience, I think it’s great. Having it integrated in the game can be very useful.”
- “ C.09: “...if you would segment players based on certain behaviours in game and then you could ask say segment specific questions (...) think it would definitely be interesting, and I think if done right it could provide valuable data. I think in of itself it becomes a thing that you were not only developing a game; you’re developing a psychology measurement tool”
- “ C11: “They’re streamlined and they can fit in the UI, I think, it would work really well. We’ve done similar things when I was working at browser gaming. When we did external link, in one test we had 5% of people respond. (...) sometimes companies don’t even have enough data, so I had to rely on more of the player survey responses.”

Data Sharing and Player Privacy

When questioned directly in relation to the willingness to share data with third parties, the answer was a strong ‘no’ from most of the participants. Further, those who did not disagree immediately followed up with concerns about not knowing what third parties would do with that data.

As a follow up, they were asked about their willingness to share their player data for research or social contribution purposes. In contrast to the previous responses, almost all said they would, and some even speculated about their willingness to cooperate with other companies and trade their data to learn more about other games. This form of data sharing was generically more appreciated and helped overcome some protectiveness. The clear exceptions were the large-sized companies, whose representatives stated that they consider this data to be too valuable for their corporation to share with anyone.

- “ C.08: “My instinct is that most people are going to consider that proprietary business information that they would be reluctant to share, my biases towards openness. It’s like, what can you learn from this?”
- “ C.05: “I think data is gold so you won’t share it probably sell it (...) we would be prepared to eventually share it or sell it. It depends on what’s the purpose behind it.”
- “ C.07: “I doubt it, pretty much (...) I think, in general, [company name] has some interest in supporting the community in a more general sense and not just thinking about it in terms of a commercial venture. I think sharing user research data might be a bit too much.”

- “ C.09: “It would be interesting to have a shared database of all kinds of games and their findings and measurements for sure.”
- “ C.10: “I think it’s a win-win. I contribute to this collaborative thing, which now becomes a service (...) By giving, I’m receiving more data so I can assert this information.”

We purposely tackled the question of player privacy, which was a hot topic around the interview dates. Both the privacy global scandal with Cambridge Analytica [30] and the introduction of the European General Data Protection Regulation (GDPR) [31] were being discussed extensively in the media at the time of the interviews.

Most did not feel that player privacy was a concern, and they “assume” players know data is being collected. Interestingly, European companies were very aware of, and had clear opinions on, player privacy but didn’t properly assess the implications of the new regulation; instead, they seemed concerned about the direct implications of the new GDPR.

For non-European companies, the scenario was completely different, but most of those who were aware of the situation were not concerned. Several contrasting comparisons were made with China, where it is mandatory to track individual player behaviour and provide it to government agencies.

- “ C.03 “Well, I live in China, even my rights are a bit different (...) It’s not going to do Cambridge Analytica thing with it. I don’t think you can do that with the information (...) I think it’s difficult to answer this question.”
- “ C.04: “I think that the new law is something that I should also read into more”
- “ C.07: “I suppose my main concern is not so much to do with cultural or ethical concerns. I think more practically like how do we deal with this kind of data.”
- “ C.09 “I think most enthusiasts at least, or most people that play our games, are aware that a lot of games nowadays try to track what players do in their game. I don’t think it will change the specific issue that much.”
- “ C.10: “It doesn’t matter if you agree or not or if, we, as users, agree or not. Data is being collected anyway.”
- “ C.11 “I don’t know if you’re aware of some of the stipulations on entering the Chinese market. We don’t really record to the minute on how long someone stays in the game, but the government of China requires that. To essentially ensure that the player can’t progress if they play more than, I think, three or four hours a day, if they’re under 18 years old or something.”

Discussion

We need to keep in mind that there are multiple factors that play a role in the product cycle of a video game. Video games are first and foremost about an experience, and an experience remains subjective, even to those who create them. Data can be interpreted in many ways, and that also applies to this study. Diverse individual perspectives and differences reported by participants reflect their different roles in the industry, and thus help reflect on the broader game industry landscape.

The main themes related to player experience evaluation that emerged from our analysis were: *i)* attitudes towards the player experience, *ii)* methodologies, *iii)* attitudes towards the GURaaS Platform and *iv)* sharing of player in-game data.

Player Experience Evaluation Attitudes

The first finding showed the size/resources that the companies have, result in clear differences in the way GUR is integrated into development. Smaller teams affect the work capacity, and workers need to be flexible and take on many different roles and tasks to complete the product. None of this information implies that the role of player experience evaluations is not seen as important. To the contrary, most of the companies stated that they think it will become more important and larger than it already is.

The combination of limited resources and the lack of clarity concerning what the return on investment will be for conducting player experience evaluations dissuades smaller companies from investing more heavily in such practices. Further, when they do invest in them, they tend to opt for quick and/or cheap return methodologies that allow them to validate their actions.

Outsourcing these tasks was mentioned multiple times, but, when asked, if they were willing to share data with external services, the question was not received well. Overall, the answers centred on protection and doubt that such services would: *i)* understand a game and its vision deeply, and believing that such a service *ii)* could not conduct a proper analysis of what the companies wanted to know specifically.

Some smaller companies, even those with large and challenging games in the making, seemed to have drifted away from the evaluation of the experience and lacked clear definitions of the structure and vision. When questioned as to these observations, the pressure of deadlines and lack of skills driven by limited resources appeared to be the main culprits.

Exploring New Methodologies

The findings from the methodology discussion are also interesting. Traditional quantitative research methodologies are applied at large companies in order to attain more objectivity in their results. They have the ability to invest in the development of in-house knowledge and tools and to integrate them into their own development pipeline.

Smaller companies use other methodologies, such as direct contact with player communities, perceptions gathered from participating in events, and internal play testing. All of these methods have clear and known drawbacks such as being representative of only the more extrovert players, lacking cultural diversity, players not being honest when facing the developers and communities that hide behind (semi) anonymous profiles and inflate situations.

Still, and according to the interviewed companies, those methods are very useful, cheaper and quicker, even if the information may not be fully reliable. The notion was that if you are conscious of the lack of reliability, then there is still a lot of usable information that can support your experience design.

“ C.08 “Well, I think the main challenge is always to take the data with a big grain of salt”

Our suggestion is not to convince those companies to change their methodologies but to understand and define them better. All research methods have their advantages and disadvantages, and researchers should clearly evidence and publicly share the reliability and validity of such methodologies. In this manner, they can support companies in making better decisions.

Attitudes Towards the GURaaS Platform

The GURaaS video was introduced to encourage a speculative discussion of existing tools and common needs. We were able to confirm positive attitudes towards the platform, with the participants describing it as bringing innovative concepts when compared with other toolsets, although the in-game questionnaire raised some concerns in terms of the validity of the player experience effects.

The video also facilitated the discussion and comparison of tool attributes, which allowed us to compile a collection of the attributes that the interviewees deemed to be important in GUR tool: *i)* affordable, *ii)* easy to integrate, *iii)* centralized information, *iv)* predefined player models, *v)* up to date multi-platform support and *vi)* handle large amounts of data.

Sharing of Player In-Game Data

As long as player privacy is not at risk, trustworthy research purposes could convince companies to share their data; some participants even suggested cooperative partnerships by sharing data with other companies.

Such sharing holds tremendous potential for future platforms and could, in theory, bring the areas and expertise in research, closer to the game companies who hold these exclusive data but are in need of the skills these data scientists possess. It could form the bridge that gives the research area access to data that has been exclusive and, in return, give game companies access to more valuable data and analyses.

Conclusion

This study proposed to explore and identify currently used practices in player experience evaluation in the game industry. We contribute to the research field through identifying various gaps and challenges within companies, in terms of investments, resources, methodologies and tools used. Moreover, this research provided insights, on tool attributes which could support the evaluation of the player experience and making the development cycle more efficient.

Finally, there are some interesting findings in this study that hold potential for exploring different methodologies and tools that could help close the gap between the academic field of research and the video game industry and encourage cooperation.

We interviewed different participants and looked at the different development cycles, methodologies and tools used to generate different games, but all avenues seemed to lead back to the overarching theme of ludology and the factors that contribute to the understanding of the player experience.

Acknowledgment

We would like to acknowledge the contribution of the participants in the study; without them, it would not have been possible. In particular, we thank them for their patience and the time they invested in the study.

References

- [1] P. Zackariasson and T. L. Wilson, "Paradigm shifts in the video game industry," *Compet. Rev.*, 2010.
- [2] ESA, "Essential Facts about the Computer and Video Game Industry 2018," Entertainment Software Association, 2018. [Online]. Retrieved 19-Jan-2019 from http://www.theesa.com/wp-content/uploads/2018/05/EF2018_FINAL.pdf
- [3] NDP Group, "More Americans Play Video Games Than Go Out to the Movies," NDP Group Press Release, 2009. [Online]. Retrieved 06-Feb-2019 from https://www.npd.com/wps/portal/npd/us/news/press-releases/pr_090520
- [4] T. Wijman, "Mobile Revenues Account for More Than 50% of the Global Games Market as It Reaches \$137.9 Billion in 2018," *NewZoo.com*, 2018. [Online]. Retrieved 12-Sep-2018 from <https://newzoo.com/insights/articles/global-games-market-reaches-137-9-billion-in-2018-mobile-games-take-half/>
- [5] A. Marchand and T. Hennig-Thurau, "Value creation in the video game industry: Industry economics, consumer benefits, and research opportunities," *J. Interact. Mark.*, 2013.
- [6] R. Bernhaupt and F. "Floyd" Mueller, "Game User Experience Evaluation," in *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '16*, 2016.
- [7] P. Mirza-Babaei, V. Zammito, J. Niesenhaus, M. Sangin, and L. Nacke, "Games user research," *CHI '13 Ext. Abstr. Hum. Factors Comput. Syst. - CHI EA '13*, p. 3219, 2013.
- [8] A. Drachen, P. Mirza-Babaei, and L. E. Nacke, *Games user research*. 2018.
- [9] J. L. G. Sánchez, N. P. Zea, and F. L. Gutiérrez, "Playability: How to identify the player experience in a video game," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2009.
- [10] R. Bernhaupt, W. Ijsselsteijn, F. "Floyd" Mueller, M. Tscheligi, and D. Wixon, "Evaluating user experiences in games," in *Proceeding of the twenty-sixth annual CHI conference extended abstracts on Human factors in computing systems - CHI '08*, 2008.
- [11] M. Seif El-Nasr, H. Desurvire, L. Nacke, A. Drachen, L. Calvi, K. Isbister, and R. Bernhaupt, "Game user research," in *Proceedings of the 2012 ACM annual conference extended abstracts on Human Factors in Computing Systems Extended Abstracts - CHI EA '12*, 2012, p. 2679.
- [12] G. McAllister, P. Mirza-Babaei, and J. Avent, "Improving Gameplay with Game Metrics and Player Metrics," in *Game Analytics*, 2013, pp. 621–638.
- [13] K. Isbister and N. Schaffer, *Game Usability*. 2008.

- [14] R. Pagulayan, K. Keeker, T. Fuller, D. Wixon, R. L. Romero, and D. V. Gunn, *User-centered Design in Games*. Boca Raton: CRC Press, 2002.
- [15] L. E. Nacke, A. Drachen, K. Kuikkaniemi, and Y. a W. De Kort, "Playability and Player Experience Research," *Proc. IEEE*, pp. 1–11, 2009.
- [16] E. Brown, "The Life and Tools of a Games Designer," in *Evaluating User Experience in Games*, no. May, R. Bernhaupt, Ed. London: Springer, 2010, pp. 73–87.
- [17] M. S. El-Nasr, A. Drachen, and A. Canossa, *Game Analytics: Maximizing the Value of Player Data*. 2013.
- [18] C. T. Tan, P. Mirza-babaei, V. Zammitto, A. Canossa, G. Conley, and G. Wallner, "Tool Design Jam : Designing Tools for Games User Research," in *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '15)*, 2015, pp. 827–831.
- [19] L. Nacke, C. Moser, A. Drachen, P. Mirza-Babaei, A. Abney, and Z. (Cole) Zhenyu, "Lightweight Games User Research for Indies and Non-Profit Organizations," in *CHI Extended Abstracts on Human Factors in Computing Systems*, 2016, pp. 3597–3603.
- [20] M. D. Gall, J. P. Gall, and W. R. Borg, *Educational Research: An Introduction*. Pearson, 2003.
- [21] C. McNamara, "General Guidelines for Conducting Research Interviews." [Online]. Retrieved 06-Feb-2019 from <https://managementhelp.org/business-research/interviews.htm>.
- [22] M. Patton, "Qualitative evaluation and research methods," in *Qualitative evaluation and research methods*, 2nd ed., Thousand Oaks, CA: Sage Publications, Inc., 1990.
- [23] M. N. Marshall, "Sampling for qualitative research," *Fam. Pract.*, 1996.
- [24] S. Kvale, "Introduction to interview research," in *Doing Interviews*, 2007.
- [25] D. R. Thomas, "A General Inductive Approach for Analyzing Qualitative Evaluation Data," *Am. J. Eval.*, 2006.
- [26] C. P. Santos, J. van de Haterd, K. Hutchinson, V. J. Khan, and P. Markopoulos, "GURaaS: An end-user platform for embedding research instruments into games," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2017, vol. 10303 LNCS, pp. 34–48.
- [27] B. G. Glaser and A. L. Strauss, "The discovery of grounded theory," *Int. J. Qual. Methods*, 1967.
- [28] F. de Boer, "The Grounded Theory Approach: updated," *Kwalon*, 2011. [Online]. Available: https://www.tijdschriftkwalon.nl/inhoud/tijdschrift_artikel/KW-16-1-25/De-Grounded-Theory-Approach-een-update. [Accessed: 07-Feb-2019].

- [29] N. K. Gale, G. Heath, E. Cameron, S. Rashid, and S. Redwood, "Using the framework method for the analysis of qualitative data in multi-disciplinary health research," *BMC Med. Res. Methodol.*, 2013.
- [30] J. Isaak and M. J. Hanna, "User Data Privacy: Facebook, Cambridge Analytica, and Privacy Protection," *Computer (Long. Beach. Calif.)*, 2018.
- [31] European Union (EU), "EU GDPR Information Portal," Retrieved 06-Feb-2019 from <https://www.eugdpr.org/>



Chapter 6

Effects of advertisements and questionnaire interruptions on the player experience

Abstract

New online stores and digital distribution methods led to the development of alternative monetization models for video-games, such as free-to-play games with advertisements. Although there are many games using such models, until now the effect on the player experience from such interruptions has not been studied. In this controlled experiment, we requested that participants (N=236) play one of three different versions of a platformer game with: 1) no interruptions, 2) 30-second video advertisements, and 3) a multiple-choice questionnaire. We then evaluated the effects on the player experience. The study shows differences in their experiences, namely in: competence, immersion, annoyance, affects, and the reliability of the questionnaire answers. The contribution of this work clarifies for game developers which player experience variables are affected by interruptions, and supports them in terms of designing a better experience for their players.

Summary:

Chapter 6 measures the impact of in-game interruptions and its effects on player experience, since questionnaire-based interruptions were hypothesized as generating strong negative experience effects in previous chapters.

Accepted in:

IEEE Conference On Games 2019, London, UK

Authors:

Carlos Pereira Santos, Niels van Gaans, Vassilis-Javed Khan, Panos Markopoulos

Introduction

Games are a thriving industry, with 2.3 billion gamers across the globe. The industry is projected to grow a further 13.3% by the end of 2018, reaching revenues of \$137.9 billion [1]. With most game companies adopting the practice of producing only downloadable games and moving away from traditional means of distribution, games have become more accessible than ever. Present-day consumers have access to hundreds of thousands of games for free; therefore, they are less willing to spend money on games than they were before [2]. Hence, companies have had to adopt free-to-play business models in order to maintain revenue streams.

As a result, the “freemium” or “free-to-play”, as it is known in the game industry, has become an increasingly relevant business model. Out of a 2017 confirmed revenue of \$108.4 billion, \$82 billion ($\approx 76\%$) came from free-to-play titles for both PC and mobile platforms [3]. The freemium business model provides users access to services or goods without requiring them to pay, although additional features within the service can be monetized. Within this model, companies may use multiple strategies to earn revenue. Two main sources of revenue are: 1) eliciting players to pay for extras in the games, e.g. cosmetic elements, time or content; 2) allowing third-party entities to embed external content, such as ads or questionnaires, for a flat rate per click or view of third-party content.

This paper focuses on games that utilize interruptions to display third-party content, i.e., the game is suspended completely, and the player is “forced” to consume incongruent external content. Most of the time, the interruptions are displayed intermittently between levels, but there are several games which use a much more visceral approach and interrupt the gameplay actively to display all sorts of content, including short videos (from 10 up to 45 seconds long, see Figure 6-1), questionnaires (e.g., PollFish or Survata), or even other interactive mini-games.

Although there are less intrusive advertisement strategies, such as in-game advertisements, i.e., placing brands and products directly in the game level or scene like: props, billboards, or street advertisements [24]. Those normally require more involvement from the game company, and not all games are able to accommodate such advertisement strategies, due to incompatibility of genres, violence driven games, lack of proper placement opportunities, or too much an alternative story/world to associate with real world brands. Lastly, creating such an intertwined relationship between the advertised brand and the game may not be desirable; hence, such methods are rarely used, and the study we report on in this paper will not address those advertisement strategies.

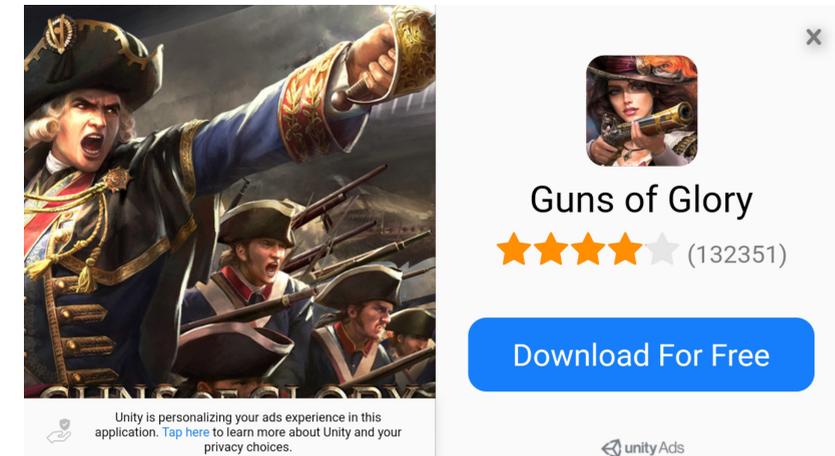


Figure 6-1: Final frames of a typical commercial interruption that showcases a 30-second video and requests that the player download the game (big blue button) or dismiss the advertisement (grey cross on the top right corner).

Even with such a strong market presence in the industry, interruption-based free-to-play games are still a bit of a black art in the sense that there are still a lot of assumptions being made. The few research studies covering this topic seem to contradict dogmatic opinions. For example, meta-critic scores seem maladjusted to free-to-play games [4], or developers’ attitudes towards free-to-play are not as negative as initially thought [5]. Those interruptions are also used for Games User Research [39] focused on the measurement of any aspect of influence the experience and perception of video games, which may influence the intended measurement.

It is our goal with this paper to address an existing gap; support game companies, designers, researchers; and study the relationship between the game experience and non-congruent interruptions. By comparing the same game under three different conditions, we take a deeper look into how interruptions might affect the gaming experience.

Background

It is fundamental that game designers consider deeply the experience that they are providing; hence, the player experience is a well-explored topic. There are multiple models and frameworks which focus on player enjoyment or satisfaction. Of these models, probably the most well-known is flow, defined by Nakamura & Csikszentmihalyi [6] as “the state in which people are so involved in an activity that nothing else seems to matter, people will even do it at great cost, for the sheer sake of doing it”.

Relative vs. Absolute Ordinal Measurements

The arguments to be made on how to represent experiences since they are linked with an individual affective context, which includes the cultural setting, edification and history of the individual [36].

Therefore, by requesting a participant to assigning absolute ordinal values to represent an emotional behavior, for example in seven-point Likert scale, will bring their individual context bias, which is not easily avoided since is part of their conceptual model. Hence, it may be arguable that relative values may introduce a comparison base line [36]. Relative scales request users to rate or compare two or more cases with regards to each other, and able to measure the anchoring the information provided and able to measure the distance between presented cases.

Although absolute values are not perfect, the statistical analysis is able to calculate associated error for that sample population. Noteworthy is the size and representability of the population. Although, there are great advantages of relative measurement approach, there also important drawbacks, namely the anchor cases are required to be evaluated, and not easily transferable or comparable between the design experiment and other cases. In short, absolute ordinal values, allow you to compare two distinct experiences independently of each other, while relative values require you to compare all cases with each other.

Measuring Game Experience

Although flow is very relevant and studied deeply in terms of its relation to games [7–9], there are multiple psychological elements that may influence whether or not gaming is perceived of as a joyful and entertaining activity which translates into a positive experience. The most-known elements include notions of presence (the subjective feeling of being in the virtual environment), challenge (the match between the skills of the player and the obstacles to be overcome), immersion (the natural attention-grabbing properties of the game), and positive or negative affects (such as relief or frustration).

Although it is outside of the scope of the paper to explain the different methodologies with which the player experience can be measured, we would like to provide reference to the most known methodologies in order to justify the selection of the Game Experience Questionnaire (GExpQ) that we used.

Game Flow

The GameFlow model [7] proposes to explain the enjoyment in games by expanding on the basic concept of flow [6] by adding a social component. The eight dimensions of GameFlow are: Concentration, Challenge, Player Skills, Control, Clear Goals, Feedback, Immersion and Social Interaction.

Heuristics of Play

Proposed by Desurvire and Wiberg [10], the tool Game Usability Heuristics is used for the evaluation of game experiences. The heuristics incorporate usability principles for game design and aid producers in their processes. The evaluation form is constructed using three main categories: 1) Gameplay, which includes dimensions from flow [6]; 2) Emotional Immersion, which covers the emotional aspects via an SCI (Sensory, Challenge, and Imagination) model [11]; and 3) Usability and Mechanics, which covers requirements for a usable system and is similar to the heuristics of Nielsen [12].

Game Engagement Questionnaire

The Game Engagement Questionnaire's (GEngQ's) [13] purpose is to measure the engagement potential, and it consists of 19 self-reported items that measure the subjective feelings of immersion, presence, flow, and absorption. In contrast to flow [6], negative effects are considered, such as anxiety and frustration. In addition, the authors also propose an ordered structuring of the experience elements, for example what is more likely to achieve immersion rather than flow or absorption.

Game Experience Questionnaire

The Game Experience Questionnaire (GExpQ) [14] is similar to the GEngQ in the sense it is a tool for measuring subjective game experiences. GExpQ uses a modular approach that covers: core gameplay, social presence, and post-gameplay experiences. GExpQ includes the majority of factors present in GameFlow [7] and the GEngQ [13]. In addition, it divides absorption into individual components by providing more refinement (tension/annoyance, positive and negative affects). In a unique move, this model splits the in-game and post-game experiences. The authors provide extensive documentation, including internal correlations between game experience constructs and demographic factors that influence the measurements, making the GExpQ a robust and frequently used measurement tool for game experiences.

Core Elements of Gaming Experience (CEGE)

The Core Elements of Gaming Experience (CEGE) [15] assesses interaction with the game as a facilitator of experience, thereby positioning it between the GExpQ and Heuristics of Play. The CEGE has two main categories, namely, Puppetry (assessing player control and how responsive the game is to player actions) and the Video Game (assessing content, mechanics, and environment).

Interrupting an Experience

Although a scarce amount of research has been done on interruptions within a game context, the disruptive effects of interruptions have been researched in a task-oriented context. Despite the contextual differences, a structural similarity can be found between the interruptions in the two environments (shown in

Figure 6-2 [16]). In both cases, interruptions require the user to switch focus between multiple tasks, thereby affecting the execution of the primary task.

Early work showed that interruptions affect behaviour and memory performance [17]. Specifically within a gaming context, Gillie and Broadbent showed there are negative performance effects from interruptions [18], they found that nature (similarity) and the complexity of the interruption are determinant in making the interruption disruptive, while control over when player is interrupted and length are less important factors. Other research [19] suggests that the interruption timing (e.g., immediate or scheduled) has a distinct effect on task performance in and of itself (e.g., task accuracy, promptness, and completeness).

Although the game and advertisement tasks are clearly disjoint, in the sense that the context, cognitive mental model, and required actions (playing vs. consuming) are unrelated, there is clear evidence they influence each other.

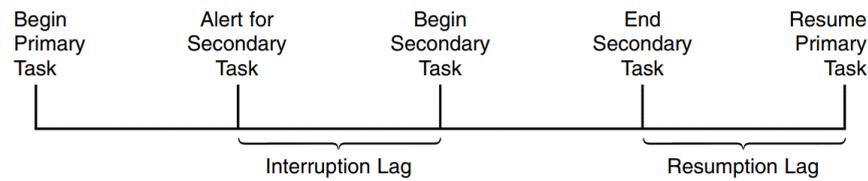


Figure 6-2: The interruption and resumption process involving tasks [16].

Several studies show this crossover: for example, video game difficulty influences the effectiveness of the in-game advertising, i.e., increasing game difficulty affects the processing and evaluation of the brands [20]. Another study showed that levels of video game violence may influence the effectiveness of brand recognition, recall, and positive attitude [21]. And yet another quasi-experimental survey [22] looked into schema incongruity to understand if advertisements in a massively multiplayer, online role-playing game (MMORPG) would change brand awareness rates, and it concluded that moderately incongruent advertising in an MMORPG attains higher awareness rates, although extremely incongruent in-game advertising can reduce the perceived sense of realism of a game and annoy players.

Ad Intrusiveness

A commonly used indicator for advertisement disruption and averting is perceived intrusiveness [23–25]. The main focus of perceived intrusiveness is to understand whether the advertisement evokes personal attitude changes in the consumer after exposure [26]. Although intrusiveness is related to traditional media, the concept can be applied to advertisements in games. Some research has been done on perceived advertisement intrusiveness in games [27]; yet, the causes of perceived intrusiveness remains rather unexplored.

There is a defined research tool which enables the measurement of the perceived intrusiveness of an advertisement (see Figure 6-3) [26]. The combination of measured intrusiveness and player experience may provide a deeper look at, and support validation on, the effects of game interruptions. Although this is a known and frequently used tool specifically designed for advertisements, there is no similar or comparable measurement tool which focuses on research questionnaires, which we are also focusing on in this study.

When the ad was shown, I thought it was....

	1	2	3	4	5	6	7
	<i>Strongly Disagree</i>						<i>Strongly Agree</i>
Distracting	1	2	3	4	5	6	7
Disturbing	1	2	3	4	5	6	7
Forced	1	2	3	4	5	6	7
Interfering	1	2	3	4	5	6	7
Intrusive	1	2	3	4	5	6	7
Invasive	1	2	3	4	5	6	7
Obtrusive	1	2	3	4	5	6	7

Figure 6-3: Example of an Ad Intrusiveness questionnaire [26].

Method

To investigate more closely how interruptions influence the game experience, we defined a between-subjects experiment in which the independent variable was the interruption setting of a video game, which was then followed by the GExpQ [40] to measure and evaluate the player experience under different conditions.

We excluded the usability-based methodologies (Heuristics of Play and CEGE) since we wanted to focus on the players' experiences. A very recent work from Law, Brühlmann and Meke re-evaluated the reliability of GExpQ and found some reliability problems, mainly to the Challenge and Negative Affect components [38] and it is advised to perform Cronbach's α , also included in the results.

We altered an existing game – Runner (see Figure 6-4) – that we developed for another study [31]. Runner is a skill-based platformer game in which the player is represented by a simple red square and the goal is simply to reach the end of the level (denoted by a green area) in the least amount of time possible by jumping onto and grabbing the available objects.

The game was intentionally designed to be a barren, 2D platform with a negligible narrative. Such a set-up exposes players to the core game mechanics only and avoids strong influences, such as a plot, character self-representation, or empathy. For the same reason, we kept the graphics simple and purposely did not add any (cartoonish) avatar to the game.



Figure 6-4: Runner game footage portraying the player's character (red square) and providing two possible paths to reach the end of the level.

For this study, the players had to play the first three levels of the game. The first level introduces visual clues that teach the game to the novice players, and subsequent levels increase the level of difficulty. All three levels took our participating players between 5 to 6 minutes to complete. Nevertheless, we recorded high variability ($M=330.7$ seconds, with a standard deviation of 333.2 seconds), which we deemed natural for a skilled-based game.

Specific performance metrics were tracked for all participants, namely: *i*) the score of the level, i.e., time taken on the first successful attempt to reach the end of the level; *ii*) total time, i.e., the total time a player spent in a level, including all failed attempts, and *iii*) number of attempts, i.e., the number of attempts required to complete the level.

Experimental Conditions

The participants were split across three groups in which they played exactly the same game, with the sole difference among the groups being the nature of the interruption.

In this section, we present the different experiment group conditions in more detail.

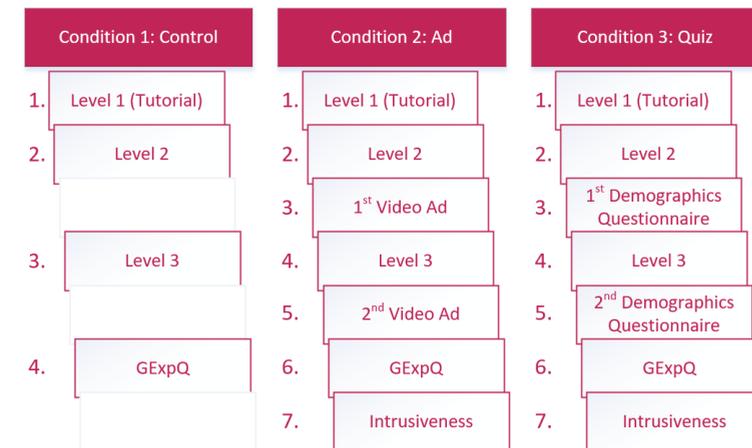


Figure 6-5: Visual representation of the different activities that the participants in each group performed.

Condition 1: Control Group

In the first condition (see Figure 6-5), the players go through all three levels, and, at the end of each level, a level-completion screen with the score (traversing time) is displayed (see Figure 6-6). If a player is not able to transverse a level successfully, they will restart from the beginning of the level. In this condition, there are no interruptions, and, after participants have successfully completed all levels, they are forwarded to an online, self-reported GExpQ [14] in order for the study to establish a comparison baseline.

Condition 2: Advertisement Group

In the second condition (see Figure 6-5), immediately after a player finishes the second and third levels, a distinct video ad is played. These two distinct video ads mimic patterns found in existing games by not allowing the player to dismiss or interact with the game until the ad ends. At the moment the ad does end, a cross appears on the upper left corner, requiring the player to dismiss it manually. Only after the ad is dismissed by the player will the level-completion screen (see Figure 6-6) display, similar to Condition 1.

In order to mimic real-life conditions, both videos were real ads for other games that were exactly 30 seconds long. Although there are many advertisement formats, ranging from small banners to fully interactive mini-games, we selected a common advertisement format present in games at the time of this study.

Similar to Condition 1, after players complete all levels successfully and view the ads, they are forwarded to an online, self-reported GExpQ [14], followed by the ad intrusiveness questionnaire [26], which requests that they evaluate the second (i.e., last) video advertisement that they saw.



Figure 6-6: Level-completion screen presented to the players with their level time score (in seconds).

Condition 3: Questionnaire Group

In the third condition (see Figure 6-5), as in the previous condition, immediately after a player completes the second and third levels, a research-like questionnaire is presented to the player. To improve readability, we will refer to it as the Demographics Questionnaire in order to avoid confusing it with the other research questionnaires we are using.

The Demographics Questionnaire questions were directly copied from the statistical bulletin from the UK Office of National Statistics on Internet access – households and individuals from 2017, licensed under the Open Government License v3.0 [32].

To improve the questions reliability of the Demographics Questionnaire, the UK Office of National Statistics kindly supplied us with the original questionnaire. Hence, we used the original questions exactly as they were formulated and presented them to our player-participants during the game interruptions. By using a known and validated questionnaire, we can measure the reliability of an in-game research tool by comparing the two sets of questionnaire results

The questions retrieved from the national survey were related to what online sources were used to arrange accommodations or transportation and what types of products were bought or ordered online (see Figure 6-7, which showcases one of the questionnaires that was given to our participants). We made sure to include a reasonable number of questions that participants could respond to in a short amount of time in order for this condition to be comparable, at least in terms of the time span, to the 30-second ad condition.

Figure 6-7: Questionnaire presented to players after the 3rd level under Condition 3 (Demographics Questionnaire group). Questions are courtesy of the UK Office of National Statistics.

Participants

For the experiment, we tried to reach a diverse and large number of participants by using a crowdsourcing service (Prolific.ac), which has been proven to be more representative than a traditional university participant pool [33]. We rewarded each participant matching the UK minimum hourly wage (€9/hour), for completing the experiment. We used the same payment for all three conditions.

Participation in the study was voluntary and done with consent. In addition, participants were informed that their in-game behaviour was going to be monitored, data would be collected, and that all collected information would be used purely for research purposes. In addition, they were told that all published information would keep the participants anonymous and that they could refuse to participate or withdraw at any time.

We recruited 236 participants from the United Kingdom in line with the target demographics of our third condition, and these participants self-reported that they play games actively an average of 8.5 (SD=3.9) hours a week. Through the detailed demographic profiles were provided by the crowdsourcing service, we found that the average age of our participants was 30.4 (SD=8.6), with 30.1% being female, and almost all were Caucasian.

The average experiment duration was slightly above 8 minutes (M=495.5 seconds; SD=414.9 seconds). In total, 212 participants completed all steps and weren't excluded because they took too long and didn't fit within the outlier bounds.

Results

In this section, we present and compare the results of each condition. After applying the aforementioned exclusion criteria, we had 70, 72, and 70 participants per condition, respectively. Based on the recommendations of [38], we analysed the internal consistency using Cronbach's α for each condition and the obtained values are considered acceptable for all conditions. The Control condition got 0.7071, the Advertisement condition 0.7348 and Demographic Questionnaire 0.7705.

Game Performance

The statistical analysis of all tracked performance metrics, namely: *i)* the score of the level, *ii)* total time, and *iii)* number of attempts didn't show any significant differences between the conditions.

Game Core Experience

In Table 6-1, we display the results of the statistical values for the game experience core variables, and, through the one-way ANOVA, it is possible to observe that there is statistical significance for multiple core variables, namely: competence, immersion (or, more accurately, Sensory and Imaginative Immersion, according to the GExpQ), tension/annoyance, negative affect, and positive affect.

We also present a graphic-based illustration (see Figure 6-8) to show more clearly how this game experience was affected by the interruptions. Of all the variables, only challenge seems to be unaltered, which makes sense since the game was not altered per condition. Although the means show some difference in terms of how players experienced flow, according to the ANOVA result, the difference is not significant.

All of the other game experience variables registered significant or highly significant differences among the three conditions. Interestingly, and looking at the overall picture, it seems that the advertisement created stronger differences when compared to the Demographics Questionnaire. In theory, the Demographics Questionnaire requires higher actions and cognition of players by forcing them to answer a questionnaire; still, it seems less disruptive of the player experience when compared with the control condition, particularly concerning the negative affect and immersion variables.

It is clearly observed (see Figure 6-8), that both the demographic questionnaire and ad player's tension/annoyance was higher than the control, while the positive affect drops for both conditions. Another very interesting result is that the player's perception of his/her own competence was altered under both Conditions 2 and 3.

Variable	Competence ^a	Immersion ^a	Flow	Tension / Annoyance ^a	Challenge	Negative Affect ^a	Positive Affect ^a
Condition 1 - Control (N= 70)							
mean	0.724	0.561	0.756	0.481	0.733	0.458	0.798
std dev	0.223	0.232	0.223	0.251	0.208	0.207	0.249
std error	0.027	0.028	0.027	0.030	0.025	0.025	0.030
Condition 2 - Ads (N=72)							
mean	0.607	0.472	0.681	0.653	0.701	0.554	0.673
std dev	0.224	0.192	0.244	0.322	0.230	0.239	0.260
std error	0.027	0.023	0.029	0.038	0.027	0.028	0.031
Condition 3 - Demographic Questionnaire (N=70)							
mean	0.617	0.523	0.742	0.593	0.729	0.477	0.724
std dev	0.252	0.223	0.250	0.280	0.225	0.195	0.265
std error	0.030	0.027	0.030	0.034	0.027	0.024	0.032
ANOVA							
F	5.396	3.036	1.954	6.488	0.447	3.921	4.131
Sig.	0.005 ^a	0.050 ^a	0.144	0.002 ^a	0.640	0.021 ^a	0.017 ^a

^a Significant result

Table 6-1: Statistical analyses of the normalized Game Core Experience Variables, including means, standard deviations, and standard errors of the means, for all three conditions as well as one-way ANOVA.

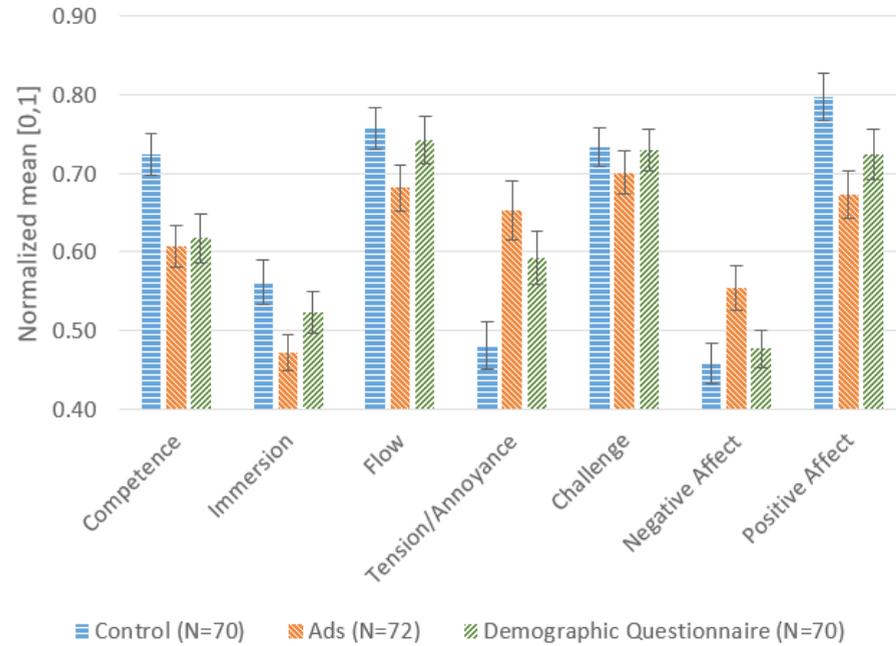


Figure 6-8: Plot of the mean value and the standard error for the core experience variables for the comparison between conditions.

Game Post Experience

The statistical results of the Game Post Experience are presented in Table 6-2 and illustrated in Figure 6-9. Overall, there are some differences in the means between the conditions, but according to the ANOVA results, only positive experience shows a significant difference in the expected direction.

Perceived Intrusiveness

As in the previous section, in Table 6-2 and as illustrated in Figure 6-10, it is possible to see the result of the statistical analysis of the perceived intrusiveness questionnaire [26]. Since the perceived intrusiveness questionnaire only makes sense when interruptions are present, they only cover the Ad and Quiz conditions, and as explained above, the perceived intrusiveness questionnaire was adapted directly to Condition 3, but, to our knowledge perceived intrusiveness has never been used to evaluate research questionnaires. However, the t-test analysis produced highly significant results, i.e., very different means between Conditions 2 and 3.

Variable	Positive Experience ^a	Negative Experience	Tiredness	Return to Reality		Perceived Intrusiveness ^a
Condition 1 - Control (N= 70)						
mean	0.620	0.358	0.379	0.375		n/a
std dev	0.187	0.130	0.174	0.159		n/a
std error	0.023	0.016	0.021	0.019		n/a
Condition 2 - Ads (N=72)						
mean	0.515	0.402	0.405	0.346		0.844
std dev	0.226	0.163	0.233	0.145		0.226
std error	0.027	0.019	0.028	0.017		0.027
Condition 3 - Demographic Questionnaire (N=70)						
mean	0.584	0.392	0.411	0.388		0.577
std dev	0.244	0.157	0.206	0.162		0.256
std error	0.029	0.019	0.025	0.020		0.027
ANOVA					t-Test	
F	4.085	1.668	0.474	1.342	t	6.556
Sig.	0.018 a	0.191	0.623	0.264	Sig.	0.000000001 a

^a Significant result

Table 6-2: Statistical analyses of the Normalized Game Post-Experience Variables, including means, standard deviations, and standard errors of the means, for all three conditions as well as one-way ANOVA and Statistical analysis of the Perceived Intrusiveness Variable, including mean, standard deviation, and standard error of the mean for conditions 2 and 3, plus a two-sample t-Test.

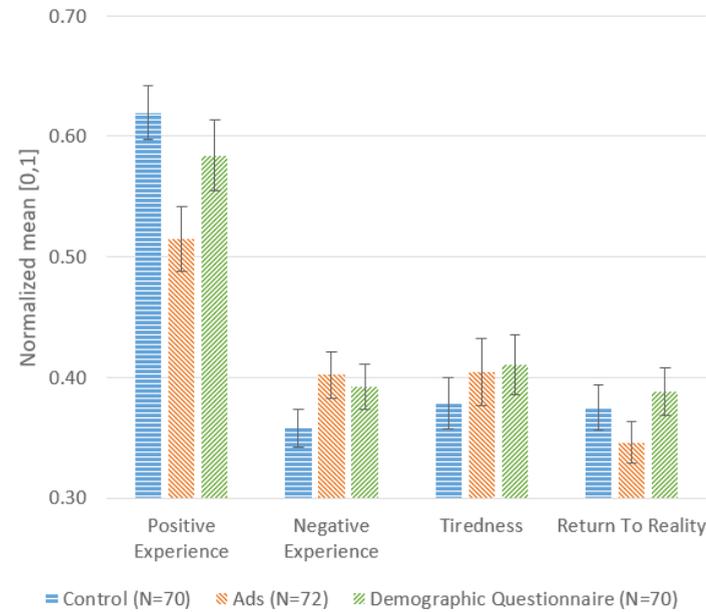


Figure 6-9: Plot of the normalized mean value and the standard error for the post experience variables for the comparison between conditions.

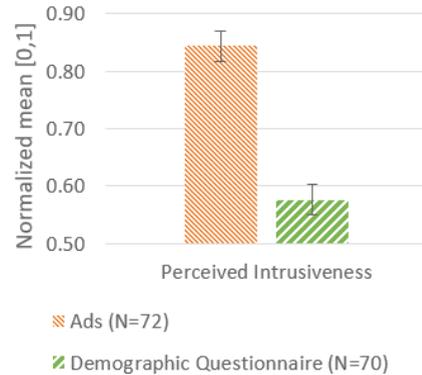


Figure 6-10: Plot of the normalized mean value and the standard error for the perceived intrusiveness comparison between the ads and demographic questionnaire conditions.

Reliability of the Demographic Questionnaire

We purposely limited the participant pool to the United Kingdom and used an existing research study with a highly reliable statistical analysis to compare the results obtained in the demographics questionnaire. Below (see Figure 6-11 and Figure 6-12), a side-by-side comparison of the questionnaire outcomes is given.

In both figures, there is a noticeable difference between the national results in comparison with the results obtained within the game. Although the results seem relative congruent in terms of order, i.e., the different options are equivalently ordered in relation to each other, there is a large disparity between the values obtained by the national survey and those obtained within the game.

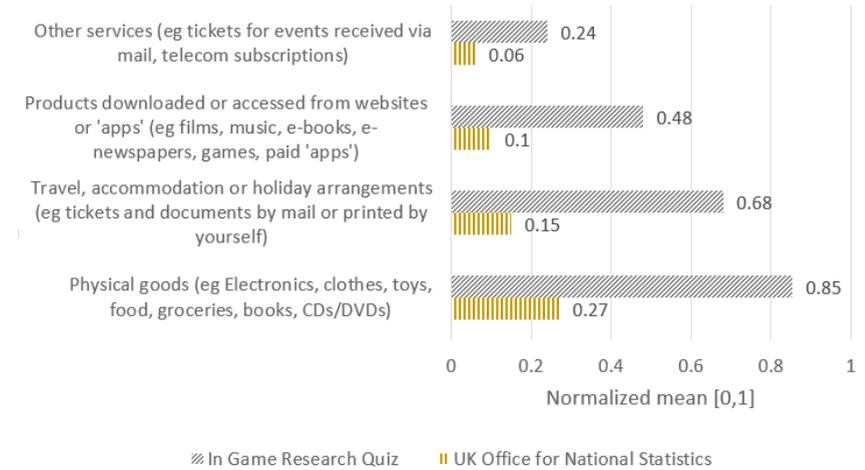


Figure 6-11: Product bought or ordered over the internet from outside the UK, comparison between the normalized results from the UK Office for National Statistics and the Research quiz placed within Runner.

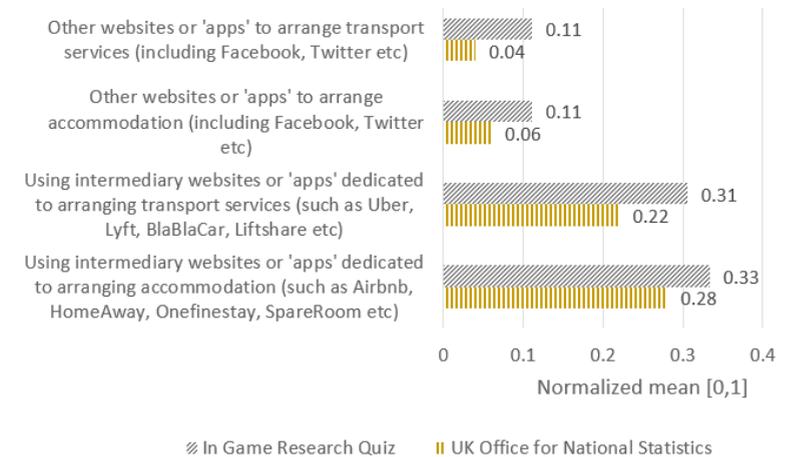


Figure 6-12: Use of Internet to arrange accommodations or transport via another individual, comparison between the normalized results from the UK Office for National Statistics and the Research quiz placed within Runner.

Discussion

The results show clearly that there are significant alterations to the game experience variables when interruptions are introduced. The games with interruptions show a decrease in perceived competence, immersion, and the positive affect, while there is an increase in the negative affect and tension/annoyance. In the game post-experience, it was also clear that the players had a better positive experience without interruptions; this is particularly important because it may lead to higher impressions of the game. From a product relationship and marketing point of view, better positive experiences may lead to improved product satisfaction, public ratings, and recommendations.

Detailed Evaluation of the Game Experience

When taking a detailed look at the experimental results and different conditions, there are some notable results, which we took a closer look at in this section.

Flow

Although, in Table 6-1, there is a noticeable difference between the flow means, the ANOVA was not significant and did not allow us to prove that player flow is affected by interruptions. Despite our interruptions being performed rather abruptly after the player had terminated a level, they did not interrupt the gameplay, and therefore perhaps minimizing their impact. We speculate that the timing of the interruption suggested a scheduled interruption [19], which may have reduced the disruption of player flow.

This result is extremely relevant for game developers in terms of understanding under which circumstances the game should be interrupted and minimize game flow interference.

Competence

There is a significant difference in competence; players felt less competent in the conditions with interruptions. At first glance, this seems a curious result since the players didn't notice a difference in the challenge while the self-perception of competence was lower.

A task interruption study conclusively found that performance decreases with the introduction of interruptions [18], which led us to extend our analysis and analyse the different times/scores of the groups, and we concluded that there wasn't any meaningful variation on the means.

We hypothesize that the interruptions may incite players to reflect on their own performances, which might lead to lower self-perceptions of competence. We advise that this lead should be analysed by a follow-up qualitative experiment.

Advertisement vs. Demographic Questionnaire

Although the in-game advertisement is a known and frequently used model and players have an acceptance attitude towards in-game advertisements [34], the study shows clearly that game experience variables for the advertisement condition diverged from the control variables more than the Demographics Questionnaire condition variables did. They showed higher impacts on immersion, tension, negative affect, positive affect, and positive experience. Similarly, players found the advertisement more intrusive (shown in Figure 6-10).

If we consider both activities (watching an advertisement video vs. answering research questions), in theory, the research questions should be more challenging and disruptive, i.e., they require more cognition, an explicit action for answering (checking the correct box), and may have privacy issues to consider. Still, the advertisement, which is a passive experience and one that can be completely ignored by the player, was clearly more disruptive.

We can only speculate why without a proper and grounded follow-up study, but our theory is that the advertisement has an explicitly commercial focus and completely breaks with the player context with a different graphic style and soundtrack, perhaps causing a consumer backlash [35].

Reliability of In-Game Questionnaires

We purposefully used a known and reliable study in order to attempt to replicate similar conditions and evaluate the reliability of performing an in-game questionnaire.

The results were mixed. It seems that the results show relative accuracy, i.e., the results' ordering and distribution are well-balanced (ratio between two results) for both questions (Figure 6-11 and Figure 6-12). In addition, for the second question, the results for the in-game questionnaire are rather close to the national office statistics; however, the results of the first question are completely different (with a difference up to 58%). Note that if players had randomly select items on the list, we would see an even distribution of all results. In addition, it would be unlikely that they would follow the same distribution as the results of a national survey.

The second question (about arranging accommodations or transport) shows a clear connection to the national results, but what caused such a big difference in the first question? The difference might have been due the selection of the target audience. Because we used online crowdsource workers and, by the nature of their work, they spend a lot of time online, and they may be more inclined to buy products online, independently of their country of origin.

Although we made an effort to target UK players in order to assure congruity in terms of the target audience, we did not consider that the nature of the question could have influenced the outcomes for this target population.

In addition, our questionnaire contained only two questions, which is a very limited sample; therefore, we cannot recommend that these game-embedded research tools be used without further research into this topic, not only to confirm the above identified hypothesis, but also to address other issues such as ordering, carryover effects or psychological influences that the game may have on the research tool since such influences are clearly present in the case of advertisement [20–22].

Summarizing, more research is required but we believe that in-game questionnaires have the potential to provide good insights into players as a target audience but that they cannot be generalized to the general population. Without an appropriate study on the reliability, companies that are using in-game questionnaires services to conduct research should carefully consider the representability of such study, due to the natural segmentation created by the player base of that game.

Implications

This study looked directly into the effects on the game experience caused by advertisement and questionnaire interruptions and identified and quantified the influences of those interruptions on players' game experiences. For game developers and designers, this information might be critical in terms of making informed decisions on how a specific business model will influence the game experience. In addition, it may provide support for those same developers to attenuate those influences by counter-balancing the negative effects through product design, for example, by timing the interruptions or allowing the player to decide when they want to interrupt their game.

The new business model, in which advertisement interruptions are replaced by a questionnaire, is becoming increasingly common. Since the focus of this study was mainly on the game experience consequences, the reliability results are considered a by-product, and, although promising, they can only be considered preliminary work since there is still much more work required in this field.

Limitations and Opportunities

In addition to the follow-up study options mentioned above, we would like to list other possible research opportunities. The target audience was limited to UK players in order to be able to look into the reliability of the demographics questionnaire. Cultural differences might interfere with the results, therefore, a cross-cultural study could analyse this facet in more detail.

There is the opportunity to expand this study by considering other interruption formats. We selected two specific incongruent interruptions: a multiple choice research questionnaire and a 30-second video advertisement. Questions remain as to whether other interruptions would manifest in a similar manner and how the effects of these other interruptions could be minimized.

Although we purposely used a neutral game without external artifices or any known intellectual property to minimize side effects and properly test the game experience, it is not possible to fully generalize this study to all mechanics and genres. We would like to promote the opportunity to redo the experiment with other types of games.

Conclusions

In this work, we examined the effects that two types of interruptions have on the player experience. For that purpose, we designed a between-subjects experiment in which we altered game interruptions and measured the player experience through GExpQ [14]. In our analyses, we found proof that the effects extend beyond tension and annoyance, affecting other variables such as immersion, competence, and positive experience. In addition, the analysis of the reliability of the in-game questionnaires showed strong scaling effects in terms of the answers.

This work contributes directly to the deep understanding of the effects of interruptions on the player experience in games, which, from a product point of view, may influence recommendations and overall product satisfaction.

Acknowledgment

We would like to acknowledge the support of the UK Office of National Statistics for expeditiously supplying us with results and the original questionnaire of their study on Internet access, households, and individuals [32].

References

- [1] T. Wijman, "Mobile Revenues Account for More Than 50% of the Global Games Market as It Reaches \$137.9 Billion in 2018," *NewZoo.com*, 2018. [Online]. Available: <https://newzoo.com/insights/articles/global-games-market-reaches-137-9-billion-in-2018-mobile-games-take-half/>. [Accessed: 12-Sep-2018].
- [2] T. Rayna and L. Striukova, "'Few to Many': Change of Business Model Paradigm in the Video Game Industry," *Digiworld Econ. J.*, vol. 94, no. 2nd Q, pp. 61–82, 2014.
- [3] J. Batchelor, "Games industry generated \$108.4bn in revenues in 2017," *gamesindustry.biz*, 2018. [Online]. Available: <https://www.gamesindustry.biz/articles/2018-01-31-games-industry-generated-usd108-4bn-in-revenues-in-2017>. [Accessed: 12-Sep-2018].
- [4] K. Alha, E. Koskinen, J. Paavilainen, and J. Hamari, "Critical Acclaim and Commercial Success in Mobile Free-to-Play Games," in *Proceedings of 1st International Joint Conference of DiGRA and FDG*, 2016.
- [5] K. Alha, E. Koskinen, J. Paavilainen, and J. Hamari, "Free-to-Play Games : Professionals' Perspectives," *Proc. Nord. Digra 2014 Gotland, Sweden*, 2014.
- [6] J. Nakamura and M. Csikszentmihalyi, "The concept of flow," in *Flow and the Foundations of Positive Psychology: The Collected Works of Mihaly Csikszentmihalyi*, 2014.
- [7] P. Sweetser and P. Wyeth, "GameFlow: a model for evaluating player enjoyment in games," *Comput. Entertain. - Theor. Pract. Comput. Appl. Entertain.*, vol. 3, no. 3, pp. 1–24, 2005.
- [8] J. Chen, "Flow in games (and everything else)," *Commun. ACM*, 2007.
- [9] B. Cowley, D. Charles, M. Black, and R. Hickey, "Toward an understanding of flow in video games," *Comput. Entertain.*, 2008.
- [10] H. Desurvire and C. Wiberg, "Game usability heuristics (PLAY) for evaluating and designing better games: The next iteration," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2009.
- [11] L. Ermi and F. Mäyrä, "Fundamental Components of the Gameplay Experience: Analysing Immersion," *Chang. Views Worlds Play*, 2005.
- [12] J. Nielsen, "10 Usability Heuristics for User Interface Design," *Nielsen Norman Gr.*, 1995.
- [13] J. H. Brockmyer, C. M. Fox, K. A. Curtiss, E. Mcbroom, K. M. Burkhart, and J. N. Pidruzny, "Journal of Experimental Social Psychology The development of the Game Engagement Questionnaire : A measure of engagement in video game-playing," *J. Exp. Soc. Psychol.*, 2009.
- [14] W. a. IJsselsteijn, Y. a. W. de Kort, and K. Poels, "Game Experience Questionnaire," *FUGA - Fun Gaming*, 2013.
- [15] E. Calvillo-Gómez, P. Cairns, and A. Cox, "Assessing the Core Elements of the Gaming Experience," in *Game User Experience Evaluation, Human-Comp.*, R. Bernhaupt, Ed. Springer, Cham, 2015, pp. 37–62.

- [16] J. G. Trafton, E. M. Altmann, D. P. Brock, and F. E. Mintz, "Preparing to resume an interrupted task: Effects of prospective goal encoding and retrospective rehearsal," *Int. J. Hum. Comput. Stud.*, 2003.
- [17] M. B. Edwards and S. D. Gronlund, "Task Interruption and its Effects on Memory," *Memory*, no. 6, 1998, pp. 665–687.
- [18] T. Gillie and D. Broadbent, "What makes interruptions disruptive? A study of length, similarity, and complexity" *Psychol. Res.*, 1989.
- [19] A. Sasse, C. Johnson, and others, "Coordinating the interruption of people in human-computer interaction," *Human-computer Interact. INTERACT*, 1999.
- [20] L. Herrewijn and K. Poels, "Putting brands into play: How game difficulty and player experiences influence the effectiveness of in-game advertising," *Int. J. Advert.*, 2013.
- [21] S.-C. Yoo and J. Peña, "Do Violent Video Games Impair The Effectiveness of In-Game Advertisements? The Impact of Gaming Environment on Brand Recall, Brand Attitude, and Purchase Intention," *Cyberpsychology, Behav. Soc. Netw.*, 2011.
- [22] B. Lewis and L. Porter, "In-game advertising effects: Examining player perceptions of advertising schema congruity in a massively multiplayer online role-playing game," *J. Interact. Advert.*, 2010.
- [23] L. Ha, "Observations: Advertising clutter in consumer magazines: Dimensions and effects," *J. Advert. Res.*, 1996.
- [24] C. Schneider, "Ad Intrusiveness , Loss of Control , and Stress : A Psychophysiological Study," *Proc. 36th Int. Conf. Inf. Syst.*, 2015.
- [25] S. McCoy, A. Everard, P. Polak, and D. F. Galletta, "An experimental study of antecedents and consequences of online ad intrusiveness," *Int. J. Hum. Comput. Interact.*, 2008.
- [26] H. Li, S. M. Edwards, and J. H. Lee, "Measuring the intrusiveness of advertisements: Scale development and validation," *J. Advert.*, vol. 31, no. 2, pp. 37–47, 2002.
- [27] S. Verberckmoes, K. Poels, N. Dens, L. Herrewijn, and P. De Pelsmacker, "When and why is perceived congruity important for in-game advertising in fantasy games?," *Comput. Human Behav.*, 2016.
- [28] B. J. Gajadhar, H. H. Nap, Y. A. W. de Kort, and W. A. IJsselsteijn, "Out of sight, out of mind: co-player effects on seniors' player experience," in *Proceedings of the 3rd International Conference on Fun and Games*, 2010, pp. 74–83.
- [29] B. Zaman, V. Vanden Abeele, P. Markopoulos, and P. Marshall, "the evolving field of tangible interaction for children: the challenge of empirical validation," *Pers. Ubiquitous Comput.*, vol. 16, no. 4, pp. 367–378, 2012.
- [30] S. Gualeni, D. Janssen, and L. Calvi, "How psychophysiology can aid the design process of casual games: A tale of stress, facial muscles, and paper beasts," in *Proceedings of the international conference on the foundations of digital games*, 2012, pp. 149–155.

- [31] C. Pereira Santos, K. Hutchinson, V.-J. Khan, and P. Markopoulos, "Measuring Self-Esteem with Games," in *ACM Intelligent User Interfaces*, 2017.
- [32] O. for N. S. (UK) and C. Prescott, "Internet access – households and individuals: 2017," 2017.
- [33] T. S. Behrend, D. J. Sharek, A. W. Meade, and E. N. Wiebe, "The viability of crowdsourcing for survey research.," *Behav. Res. Methods*, vol. 43, no. October 2015, pp. 800–813, 2011.
- [34] K. Poels, W. Janssens, and L. Herrewijn, "Play buddies or space invaders? players attitudes toward in-game advertising," *J. Advert.*, 2013.
- [35] S. Peterson, "Games With Advertising: The Way Forward," *Business Insider*, 2011. [Online]. Available: <https://www.businessinsider.com/games-with-advertising-the-way-forward-2011-9>. [Accessed: 10-Oct-2018].
- [36] Georgios N. Yannakakis, Roddy Cowie, and Carlos Busso. "The ordinal nature of emotions." In *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*, pp. 248-255. IEEE, 2017.
- [37] ESA. 2018. Essential Facts about the Computer and Video Game Industry 2018. Entertainment Software Association 1. Retrieved 19 Jan 2019 from http://www.theesa.com/wp-content/uploads/2018/05/EF2018_FINAL.pdf
- [38] Effie Law, Florian Brühlmann, and Elisa Mekler. 2018. Systematic Review and Validation of the Game Experience Questionnaire (GEQ) - Implications for Citation and Reporting Practice. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '18)*. ACM, NY, USA, 257-270. DOI: <https://doi.org/10.1145/3242671.3242683>
- [39] Anders Drachen, Pejman Mirza-Babaei, Lennart Nacke. 2018. *Games User Research*. Oxford University Press
- [40] IJsselsteijn, W. A., de Kort, Y. A. W., & Poels, K. 2013. *The Game Experience Questionnaire*. Technische Universiteit Eindhoven

Chapter 7

Conclusion

This thesis took a very close look at capturing, analysing and evaluating players' in-game behaviour, and, in this chapter, it will evaluate the research questions introduced in Chapter 1 based on the summaries of all of the studies detailed in Figure 7-1.

Profiling Need for Cognition Case Study	Data Correlation Research	Ch. 2
Profiling Self-Esteem Case Study	Data Correlation Research	Ch. 2
Profiling Ethical Norms Case Study	Data Correlation Research	Ch. 3
Game User Research as a Service Platform	Design Science	Ch. 4
Player Experience Evaluation within the Games Industry	Semi-Structured Interviews	Ch. 5
Player Experience Effects of Interruptions	Between-Subjects Experiment	Ch. 6

Figure 7-1: Enumeration of the studies performed for the thesis and their associated methodologies

Profiling the Need for Cognition Case Study

In this controlled experiment, I proposed to develop a game which could profile the Need for Cognition through a commonly used game mechanic – hints. Need for Cognition is a personality trait that can predict the effectiveness of different persuasion strategies upon users [17], and a controlled experiment with 188 participants provided empirical evidence that Need for Cognition has a negative correlation with the number of hints players follow during the game.

Not only was Nanobots the first purposely developed game able to predict a psychologic trait, but, in addition it confirmed that is possible to develop a player model through a designed game mechanic while simultaneously defining a player model able to predict the Need for Cognition.

Profiling the Self-Esteem Case Study

In the second case study, I investigated Self-Esteem and how it is linked with in-game behaviour. Because self-esteem is a reflective psychological trait and is not linked to an individual's abilities [1], it provided an interesting design challenge.

By analysing the game metrics of two game versions (98 and 85 participants), I was able to define a new metric – Self-Evaluation Bias – as a measurable variable able to predict the results of the Rosenberg Self-Esteem Scale [24].

By exposing how player self-esteem is linked with in-game behaviour, I was able to increase game designer awareness and demystify some of the influences of self-esteem, for example, players' self-esteem does not influence player choices or affect game performance.

Moreover, the results of these two case studies allowed me to propose a method for designing game mechanics to infer player traits and provide broader recognition of games as implicit measurement tools for players' psychology traits, with potential applications in healthcare.

Profiling the Ethical Norms Case Study

In Chapter 3, a new controlled experiment is presented to measure ethical orientation by analysing the player choices in a non-linear branching narrative presenting several moral dilemmas. This study brought new challenges when compared with the previous ones, namely, the multi-dimensionality aspects of the norms [23].

By using a genetic algorithm, it was possible to calculate the influence of each narrative branch on an initial group of 80 participants and to find strong correlations between player choices and their ethical inclinations with a new cohort of 20 participants.

Some games have a strong narrative intertwined with their game experience, and often writers and designers spend a lot of effort in developing a narrative with strong emotional content. Hence, it is conceivable that ethical norms and the presented evaluation technique can be used as a tool to validate and personalize the player experience and deliver meaningful consequences.

This work also demonstrates how machine learning (namely, genetic algorithms) can be applied to weighing individual game variables in order to develop player models, and, although there are studies that are able to create models of multi-dimensional traits [18], modelling ethical norms based on player choices had not yet been done. In addition, being able to do so, confirmed that designing game mechanics in order to infer player traits may also be possible for more complex traits.

Game User Research as a Service Platform

Chapter 4 reviewed the needs from the stakeholders involved in behaviour player research and detailed a scalable software solution which would enable game developers to collect metrics from their players. Technically, defining a platform able to gather millions of entries in a short time span is a complex feat; however, the designed high-level architecture using a micro-services design pattern can accommodate the required data throughput, in a manner, similar to other existing continuous streaming services.

The in-game recording features allow developers to tailor the telemetry data to their game and company needs while providing analytical tools to evaluate results, for example, to plot graphical information or filter data by different categories, such as player platforms, system performance, player skill, and so forth.

For game studios, which have the capability of performing their own behavioural player analytics, this toolset provides a clear added value since it also empowers internal researchers by allowing them to run experiments, such as embedded surveys within specific sections of a game to perform behavioural research. If game studios do not have the capability of doing such things, then they can willingly share or outsource particular data to third-party institutions, by picking and choosing if and what they are willing to share.

Nonetheless, this separation between game data gathering and the research that can be performed on may have much deeper and more fundamental impacts. It is frequently the case that researchers have difficulties reaching an appropriate number of participants with which to perform all sorts of case studies that require human subjects. The GURaaS platform enables research through games by allowing researchers to embed research instruments in games and target specific audiences linked with in-game behavioural data. Such research would allow us to not only better understand games and experiences, but it has the potential to help us better understand society.

In short, researchers using games as a means to reach the power of the crowd, coupled with the player profiles and models which can be gathered directly by in-game behaviour analysis, is a powerful concept and one of the most important contributions of this thesis, one which defines strongly the focus and direction of follow-up studies.

Player Experience Evaluation within the Games Industry

In Chapter 5, I describe a qualitative inspection of player experience evaluation within the game industry in interviews with 11 different (small to large) companies. The main goal of these interviews was to identify if and how player experience evaluation is being used during the game lifecycle. The results were then used to survey the state-of-the-art tools and methodologies, distinguish between company needs, and understand their take on sharing game data.

It was clear that independent of the size of the company, developers find player experience evaluation a very important topic, and they feel very passionate about it, but this is where the similarities stop. There was a clear distinction between large and small developers in terms of time spent, systematic approaches and methodologies and tools used. Expert skillset and resources are clearly the main gap, leading companies with fewer resources to adapt or develop ad-hoc tools frequently and to use crude or vague methods. Meanwhile, companies with more resources have specialized teams working exclusively on this topic and use an iterative process to hypothesize, test, evaluate and input the findings into the products life cycle.

Generically, the concept of the GURaaS platform as a tool for companies to evaluate player experience was very well received, with some participants clearly engaged and interested in knowing more. The most debated feature was the in-game questionnaire, with some praising its integration, others arguing about its impact on the player experience and multiple participants specifically addressing flow.

When discussing sharing data with external third parties, many were reluctant to share any player data because they considered the data to be a valuable asset for their corporation. Nonetheless, when asked about sharing the data for social or research purposes, almost all small and mid-sized companies were very open to the idea. Some even romanticized the idea of a cooperative venture in which companies could share data and findings among themselves and other trusted institutions.

Player Experience Effects of Interruptions

In both Chapters 4 and 5, the outcomes pointed to the need to examine how in-game questionnaires would affect the player experience [13]. When attempting to compare this intrusion with similar in-game interruptions, namely, advertisement, there was a noticeable lack of literature on the topic.

In this last chapter, the effects of interruptions on the player experience were evaluated by having 236 participants play one of three variations of the same game, with only the interruption condition having been altered, and evaluating the effects on the player experience. The study results proved that there are significant negative effects in the game experience when interruptions are introduced. In addition, flow was one of the few variables that was not influenced by the interruptions, although this was a repeated concern of game developers in Chapter 4.

Above all, it was possible to quantify the disruption, but also it was clearly shown that the 30-second, video-based advertisement had a greater disruption effect than the multiple-choice demographic questionnaire, reducing some of the concerns identified in previous chapters. Obviously, these results do not mean that an individual player will accept such an interruption. Hence,

it is proposed that an Technology Acceptance Model [4] study be performed. Moreover, I also measured the reliability of the in-game questionnaire and found some concerns: one of the results showed pronounced scaling effects, which need to be studied properly in future work to assure the reliability and validity of such measurements [9].

Reflections on Methodology

In this thesis, multiple research methods were applied across the different studies: statistical comparisons, design science, a between-subjects experiment and semi-structure interviews. All of these methods have known advantages and disadvantages; hence, I attempted to balance and use the appropriate methodologies to answer the specific questions that were identified in the different chapters and leading to a unified research goal.

To test the hypothesis of the possibility of developing games to profile player traits, required that we purposely designed and developed games for that purpose and use quantitative measurements to validate them. A post-positivism stance was adopted in the attempt to minimize any possible research effects of biases, and the used philosophical and psychologic constructs like self-esteem, need for cognition, or even the ethical norms, which are generally well accepted theories. Furthermore, it was a common occurrence to consult experts on the various topics to support and identify golden standards or reference methods in the appropriate fields of expertise.

The games used in the study were developed in collaboration with external companies or students. The versions and content were kept consistent across the different groups except when explicitly mentioned, and before the data collection phase, multiple dry runs were performed to assure the experiment would be consistent, and without software faults or other problems. Most of the recruited player participants was done through a crowdsourcing platform to reach higher diversity of the population [30]. In the participant selection criteria, it was required that they play video-games frequently (more than 2 hours per week). They were informed of the fact that data was being collected for research purposes, all logged data was anonymous, and fair compensation was provided.

In later chapters, to address issues that are more contentious and because I was after a deeper examination of current industry practices which required to shift my methodological approach to qualitative research. In addition, for the feasibility analysis of GURaaS, a design science directly looking to a user centric methodology was adopted.

In the end, this is a mixed methods approach, taking a very pragmatist view on the research goals. Further, it included a strong research through design approach, as these studies involved designing and making solutions to very open questions. I think this has been essential in my research trajectory, and also has

driven the direction when the opportunities emerged in relation to the platform in the late chapters.

Revision of Research Goals

We acknowledge that our personality traits are reflected in our everyday behaviour and influence the experiences, actions and social interactions. Hence, it is natural that such behaviour is reflected in the way we play games, even if such video games encourages the enactment of other roles; the developed work that is presented in this thesis explores in depth the existing links between game design, experiences and player models. In this section, I support existing literature by evaluating and summarizing the combined results from the different studies and answering the research questions introduced in the first chapter.

Design games to infer specific player traits

To the question:

To what extent is it possible to purposely design a game which would be able to infer a specific player trait?

I conclude that it is viable to design games to infer specific player traits; such results were presented in three different profiling cases studies (Chapters 2 and 3), in which three new games were purposely designed to infer three different personality traits which had not yet been profiled in the literature.

A systematic methodology with which to link game mechanics to player models was also introduced and used to map different mechanics to traits, even for difficult traits such as self-esteem with its reflective nature. Further, multidimensional ethical norms were able to be inferred through player models. In addition, I introduced a novel use for machine learning (through a genetic algorithm), i.e., it was used to apply weights to player behaviour data in order to increase the significance of the inference.

Player experience evaluation within industry

To the question:

What is the current role of player experience evaluation within game companies?

I concluded that it is a relevant topic in the industry, but one with large differences in terms of methodologies, tools and resources and how they are applied, even if we take into account the fact that the tools and methodologies are influenced by game genre, platform, target audience, and so on.

Frequently, ad-hoc approaches were used to conduct evaluations, especially when companies with fewer resources were involved. However, there were still clear results, improvements to the games, and high returns on investments. Companies were generically reluctant to share data with other than trusted research partners, although the definition of trusted was unspecified in the interviews.

I believe that the industry is aware that there are clear improvements that can be made; hence, some tools are being made available through open lectures at industry conferences, and so forth. On the other hand, companies with more resources are clearly investing in evaluation and using it to create differentiating factors. As an exploratory study, the information gained in the interviews revealed many interesting research topics to expand upon, such as the development of new methodologies, industry cooperation and the usability of tools, to name a few.

Player-focused research through games tool

To the question:

How can we operationalize a tool which would allow games to support player-focused research?

I conclude that is possible to develop a tool which is able to enhance games to support player-focused research. In Chapter 4, the technical architecture of such tool was introduced, and I documented a case study. However, the platform was also successfully used in other studies, namely the Ethical Norms study (Chapter 3) and the study on the effects of interruptions (Chapter 6).

The platform was also presented to game developers (Chapter 5), and almost all companies were interested in and positive in terms of the presented features, although the in-game questionnaires generated some concerns about the effects on the game experience. These effects were the focus of the last study presented (Chapter 6), which showed and quantified clearly the negative effects of such an interruption, which was comparable to in-game advertisement interruptions.

Impacts to the games industry

The impact of being able to design games to profile specific traits is extremely relevant to the industry to close the affective loop [40]. Using manually authoring processes to generate new content, or by using automated processes and procedural content generation can create custom and new game narratives, features and elements linked with specific player preferences and traits to improve the player experience [33].

By having demonstrated the possibility to design games to profile players, opens the door for such games to be applied to other purposes than entertainment. A perfect example is the recent development of Sea Hero Quest provides a means of discriminating healthy aging from genetically at-risk individuals of Alzheimer's disease [34]. In addition, terms like gamification [35] and serious or applied games [36] are being recognized by other fields, and there is a thriving new sector within the games industry sector, mostly dedicated in this field. By pushing and perfecting the proposed method in chapter 2 to designing games to profile player traits have clear impacts for research purposes with direct applications on other fields like social, healthcare and business.

Furthermore, games scenes and levels are becoming increasingly larger and more complex, meaning quality assurance is very hard to perform, since they are becoming virtually impossible to extensively perform manual tests, and automated tools are not able to provide complete coverage of all situations. Hence, game companies are now understanding the power of reaching directly their players and allow them to participate the development process. A good example of this is Tom Clancy's Ghost Recon: Wildlands [37] which clearly incorporated in their product, a toolset to support player participation and communication between players and game designers. It is my belief that tool will be just a start: more elaborate communication mechanisms will allow game companies to gather information from the players.

Obviously that community driven participatory research is not for everyone, but there are great examples of games like EyeWire [38], FoldIt [39] which were able to achieve such status. Game companies with the correct approach and the right messages can drive players and communities to engage in voluntarily support worldwide research endeavours. The implications can be huge, both for the research community, which is able to scale their reach globally; in addition to the directly positive reflection to the game industry finances, marketing strategies, social responsibility and the global image of the gaming community.

Future Challenges

Game Design

There are four main roles in game development: programming, art, management and design. While the first three roles are rather hard but well defined, game design is hard and difficult role to outline [26]. Not only is game design viewed as a multidisciplinary role for product designers, storytellers, testers, marketers, ethnographers, psychologists, and philosophers, but it is also the creative force behind and the vision holder of the game concept.

Game design in a nutshell is the design of the game experience. Experience design [11] is linked with multiple human cognition levels [22]; hence, the need for game designers to better understand human nature in order to allow them to be repeatedly able to create engaging and memorable experiences. The possibility of designing mechanics that create specific player models opens a new door to being able to deeply understand, test and effectively use player behaviour.

The question that still remains is how. How can game designers take full advantage of this new approach? In this thesis, the feasibility and methodologies for designing games to profile player traits were proven. What remains unknown is how this tool will be applied by game designers around the world and how this new technique could be integrated within an already complex role while considering the effect and implications.

Ethical Grounds

The applicability of profiling players through games has great potential, but researching player experiences is only one out of multiple possible applications. If you know a detailed profile of a player, you can influence their behaviour, for example, to increase sales. It is true that defining the ethical line between improving the player experience and profits is hard [6]. In addition, there might be numerous other ways to exploit such sensitive information. Moreover, the fact that such databases exist and are linked with individuals is very controversial, given the possibility that they could be traded, illegally accessed or employ unethically.

In today's society the debate between security and privacy is frequently in the headlines, and government agencies are defining their own strategies on how to regulate [7, 14] or, on the other hand, use them [12, 20]. This subject is also under review in other data science fields, such as Persuasive Technologies [27, 28] or Recommender Systems [16, 31], and, although ethical codes [3, 5, 10] related to data science exist, there is no widely agreed upon ethical norm in the industry.

This thesis explores the potential of analysing players' in-game behaviours in depth and examined some ethical considerations in some of its chapters, but ethics were never tackled as a central discussion point. Still, I hope that this work encourages such discussion, at least of the big and most necessary challenges.

A security and ethical code that could be applicable and widely accepted in the industry would increase trust while assuring that both developers and players are more conscious of and responsible for their actions.

GURaaS Research

I believe that GURaaS has potential and the ability to make a meaningful social contribution, but the concept of allowing entertainment games to be used as research platforms has still not matured fully. As a concept may seem like a utopian view on the topic; therefore, it is important to tackle GURaaS appropriately, not only to strengthen the concept by using research to shape and determine the tool and usage framework. Despite that I tackled some of the salient questions related to GURaaS, still there is a lot of work to be done. During chapter 4, three main actors around GURaaS were highlighted, namely researchers, game companies, and the players. On chapter 5 and 6, we evaluated two of those three actors, by looking into industry needs and practices of tools like GURaaS, and the effects to the player experience by introducing questionnaires within games. Although the work presented in chapters 2 and 3 showed us the needs of the research actors, the platform usage, benefits and the drawbacks GURaaS provides weren't evaluated, like many other topics that require appropriate corroboration.

From the many questions which are still unanswered, and I would like to highlight a few. On the research side: result validation, user experience, confounding effects and abuse prevention. Similarly and in relation to just the game industry, development cycle implications, acceptance, business model, trust and access to results. At last, but not least, on the player side, there are questions regarding: privacy concerns, acceptance, brand engagement and reflection on the social impact of GURaaS.

Final Statements of the Thesis

Design profiling games

Not only is it possible to develop games for the purpose of profiling specific traits in players, but I have proposed a methodical approach for doing so. This work opens opportunities in terms of the development of a new breed of serious games focused on personal traits, and there are several good reasons:

Implicit behavioural analysis

First, explicit self-reported questionnaires have a known drawback, namely, response bias [21, 25], which occurs when individuals skew self-assessed measures, for instance, to improve social desirability. Games define controlled and safe environments, which offer opportunities for individuals to exhibit a range of behaviours freely. In addition, the implicit measurement of behaviour is much harder for participants to deliberately deceive.

Over time

Second, depending on the game, it may offer a recurring temporal dimension, i.e., players can concentrate on measuring tasks more regularly and for longer periods, provided that these are engaging and fun. In addition, players that spend more time playing will find their achievements more meaningful [19, 29], which can be used as positive reinforcement in real life.

Engagement leads to participation

Third, such games can become a channel between the players and organizations. This channel can be a very cost-effective tool for addressing targeted groups, community building, educational purposes, and inviting participants to contribute, be engaged and participate in real life [15].

Games User Research as a Service

Allowing published games to be used as third-party research platforms is an innovative and challenging concept. Therefore, I focused on analysing it through Tim Brown's design- thinking framework, which is used to evaluate innovative concepts [2] (see Figure 7-2).

Desirability

Only now are we becoming aware of how broadly and deeply video games influence society as we know it, not only from the entertainment point of view but also in terms of group and family dynamics, education, cultural and even individual psychology and beliefs.

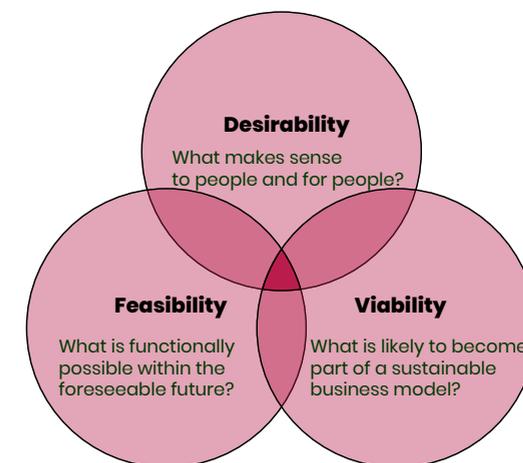


Figure 7-2: Essential criteria for successful concepts when using design thinking [2].

Platforms that empower game developers and researchers with the appropriate tools to enable them to explore player behaviour are a clearly a need. Psychologists, physicians, social workers, teachers, marketers and researchers, in general, would have another solid tool with which to reach and engage specific audiences.

When considering the social impact of such a platform, it is also essential to take into account the strong privacy and ethical concerns mentioned above. Still, those concerns are attached to most of the technology- based products we all are using, and games are and will be part of such a discussion.

Viability

The video game industry has always faced the problem of social responsibility. Like most newly created entertainment-focused industries, it still suffers from severe criticisms in terms of being harmful to society, such as inciting violence, being addictive and causing depression. Despite the efforts of many researchers who have been working on proving or disproving such theories, the game industry is longing to improve its image with society and show corporate social responsibility [8, 32].

Trading player in-game behaviour data in order to increase social responsibility seems to me to be an easy and inexpensive exchange because the data is a by-product created by players which is collected by game companies and, most of the time, eventually deleted because it has served its internal purpose. Moreover, there are already companies reformatting advertisements and replacing it with questionnaires. Plus, when questioned about the willingness to cooperate and share data with researchers, game developers embraced the concept on a generic level. A generic approach where game developers and researchers work together might not be that big of a leap.

Feasibility and revision

In Chapter 4, I demonstrated that not only is it technically feasible to design an online platform that enables games to deliver customizable research instruments, but I have also provided a detailed architecture on how to attain such a platform.

In short, I believe that there are clear and strong arguments in all of the core criteria in the design-thinking framework to support Games User Research as a Service, and I hope to continue to work on and research this topic for years to come.

References

- [1] J. Blascovich and J. Tomaka. 1992. Measures of Personality and Social Psychological Attitudes: Measures of Self-Esteem. *Behavioral & Social Sciences Librarian* 11, 2: 115–160.
- [2] Tim Brown. 2009. *Change by design: How Design Thinking Transforms Organizations and Inspires Innovation*.
- [3] Data Science Association. *Data Science Code Of Professional Conduct*. 12. Retrieved 02/03/2019 from <https://www.datascienceassn.org/code-of-conduct.html>
- [4] Fred D. Davis, Richard P. Bagozzi, and Paul R. Warshaw. 1989. User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. *Management Science*.
- [5] Digital Analytics Association. *The Web Analyst's Code of Ethics*. Digital Analytics Association. Retrieved 02/03/2019 from <https://www.digitalanalyticsassociation.org/codeofethics>
- [6] Magy Seif El-Nasr, Anders Drachen, and Alessandro Canossa. 2013. *Game Analytics: Maximizing the Value of Player Data*.
- [7] European Union (EU). 2018. *EU GDPR Information Portal*. <https://www.eugdpr.org/>.
- [8] Valentino Gandolfi. 2012. *Corporate social responsibility evaluation. Quality - Access to Success*.
- [9] Nahid Golafshani. 2003. *Understanding Reliability and Validity in Qualitative Research*. *The Qualitative Report*.
- [10] Peter Grindrod and Juan Moreno. 2018. *Code of conduct Professional Data Scientist. Code of conduct for professional Data Scientists*. Retrieved 02/03/2019 from <http://www.code-of-ethics.org/code-of-conduct/>
- [11] Marc Hassenzahl. 2005. *User Experience and Experience Design*. *Encyclopedia of Human-Computer Interaction*.
- [12] Celia Hatton. 2015. *China "social credit": Beijing sets up huge system*. *BBC News*.
- [13] W.a. IJsselsteijn, Y.a.W. de Kort, and K. Poels. 2013. *Game Experience Questionnaire. FUGA - Fun of Gaming*.
- [14] Information Commissioner's Office. 2018. *Guide to the General Data Protection Regulation (GDPR)*.
- [15] Martin Johansson and Per Linde. 2005. *Playful collaborative exploration: New research practice in participatory design*. *Journal of Research Practice*.
- [16] Maurits Kaptein and Dean Eckles. 2010. *Selecting effective means to any end: Futures and ethics of persuasion profiling*. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*.

- [17] Maurits Kaptein, Panos Markopoulos, Boris de Ruyter, and Emile Aarts. 2009. Can You Be Persuaded? Individual Differences in Susceptibility to Persuasion. In *Human-Computer Interaction – INTERACT 2009* (Lecture Notes in Computer Science).
- [18] Giel van Lankveld, Pieter Spronck, Jaap van den Herik, and Arnoud Arntz. 2011. Games as personality profiling tools. 2011 IEEE Conference on Computational Intelligence and Games (CIG'11): 197–202.
- [19] Graham McAllister, Pejman Mirza-Babaei, and Jason Avent. 2013. Improving Gameplay with Game Metrics and Player Metrics. In *Game Analytics*. 621–638.
- [20] Mirjam Meissner. 2017. China's Social Credit System: A Big-data Enabled Approach to Market Regulation With Broad Implications for Doing Business in China.
- [21] Thea F. van de Mortel. 2008. Faking it: Social desirability response bias in self-report research. *Australian Journal of Advanced Nursing*.
- [22] Don Norman. 2004. *Emotional Design: Why We Love (or Hate) Everyday Things*. Basic Civitas Books.
- [23] R. E. Reidenbach and D. P. Robin. 1990. Toward the development of a multidimensional scale for improving evaluations of Business Ethics. *Journal of Business Ethics* 9, 8: 639–653.
- [24] M. Rosenberg. 1965. Rosenberg self-esteem scale (RSE). *Acceptance and commitment therapy.*, Measures package 61: 52.
- [25] Robert Rosenman, Vidhura Tennekoon, and Laura G. Hill. 2011. Measuring bias in self-reported data. *International Journal of Behavioural and Healthcare Research*.
- [26] Jesse Schell. 2008. *The Art of Game Design: A Book of Lenses*. Morgan Kaufmann.
- [27] Pp Verbeek. 2006. Persuasive Technology and Moral Responsibility Toward an ethical framework for persuasive technologies. *Persuasive06*: 1–15.
- [28] C Versteeg. 2013. *Ethics & Gamification Design*. Utrecht University.
- [29] Hao Wang and Chuen-Tsai Sun. 2011. Game reward systems : Gaming experiences and social meanings. In *Digital Games Research Association Conference*.
- [30] Jill Weinberg, Jeremy Freese, and David McElhattan. 2014. Comparing Data Characteristics and Results of an Online Factorial Survey between a Population-Based and a Crowdsourced-Recruited Sample. *Sociological Science* 1:292–310.
- [31] Pinata Winoto and Tiffany Tang. 2014. The ethics of a recommendation system. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*.
- [32] Janina Witkowska. 2016. *Corporate Social Responsibility: Selected Theoretical and Empirical Aspects*. Comparative Economic Research.

- [33] Georgios N. Yannakakis and Julian Togelius. 2011. Experience-driven procedural content generation. *IEEE Transactions on Affective Computing* 2.3 (2011): 147–161.
- [34] Gillian Coughlan, Antoine Coutrot, Mizanur Khondoker, Anne-Marie Minihane, Hugo Spiers and Michael Hornberger. 2019. Toward personalized cognitive diagnostics of at-genetic-risk Alzheimer's disease. *Proceedings of the National Academy of Sciences*. National Academy of Sciences
- [35] Zichermann, Gabe, and Christopher Cunningham. 2011. *Gamification by design: Implementing game mechanics in web and mobile apps*. O'Reilly Media.
- [36] Igor Mayer, Harald Warmelink, and Qiqi Zhou. 2016. A frame-reflective discourse analysis of serious games. *British Journal of Educational Technology*.
- [37] Tom Clancy's Ghost Recon: Wildlands. 2017. Ubisoft.
- [38] Sebastian Seung. 2013. *Eyewire*. Eyewire.
- [39] Firas Khatib, Frank DiMaio, Seth Cooper, Maciej Kazmierczyk, Mirosław Gilski, Szymon Krzywda, Helena Zabranska, Iva Pichova, James Thompson, Zoran Popović, Mariusz Jaskolski, and David Baker. 2012. Crystal structure of a monomeric retroviral protease solved by protein folding game players. *Nature Structural & Molecular Biology* 19, 364–364.
- [40] Petra Sundström. 2006. *Exploring the affective loop*, Stockholm University, Technical Report.

Resumo em Português

Compreender pessoas através de jogos

Video jogos são mais que uma industria criativa e global, eles contribuem para a divulgação de diferentes ideias, crenças culturais, pontos de vista éticos e políticos, e são consumidos por bilhões de pessoas no mundo. Só recentemente é que o mundo científico se apercebeu o quanto os jogos estão a ter um impacto não só como entretenimento mas também nas nossas vidas, indústria, sociedade e cultura.

Capturar e analisar informação comportamental dos jogadores é um novo método que está a gerar bastante interesse, que pode ser aplicado para melhorar a experiência ou criar modelos comportamentais dos jogadores. Este trabalho é focado nessa informação comportamental dos jogadores e como se pode melhorar métodos existentes nas áreas de investigação (Games User Research) e modelação comportamental (Player Modelling) do jogador.

Nas primeiras experiências, três video jogos foram criados para modelar diferentes aspectos comportamentais de jogadores, cada com o seu desafio: Necessidade de Cognição é um traço dicotômico de personalidade (alto vs. baixo), Auto-Estima é um traço reflectivo e não ligado à habilidade individual, e o perfil ético do jogador que é multi-dimensional. Através dessas experiências demonstrou-se que é possível desenvolver jogos para captar e representar traços de personalidade dos jogadores através da forma como jogam, e propôs-se uma metodologia para desenvolver esse tipo de jogos.

Os estudos anteriores identificaram a necessidade de soluções técnicas e escaláveis para a avaliação da experiência e modelação dos perfis dos jogadores. Como tal, foi desenvolvida a plataforma GURaaS, com foco em melhorar a colaboração entre investigadores e profissionais da indústria. Os dois estudos seguintes concentram-se em examinar os benefícios e desvantagens da plataforma: no primeiro, identificamos como a plataforma GURaaS poderá ajudar a colmatar as falhas e necessidades no processo de desenvolvimento nas empresas de video jogos, e no segundo estudo, medimos o impacto causado na experiência dos jogadores por serem interrompidos pelas ferramentas de investigação científica.

A contribuição deste trabalho, demonstra o potencial de usar video jogos como instrumentos de investigação científica e demonstra a viabilidade de uma plataforma que facilita a desenvolvimento de jogos para esse fim.

List of Publications

Thesis Related Publications 10

(submitted) Carlos Pereira Santos, Heleen van Geelen, Vassilis-Javed Khan, Panos Markopoulos. 2019. *Player Experience Evaluation in the Games Industry*. ACM Annual Symposium on Computer-Human Interaction in Play (CHIPLAY), Spain

(accepted) Carlos Pereira Santos, Niels van Gaans, Vassilis-Javed Khan, Panos Markopoulos. 2019. *Effects of advertisements and questionnaire interruptions on the player experience*. IEEE Conference On Games 2019, UK

Carlos Pereira Santos, Kevin Hutchinson, Vassilis-Javed Khan, Panos Markopoulos. 2019. *Profiling Personality Traits with Games*. ACM Transactions on Interactive Intelligent Systems Journal (TiiS), Volume 9 Issue 2-3, Article No.: 11, March 2019, ACM New York, NY, USA. ISSN:2160-6455 EISSN:2160-6463, doi: 10.1145/3230738

Carlos Pereira Santos, Vassilis-Javed Khan & Panos Markopoulos. 2018. *Profiling ethics orientation through play*. Behaviour & Information Technology Volume 37, 2018 - Issue 9: Highlights from ECCE 2018, September 2018

Carlos Pereira Santos, Vassilis-Javed Khan, Panos Markopoulos. 2018. *Interactive Narratives for Profiling Ethics Orientation*. 36th European Conference on Cognitive Ergonomics (ECCE 2018), September 2018, the Netherlands

Carlos Pereira Santos, Jeroen van de Haterd, Kevin Hutchinson, Vassilis-Javed Khan, Panos Markopoulos. 2017. GURaaS: An End-User Platform for Embedding Research Instruments into Games. 6th International Symposium on End User Development (IS-EUD 2017), the Netherlands

Carlos Pereira Santos, Kevin Hutchinson, Vassilis-Javed Khan, Panos Markopoulos. 2017. *Measuring Self-Esteem with Games*. ACM Intelligent User Interfaces (IUI 2017), March 2017, Cyprus

Carlos Pereira Santos, Vassilis-Javed Khan, Panos Markopoulos. 2016. *Game User Research as a Service*. ACM CHI Conference on Human Factors in Computing Systems 2016 Workshop (CHI 2016): Lightweight Games User Research for Indies and Non-Profit Organizations, May 2016, USA

Carlos Pereira Santos, Vassilis-Javed Khan, Panos Markopoulos. 2016. *Inferring A Player's Need For Cognition From Hints*. ACM Intelligent User Interfaces (IUI 2016), March 2016, USA

Carlos Pereira Santos, Vassilis-Javed Khan, Panos Markopoulos. 2014. *On Utilizing Player Models to Predict Behavior in Crowdsourcing Tasks*. International Conference on Social Informatics (SocInfo 2014), November 2014, Spain

Other Publications

21

(submitted) Phil de Groot, Wilco Boode, Carlos Pereira Santos, Harald Warmelink, Igor Mayer. 2019. *A Shipping Simulation Through Pathfinding: SEL within the MSP Challenge Simulation Platform*. 20th annual European GAME-ON® Conference (GAME-ON®'2019) on Simulation and AI in Computer Games.

(accepted) Magali Goncalves, Jeroen Steenbeek, Maciej Tomczak, Giovanni Romagnoni, Rikka Puntilla, Ville Karvinen, Carlos Pereira Santos, Xander Keijser, Lodewijk Abspoel, Harald Warmelink, Igor Mayer. 2019. *Food-web modeling in the Maritime Spatial Planning Challenge Simulation Platform: Results from the Baltic Sea Region*. 20th annual European GAME-ON® Conference (GAME-ON®'2019) on Simulation and AI in Computer Games.

(accepted) Jessika Weber-Sabil, Lidija Lalicic, Thomas Buijtenweg, Kevin Hutchinson, Carlos Pereira Santos, Melissen Frans, Ko Koens and Igor Mayer. 2019. *Managing Competing Values in Sustainable Urban Tourism: a Simulation-Gaming Approach*. International Simulation and Gaming Association. (ISAGA 2019), Warsaw, Poland.

(accepted) Igor Mayer, Harald Warmelink, Xander Keijser, Carlos Pereira Santos, Wilco Boode and Magali Gonçalves. 2019. *MSP Challenge simulation platform - demonstration*. International Simulation and Gaming Association. (ISAGA 2019), Warsaw, Poland.

(accepted) Jeroen Steenbeek, Giovanni Romagnoni, Jacob Bentley, Johanna Jacomina Heymans, Natalia Serpetti, Magali Gonçalves, Carlos Pereira Santos, Harald Warmelink, Igor Mayer, Xander Keijser, Rhona Fairgrieve and Lodewijk Abspoel. 2019 *Combining ecosystem modelling and serious gaming to aid transnational management of marine space*. Managing local and global fisheries in the Anthropocene in Ecology and Society Journal

Marnix Van Gisbergen, Hans Ladders, Carlos Pereira Santos, Dennis Hoogervorst. 2018. *How AR visualisation can help us understand data*. Data Driven Digital Marketing Event (DDME2018), 2018, Huizen, Netherlands.

Kevin Hutchinson, Harald Warmelink, Wilco Boode, Carlos Pereira Santos, Igor Mayer. 2018. *An Offshore Energy Simulation through Flow Networks: CEL within the MSP Challenge 2050 Simulation Game Platform*. European Simulation and Modelling Conference (ESM 2018), October 2018, Belgium

Licia Calvi, Carlos Santos, Joey Relouw, Bojan Endrovski, Serena Lucrezi, Antonio Sarà, Ubaldo Pantaleo. 2017. *A VR game to teach underwater sustainability*. The Fifth IFIP Conference on Sustainable Internet and ICT for Sustainability (SustainIT 2017), December 2017, Madeira, Portugal.

Licia Calvi, Carlos Pereira Santos, Serena Lucrezi, Antonio Sarà, Joey Relouw, Bojan Endrovski, Chris Rothwell, Marco Palma, Ubaldo Pantaleo. 2017. *Can a VR game increase awareness in a sustainable marine behaviour?* European Marine Science Educators Association Conference (ESMEA 2017), October 2017, Valletta, Malta.

Erwin Peters, Bram Heijligers, Josse de Kievith, Xavier Razafindrakoto, Ruben Van Oosterhout, Carlos Pereira Santos, Igor Mayer, Max M Louwerse. 2016. *Design for Collaboration in Mixed Reality Technical Challenges and Solutions*. 8th International Conference on Virtual Worlds and Games for Serious Applications (VS-games 2016), November 2016, Spain

Giuseppe Maggiore, Carlos Santos, Aske Plaat. 2014. *Smarter smartphones: understanding and predicting user habits from GPS sensor data*. Procedia Computer Science Volume 34, 2014

Giuseppe Maggiore, Carlos Santos, Dino Dini, Frank Peters, Hans Bouwknecht, Pieter Spronck. 2013. *LGOAP: adaptive layered planning for real-time videogames*. IEEE Conference on Computational Intelligence in Games (CIG2013), August 2013, Canada

José Danado, Eduardo Dias, Teresa Romão, Nuno Correia, Adelaide Trabuco, Carlos Santos, João Serpa, Manuel Costa, António Câmara. 2005. *Mobile Environmental Visualization*. The Cartographic Journal Vol. 42 No. 1, June 2005

Teresa Romão, Nuno Correia, Eduardo Dias, Jose Danado, Adelaide Trabuco, Carlos Santos, Rossana Santos, Antonio Camara, Edmundo Nobre. 2004. *ANTS - Augmented Environments*. Computers & Graphics, Special Issue on Pervasive Computing and Ambient Intelligence - Mobility, Ubiquity and Wearables Get Together, October 2004

Miguel Boavida, Anacleto Correia, Susana Cabaço, Adelaide Trabuco, Carlos Santos and Jorge Pereiros. 2004. *Tools for Experiential Learning: The "E2-LEARNING" Approach*. International Conference on Education (IADAT-e2004) Volume: IADAT-e Technological Issues, July 2004, Spain

José Danado, Eduardo Dias, Teresa Romão, Nuno Correia, Adelaide Trabuco, Carlos Santos, João Serpa, Manuel Costa, António Câmara. 2004. *A Multi-user mobile system to visualize environmental processes*. Proceedings of the 12th International conference on Geoinformatics-Geospatial Information Research: Bridging the Pacific and Atlantic, June 2004, Sweden

José Danado, Eduardo Dias, Teresa Romão, Nuno Correia, Adelaide Trabuco, Carlos Santos, David Araújo, Pedro Duarte, Rui Rebocho, José Palmeiro, João Serpa, Manuel Costa, António Câmara. 2003. *Mobile Augmented Reality for Environmental Management (MARE)*. Eurographics 2003 (EG'03), September 2003, Spain

Carlos Santos, Nuno Correia. 2003. *A Navigation and Registration System for Mobile and Augmented Environments*. Workshop on the Design Specification and Verification of Interactive Systems (DSV-IS 2003), June 2003, Portugal

Miguel Boavida, Rossana Santos, Carlos Pereira Santos, Nuno Correia. 2003. *A Framework for Shared and Personalized Video Based Documents*. The Twelfth International World Wide Web Conference (WWW 2003 Posters), May 2003, Hungary

Nuno Correia, Teresa Romao, Carlos Santos, Adelaide Trabuco, Rossana Santos, Luis Romero, Jose Danado, Eduardo Dias, Antonio Camara, Edmundo Nobre. 2003. *A Flexible Augmented Reality Architecture Applied to Environmental Management*, SPIE Electronic Imaging, January 2003, USA

Teresa Romão, Eduardo Dias, José Danado, Nuno Correia, Adelaide Trabuco, Carlos Santos, Rossana Santos, Edmundo Nobre, António Câmara, Luís Romero. 2002. *Augmenting Reality with Geo-Referenced Information for Environmental Management*. Proceedings of the 10th ACM international symposium on Advances in geographic information systems (GIS'02), November 2002, USA



Carlos Pereira Santos
Research and Development Coordinator
Senior Lecturer in Games

[linkedin.com/in/carlospereirasantos](https://www.linkedin.com/in/carlospereirasantos)



Carlos Pereira Santos was born on 13-08-1977 in Faro, Portugal. He graduated in 2000 from the Universidade Nova de Lisboa winning the best student award from the Software Engineering program. In 2004 he finished a Master in Computer Science in the same university with a dissertation focused on Contextual Information support for Augmented Environments.

Carlos joined the software development industry from 2003 to 2009 where he was responsible for the technical development of several large international solutions in Multimedia and Critical Software for the Air-Force, Navy, and the European Space Agency.

Currently, he is a Senior Lecturer and Research and Development Coordinator in the Academy for Digital Entertainment in the Breda University of Applied Sciences, where he is responsible for the technical development of multiple games and other entertainment projects involving cutting-edge research and technology like Mixed Reality, Games and User Experience.

Since 2012 and parallel to his full time position, he became a PhD student at Eindhoven University of Technology, which the results are presented in this dissertation, focusing on player behaviour, and has several published works in the areas of Player Modelling, User Experience and Artificial Intelligence.