20TH INTERNATIONAL CONFERENCE

ON

INTELLIGENT GAMES AND SIMULATION

GAME-ON® 2019

EDITED BY

Robert Grigg

September 18-20, 2019

Breda University of Applied Sciences
Breda
THE NETHERLANDS

A Publication of EUROSIS-ETI

20TH International Conference

on

Intelligent Games and Simulation

BREDA, THE NETHERLANDS

SEPTEMBER 18 - 20, 2019

Organised by

ETI

Sponsored by

EUROSIS

Co-Sponsored by

Ghent University

University of Skövde

University of Žilina

GODAN

Breda University of Applied Sciences

Breda, The Netherlands

EXECUTIVE EDITOR

PHILIPPE GERIL
(BELGIUM)

EDITOR

General Conference Chair

Robert Grigg
Head of Programming and Senior Lecturer Games
Breda University of Applied Sciences, Breda, The Netherlands

Local Programme Committee

Mata Haggis-Burridge
Professor of Creative and Entertainment Games
Breda University of Applied Sciences, Breda, The Netherlands

Josephine Lappia
Manager Research and Business Innovation, Games and Media
Breda University of Applied Sciences, Breda, The Netherlands

Past Conference Chair

David King, SFHEA
Division of Computing & Mathematics, School of Design and Informatics
University of Abertay, Dundee, United Kingdom

INTERNATIONAL PROGRAMME COMMITTEE

Game Development Methodology

Óscar Mealha, University of Aveiro, Portugal
Jounie Smed, University of Turku, Finland

Gamification and Social Game Mechanics

Liliana Vale Costa, University of Aveiro, Portugal
Sanaul Haque, Institute of Technology Carlow, Carlow, Ireland
Joseph M. Saur, Regent University, Virginia Beach, USA

Physics and Simulation

Graphics Simulation and Techniques
Pieter Jorissen, Karel de Grote Hogeschool, Hoboken, Belgium
Ian Marshall, Coventry University, Coventry, United Kingdom
Marco Roccetti, University of Bologna, Bologna, Italy

Facial, Avatar, NPC, 3D in Game Animation
Yoshihiro Okada, Kyushu University, Kasuga, Fukuoka, Japan
Marcos Rodrigues, Sheffield Hallam University, Sheffield, United Kingdom

Artificial Intelligence

Artificial Intelligence and Simulation Tools for Game Design
Stephane Assadourian, UBISOFT, Montreal, Canada
Flavio Soares Correa da Silva, USP, Sao Paulo, Brazil
Liliana Vale Costa, University of Aveiro, Portugal

INTERNATIONAL PROGRAMME COMMITTEE

Antonio J. Fernandez, Universidad de Malaga, Malaga, Spain
Marc-Philippe Huget, University of Savoie, Le-Bourget-du-Lac, France
Pieter Jorissen, Karel de Grote Hogeschool, Hoboken, Belgium
Joseph Kehoe, Institute of Technology Carlow, Carlow, Ireland
David Moffat, Glasgow Caledonian University, Glasgow, United Kingdom
Joseph M. Saur, Regent University, Virginia Beach, USA
Miguel Tsai, Ling Tung University, Taichung, Taiwan

Deep Learning & Adaptation

Christian Bauckage, University of Bonn, Sankt Augustin, Germany
Christos Bouras, University of Patras, Patras, Greece
Pieter Jorissen, Karel de Grote Hogeschool, Hoboken, Belgium

Intelligent/Knowledgeable Agents

Wenji Mao, Chinese Academy of Sciences, Beijing, China P.R.

Opponent Modelling

Ingo Steinhauser, Binary Illusions, Berlin, Germany

Peripheral

Psychology, Affective Computing and Emotional Gaming

Liliana Vale Costa, University of Aveiro, Portugal
Sanaul Haque, Institute of Technology Carlow, Carlow, Ireland
Eva Hudlicka, Psychometrix Associates, Blacksburg, USA

Artistic input to game and character design

Pieter Jorissen, Karel de Grote Hogeschool, Hoboken, Belgium
Liliana Vale Costa, University of Aveiro, Portugal
Anton Eliens, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
Richard Wages, Nomads Lab, Koln, Germany

Storytelling and Natural Language Processing

Jenny Brusk, Gotland University College, Gotland, Sweden
Liliana Vale Costa, University of Aveiro, Portugal
Ruck Thawonmas, Ritsumeikan University, Kusatsu, Shiga, Japan
Marcus Toftedahl, University of Skövde, Skövde, Sweden
Joseph M. Saur, Regent University, Virginia Beach, USA
Torbjörn Svensson, University of Skövde, Skövde, Sweden
Clark Verbrugge, McGill University, Montreal, Canada

Online Gaming and Security Issues in Online Gaming

Liliana Vale Costa, University of Aveiro, Portugal
Marco Furini, University of Modena and Reggio Emiliano, Modena, Italy
Sanaul Haque, Institute of Technology Carlow, Carlow, Ireland
Michel Simatic, Telecom Sud-Paris, Evry Cedex, France

MMOG's

Liliana Vale Costa, University of Aveiro, Portugal
Sanaul Haque, Institute of Technology Carlow, Carlow, Ireland
Michael J. Katchabaw, The University of Western Ontario, London, Canada

INTERNATIONAL PROGRAMME COMMITTEE

Serious Gaming

Serious Games and Gamification
Oisin Cawley, Institute of Technology Carlow, Carlow, Ireland
Liliana Vale Costa, University of Aveiro, Portugal
Sanaul Haque, Institute of Technology Carlow, Carlow, Ireland
Joseph Kehoe, Institute of Technology Carlow, Carlow, Ireland
Peter Lowe, Institute of Technology Carlow, Carlow, Ireland
Daire O'Broin, Institute of Technology Carlow, Carlow, Ireland
Martin Prause, WHU Otto Beisheim School of Mgmt, Vallendar, Germany

Games Analysis
Joseph M. Saur, Regent University, Virginia Beach, USA

Wargaming Aerospace Simulations, Board Games etc....
Roberto Beauclair, Institute for Pure and Applied Maths., Rio de Janeiro, Brazil
Joseph M. Saur, Georgia Tech Research Institute, Atlanta, USA

Games for Training
Sanaul Haque, Institute of Technology Carlow, Carlow, Ireland
Pieter Jorissen, Karel de Grote Hogeschool, Hoboken, Belgium
Liliana Vale Costa, University of Aveiro, Portugal
Michael J. Katchabaw, The University of Western Ontario, London, Canada

Games Applications in Education, Government, Health, Corporate, First Responders and Science
Sanaul Haque, Institute of Technology Carlow, Carlow, Ireland
Pieter Jorissen, Karel de Grote Hogeschool, Hoboken, Belgium
Laura Lima, Universidade Federal de Juiz de Fora, Minas Gerais, Brazil

Games Interfaces - Playing outside the Box

Games Console Design, Mobile Gaming, Perceptual User Interfaces for Games
Liliana Vale Costa, University of Aveiro, Portugal

GAME ON®
2019

All articles published in these Proceedings have been peer reviewed

EUROSIS-ETI Publications are ISI-Thomson and IET referenced

GAMEON proceedings are indexed on SCOPUS and Elsevier Engineering Village

A CIP Catalogue record for this book is available from the Royal Library of Belgium under nr.12620

For permission to publish a complete paper write EUROSIS, c/o Philippe Geril, ETI Executive Director, Greenbridge Science Park, Ghent University, Ostend Campus, Wetenschapspark 1, Plassendale 1, B-8400 Ostend, Belgium.

EUROSIS is a Division of ETI Bvba, The European Technology Institute, Torhoutsesteenweg 162, Box 4.02, B-8400 Ostend, Belgium

Printed in Belgium by Reproduct NV, Ghent, Belgium
Cover Design by C-Design Bvba, Wielsbeke, Belgium
Divinity: Fallen Heroes Cover Art by Larian Studios

GAMEON® is a registered trademark of the **European Technology Institute** under nr: 1061384-761314

EUROSIS-ETI Publication

ISBN: 978-9492859-08-2
EAN: 978-9492859-08-2

Preface

Dear Participants,

A warm welcome to everyone attending the 20th GAME-ON conference at Breda University of Applied Sciences in the south of the Netherlands. We like to think that Breda is a mini Amsterdam but with the friendly warmness found “south of the river”. The city is full of interesting stories, from a Dutch “trojan horse” boat that surprised the occupying Spanish in 1590, to a German tank monument of the very same one stolen by the liberating Polish army in 1945. We hope you enjoy your time here and have a chance to see some of this history and architecture.

For game developers and researchers these are exciting times. Technology available is amazingly capable, yet accessible and enabling ever more innovative and realistic experiences. Technology of virtual and augmented reality continues to evolve and bring new dimensions to what a game can be. Next generation hardware moving to stream ever more realistic environments and characters - maybe even streaming your presence into the game world itself. And how do we make such large and innovative experiences, well it may leverage the power of procedural practises and intermix this with the maturing of deep learning approaches. This is fast evolving.

GAME-ON’2019 is being hosted by Games @ Breda University and we hope we can inspire you in some of these fast evolving areas. Established in 2006 the Games programme set out to create graduates that meet the quality needed to be immediately effective in the games industry. To do this we always aim to have applied research that interests and attracts industry game developers, this illustrated by our strong relationships with Sony, Ubisoft and SideFX to name a few. And that is why hosting GAME-ON is important in sharing the exciting developments in the area of game development and research.

We are looking forward to a number of exciting keynotes. Phoenix Perry, from Goldsmiths, will be sharing how games can smartly make better use of various hardware sensors to improve the gaming experience. Olivier Dauba, the VP at Ubisoft Editorial, will be sharing insights on how games could better support cognitive development. Liliana Vale Costa, from Universidade de Aveiro, will present how games and the psychotherapy process can intermix and the challenges that arise. Iris van der Meule, from AKV St.Joost, will explore how the new medium of VR brings new creative challenges and opportunities for environmental story-telling.

The time and effort of many people have helped make this conference happen. We would particularly like to thank Josephine Lappia, Professor Mata Haggis-Burridge and Mariska Kusters for their time and effort in helping our academy host the conference. And last but not least, none of this would not happen without Philippe Geril coordinating and bringing everything together.

We welcome everyone to GAME-ON'2019 and hope you enjoy the time in Breda and leave inspired.

Breda, September 2019

Robbie Grigg
Breda University, Breda, The Netherlands
GAME-ON'2019 General Conference Chair

CONTENTS

CONTENTS

REAL WORLD GAMIFICATION

SCIENTIFIC PROGRAMME

GAME
DESIGN
FUNDAMENTALS

GAMEPLAY DEFINITION: A GAME DESIGN PERSPECTIVE

Emmanuel Guardiola
Cologne Game Lab
TH Köln
51063, Cologne,
Germany
E-mail: eg@colognegamelab.de

KEYWORDS
Gameplay, Game Design, Handbooks

ABSTRACT

Game design research and game studies regularly define the notion of gameplay even if it is sometime considered an elusive word. These definitions are built on past *game* and *play* research in humanities or on a few game designers' opinions. In the second case, the handbooks or material used are limited. In this paper, we analyse a sample of 24 game design handbooks to compare their formal definitions of Gameplay and the way they use the word. We identify several common notions or characteristics in their approaches that may help us to build a more accurate definition of gameplay.

INTRODUCTION

In the game design research field *Gameplay* seems to be a magic word that avoids clear meaning. Its elusive aspect is mentioned in game design books as in Rollings and Adams (2003) or Crawford (1982). For Hiwiller (2016) it should even be totally ban from game design discussion during the making of a game: instead of saying that the gameplay is not working it's more efficient to identify the precise source of an issue, as for instance the camera setting, or a parameter of a game object. If this position sounds legitimate, the word gameplay is widely used in discussion about game. As an illustration, in the 24 game design handbooks that are the core material of this study, the word gameplay is cited more than 4.500 times, with an average one citation per 2.5 page. Even if this word seems elusive, it is also necessary to communicate about game. It has a role in the game lexical, for design and production, but also for critics, players and game studies.
When game studies delimit the meaning of gameplay it often starts with definitions from game design handbooks, considered as representative material from the practice. For instance Juul (2005) starts with Rouse (2004), Ermi (2005) used Crawford (1982). If sometimes the authors use multiple game design books as sources as Guardiola (2016), it is still far from an exhaustive state of art. A good starting point to improve the understanding of gameplay can be to explore a larger sample of game design handbooks instead of using a few. In our research we propose to go deeper into the resource coming from the practice and see how their authors are dealing with gameplay. Are there common notions or characteristics in their approaches that could help us to build a more accurate definition?

METHODOLOGY

The core method is to explore a large sample of game design books to compare their use of the word "gameplay" or, if available, compare their formal definitions of it. This section presents how we choose the books and collect the material.

A Selection Of 24 Handbooks

The selection have to be composed of game design books in English and available in a digital format to allow research in the text. We try to be as exhaustive as possible.
The eligible titles must offer tools, models or methodologies for game designers, in other words handbooks providing examples from the industry. The content must concern game design in general, without specialization in a type of game (ie board games, serious game, casual games etc.). They should avoid partial approach of the task. For instance it excludes books exclusively on narrative or level design. One of the author must have released at least one game on the market. This fact was checked through the author bio and, when possible, with web site dedicated to game credits as Moby Games (www.mobygames.com)
The selection includes game design books widely cited by academics works and popular titles. By popular we mean that users are recommending them on online reviews or rankings
We get digital copies of most of the identified titles (.pdf, .epub, .chm). When the digital versions were not available, we contacted directly the authors to ask for these files. On the 26 book identified, we end up with 24 useful files.

Analyse The Books

The digital copies of the books allow to track the presence of the words "gameplay", "game play" or "game-play". The hits includes the variations "gameplayer" and "gameplaying".
The main goal is to extract a definition from each of the 24 books. There are three situations. Some of the books offer formal definitions, for instance in specific section on gameplay or in the glossary. Some others don't. For these ones we have to go through the numerous citations looking for an explicit reference that could be used as a definition. For instance, the reference could be integrated in the description of the template of a game design document, in which the author explains what is expected in the gameplay section. Some books does not provide any definition or side definition. Then the method consists in looking for typical use of the word *gameplay* as in sentences like "the gameplay of this game is".
During the analysis, the method is to identify common notions, terms or principles shared by several definition.

Then we want to weight each these notion by counting how many time they appear among the 24 definitions.
The definitions are split in two tables in annexes of this papers. Annexe 1 for the books with formal definitions of gameplay, and Annexe 2 for the books without it.

ANALYSIS

General Comments
The final selection of 24 books is covering a period of time from 1982 to 2016. There are 26 different authors involved, with multiple apparitions of Adams, Rollings or Crawford. On the 24 books 11 provide formal definitions, 6 have side definitions, and the remaining 7 require to use other type of citations, most of the time an extract where the author describes an example of gameplay from a game.
Annexe 3 shows the books with their number of pages, the number of "gameplay" citations, and if it is present in the index or in the glossary. Note that the citations include the ones in summaries, index, or chapters titles repeated on several pages. Nevertheless the first observation is the surprising low number of formal definitions (7) or presence of "Gameplay" in the glossary (2) regarding the important amount of citations in the books (4566). "Gameplay" is widely used, sometime presented as a critical part of the design and production work, even if it is not often explicitly defined. Rouse, Adams and Rollings are the most likely to cite it, and they also provide definitions. Hiwiller has the lowest ratio of "gameplay" citation per pages (1 per 34 pages) due to his position on the elusive aspect of the word.

From The Definitions
Exploring the 24 definition (or use of the word gameplay) we identify several notions that are shared. Sometimes the notion is represented by a word (ie. "player" "Challenge"), sometimes it's a principle that takes many forms (ie "Permitted by or emerging from").
The "Player" is directly mentioned in most of the definitions. With 17 occurrences, the player is the most shared aspect of gameplay. The notion of player is also present indirectly. For instance "you" in Todd (2007) definition could refer to the player.
By "Action, verbs" we consider several aspects of it. It could be a cognitive or sensorimotor task (thinking, choosing, looking), it could describe the use of an input (press fire), or can be cited as an in-game manifestation (jumping). Action or verbs are directly present in 15 definitions.
"Interaction with" includes expressions like "interaction with an object", but also the presence word like "interactivity". It is cited 8 times directly, and could potentially be interpreted from Koster (2013) citation "exercising power over content".
"Challenge, performance" are named 7 times directly. It could also been perceived in Crawford (1984) "Cognitive effort" or Koster (2013) "Mastering responses to situations"
"Permitted by or emerging from" is a more elaborated notion. This principle is evoked in different manners in the definitions. It refers to the fact that gameplay is permitted by or emerges from mechanics/ rules/ object. It could sometime be directly stated as in Adams and Dormans (2012) "The actions that are related to challenges are governed by the game mechanics", or in Sylvester (2013) "Core Gameplay is what emerges from the irreducible mechanics of a game", or in Rogers (2014) "Video game mechanics are objects that create gameplay when the player interacts with them". With different wordings, other authors evoke the causality between certain game elements and the gameplay. In Bartle (2003) "the means by which the environment introduces goals for the players is called gameplay". For Rouse (2004) "gameplay is the degree and nature of the interactivity that the game includes". For Anthropy and Clark (2014) gameplay could be seen as the result of the combinations of verbs and objects, gameplay emerging from rules. 7 extracts from the definitions seems to share this causality.
The 6 "Environment" or "game world" or "simulated environment" mean the space where the gameplay takes place. It could also be the target of the interaction.
The 5 "Emotion" cover any sort of reference to emotion. It includes "pleasure" "felt" "enjoyable".
Other words are shared a very few times by some definitions. For instance, "Goal" "Choices" and "Feedbacks" appeared 3 times each.

Some Insight From The Others Citations Of Gameplay
A large set of gameplay citations was explored in the 24 books. Aside the main work on definition we also investigate the connection of gameplay to other topics as for instance level design or narration for further researches. Doing this we cross some other interesting notions related to gameplay.
Some aspects of gameplay are presented by several authors as obvious statements. Gameplay has a qualitative level, could be good, bad, and everything in the middle. The nature of the criteria could diverge. Also gameplay is a critical aspect of the game design task, but not always the most important one. Rollings and Morris (2003) rank interactivity first.
We also regularly found expressions like "during gameplay" "duration of gameplay" that frame it into a moment in time. Mention to space as "gameplay areas" or "arena" is also frequent, in particular when it comes to level design. Time and space perception of gameplay is formally connected to the game content. It seems natural that the game design defines phases, durations, triggers, environments where the gameplay takes place. This position reinforce the "Permitted by or emerging from" notion that we identify in several definitions.

DEFINITION PROPOSAL AND DISCUSSION

From the 7 most shared notions identified in the definitions we can try to set up a new one, reflecting the game design perspective on gameplay. A first try is to simply cite the list of elements. Something like: *Gameplay consist in the player, actions, challenges, interaction, emergence, environment and emotion.*
To come with a more meaningful one, we try to articulate the notions. We order the elements to be in line with the way theses notions are expressed in the original set of definitions. This is our proposition:

The Gameplay consists of the actions performed by the player when involved in a challenge. It emerges from the *emotionally-charged interaction between the player and the game components.*

"Actions" should be understood as all type of player's intentional activities, including pure cognitive ones, making choices, use of the senses (etc). Also, the expression "game components" suggests the game world, the rules, the objects and other potential constitutive formal elements. If these terms do not fit well, we would appreciate suggestions for more inclusive solutions. Another approach could be to expend the definition to integrate all the meaning of notion.

In the field is game design research, one of the goal is to offer a better understanding of the models involved in the design processes. Did these components and this definition reflect the practice?

During the development of a game, it could be asked to define the gameplay of it. The answer is often an action in a challenge like "jumping over enemies to reach the end of the level". These "player's actions" are visible in formal representations of gameplay. For instance in some game design documents this visualisation takes the form of a flow chart that connects cells named by the player's actions (Guardiola 2016). The "interaction with" is an obvious central aspect of the production of gameplay. As an example among many others: the design, coding and art production of feedbacks. These formal components of the interaction help the player to understand the impacts of her/his actions in-game or, should I say, during gameplay. About the "emerging from game elements" characteristic: the designers set up the challenges by assembling elements of the game as objects, mechanisms, and balance their parameters. This process of facing player's skills to game elements is sometime rationalize into processes, as for instance the rational design method applied *Rayman Origins* (McEntee 2102).

If all the previous characteristics of the definition have some visible existence in game production, the emotion aspect is not systematically documented or managed. Creating the engagement, the tension of challenge or the pleasure of play is less formalized but still present. In most of the cases developers want to induce these emotions. Playtests and focus groups are eventually conducted to try to evaluate these aspects of the experience, using questionnaires.

To open the discussion, we can compare the definition and components to those given by several books with a high impact in game studies. Are there interesting mismatches?

"Game play is the formalized interaction that occurs when players follow the rules of a game and experience its system through play." (Salen and Zimmerman 2004)

"The way the game is actually played when the player tries to overcome its challenges it its gameplay. The gameplay is an interaction between the rules and the player's attempt at playing the game as well as possible." (Juul 2005)

"For the sake of this discussion we define gameplay simply as the structures of player interaction with the game system and with the other players in the game. Thus, gameplay includes the possibilities, results, and the reasons for the players to interact within the game." (Bjork and Holopainen 2005)

The "permitted/emerging" aspect is never mentioned explicitly in these definitions. The challenge is only named in Juul's proposal. The emotional aspect of gameplay does not appear in any of these three contributions. The reason of these absences can be the small amount of resources on gameplay available at the time. Exploring a larger sample of game design handbooks, with many published after 2005, can add some characteristics to the previous game studies definitions of gameplay.

We can also confront this definition or this ensemble of notions to the current gameplay analysis methodologies. For instance: in their proposal of *formal analysis of gameplay* Lankoski and Björk (2015) introduce "Components", "Actions" and "Goals" as primitives. If "Components" and "Actions" are resonning with the findings from the game design books, the "goals" is not. We might be abble to investigate new approach for analysis, taking in accounts the characteristics identified.

CONCLUSION

"Gameplay" is considered as a word that covers a large range of meaning but is widely cited in the game design work. Investigating how it is used and defined in a large sample of handbooks, we find many common characteristics and a possible common definition.

24 game design books were selected to conduct our analysis. From their definitions of gameplay, or their typical use of the word, we found 7 characteristics that are shared from 5 to 17 times. From this material, a new definition is suggested: *The Gameplay consists of the actions performed by the player when involved in a challenge. It emerges from the emotionally-charged interaction between the player and the game components*. Compared to the game design task, this definition and the characteristics fit with the practice.

Even if our contribution provides a framework to think about the gameplay, it does not reduce neither handicap the creative aspect of it. The player's actions, the nature of the interaction, the game elements, and the emotions are as many elements that could be interpreted subjectively or in an infinite number of perspectives.

In comparison to some past definitions from the game studies this contribution increases the range of the characteristics of gameplay. We hope also that it can help for the design of future methods for the visualization, formalisation, and analyse of gameplay.

REFERENCES

Anthropy, A., & Clark, N. 2014. *A game design vocabulary.* Addisson Wesley, Upper Saddle River, NJ, USA.

Adams, E. 2009. *Fundamentals of game design - 2nd edition.* New Riders, Berkeley, CA, USA.

Adams, E., and Dormans, J. 2012. *Game Mechanics Advanced Game Design*. New Riders, Berkeley, CA, USA.

Bartle, A. R. 2003. *Designing Virtual Worlds.* New Riders, Berkeley, CA, USA.

Batemann, C., and Boon, R. 2005. *21st Century Game Design.* Course Technology / Cengage Learning, Boston, MA, USA.

Bates, B. 2004. *Game Design Second edition,* Course Technology / Cengage Learning, Boston, MA, USA.

Bjork, S., and Holopainen, J. 2005, *Patterns In Game Design*. Charles River Media, Hingham, MA, USA.

Crawford, C. 1984. *The art of computer game design.* Osborne / Mc Grawhill, Berkeley, CA, USA. Digital copy: https://archive.org/details/artofcomputergam00chri

Crawford, C. 2003. *Chris Crawford on Game Design.* New Riders, Indianapolis, IN, USA.
Ermi, L., and Mäyrä F. 2005. "Fundamental Components of the Gameplay Experience: Analysing Immersion". DiGRA'05. In *Proceedings of the 2005 DiGRA International Conference*: Changing Views: Worlds in Play.
Fullerton, T. 2008. *Game Design Workshop: A playcentric approach to creating innovative games-2nd*. Morgan Kaufmann / CRC press, Burlington, MA, USA.
Guardiola, E. 2016. "The Gameplay Loop: a Player Activity Model for Game Design and Analysis". ACE'16. In *Proceedings of the 13th International Conference on Advance in Computer Entertainment Technology*, Osaka, Japan, November 2016. New York, NY, USA: ACM
Hiwiller, Z. 2015. *Players Making Decisions: Game Design Essentials and the Art of Understanding Your Players.* New Riders, USA.
Juul, J. 2005. *Half Real.* MIT press, Cambridge, MA, USA.
Koster, R. 2013. *Theory of Fun for Game Design, 10th anniversary*, 2nd edition. O'Reilly Sebastopol, CA, USA.
Lankoski, P., & Björk, S. 2015. "Formal analysis of gameplay". In *Game Research Methods: An overview,* P. Lankoski, & S. Björk. ETC Press, Pittsburgh, 23-35.
Macklin, C., and Sharp, J. 2016. *Games, Design and Play: A detailed approach to iterative game design.* Addison-Wesley, Boston.
McEntee., C. 2012. *Rational Design: The Core of Rayman Origin*. Gamasutra. https://www.gamasutra.com/view/feature/167214/rational_design_the_core_of_.php
Moore, M. 2011. *Basics of Game Design.* A K Peters / CRC Press, Boca Raton, FL, USA.
Oxland, K. 2004. *Gameplay and Design.* Addisson Wesley / Pearson Education, Harlow, UK.
Perry, D., and DeMaria, R. 2009. *David Perry on Game Design: A Brainstorming ToolBox.* Course Technology Cengage Learning / Charles River Media, Boston, MA, USA.
Pedersen, R. A. 2003. *Game Design Foundations.* Wordware Publishing Inc, Plano, TX, USA.
Rogers S. 2014 *Level Up! The Guide to Great Video Game Design 2nd edition.* John Wiley & Son, Chichester, U.K.
Rollings, A., and Morris, D. 2003. *Game Architecture and Design - A New Edition.* New Riders, Indianapolis, IN, USA.
Rollings, A., and Adams, E. 2003. *Andrew Rollings and Ernest Adams on Game Design.* New Riders, Indianapolis, IN, USA..
Rouse III, R. 2004 *Game Design Theory and Practice- 2nd Edition.* Wordware Publishing Inc, Plano, TX, USA.
Salen, K., and Zimmerman, E. 2004. *Rules of play: game design fundamentals.* MIT press London, UK.
Schell, J. 2014. *The Art of Game Design, Second Edition*. A K Peters / CRC Press, Boca Raton, FL, USA.
Swink, S. 2008. *Game Feel.* Morgan Kaufmann / CRC Press, Burlington, MA, USA.
Sylvester, T., 2013. *Designing Games: A Guide to Engineering Experiences*. O'Reilly, Sebastopol, CA, USA.
Todd, D. 2007. *Game Design From Blue Sky To Green Light.* A K Peters / CRC Press, Boca Raton, FL, USA.
Ubisoft Montpellier. 2011. *Rayman Origins*. Playstation 3. Ubisoft

ANNEXES

Annexe 1: Game design books with formal definition of gameplay

Books with definitions	**Gameplay definitions**
Andrew Rollings and Ernest Adams on Game Design (Rollings and Adams 2003)	*chapter 7 in the section Defining Gameplay*: One or more causally linked series of challenges in a simulated environment *(later in chap 7:)* You will recall from Chapter 2 that gameplay consists of the challenges the player faces, plus the actions she can take to overcome them. As we said previously, designing the gameplay is one of your most important design tasks.
Game Mechanics Advanced Game Design (Adams and Dormans 2012	*Chap3 p43:* We define gameplay as the challenges that a game poses to a player and the actions the player can perform in the game. Most actions enable the player to overcome challenges, although a few actions (such as changing the color of a racing car or chatting) may not be related to challenges. The actions that are related to challenges are governed by the game mechanics. An avatar can jump only when a jumping mechanic has been implemented in the game, for example.
Designing Virtual Worlds (Bartle 2003)	*Chap 1, Section Some Definitions, p2*: the human beings who interact with the simulated environment are known as players rather than users; **the means by which the environment introduces goals for the players is called gameplay**; the activity of interacting with the environment is referred to as playing
Gameplay and Design (Oxland, 2004)	*p7:* I believe gameplay is the components that make up a rewarding, absorbing, challenging experience that compels the player to return for more, time and time again. It sits at the heart of a game that cannot be seen as a dimensional entity, but only felt from a superbly woven and captivating world of interactive challenges that stimulates your every sense.
Game Design Theory and Practice, 2nd Edition (Rouse 2004)	*pXX, section What Is Gameplay?:* A game's gameplay is the degree and nature of the interactivity that the game includes, i.e., how players are able to interact with the game-world and how that game-world reacts to the choices players make.
The art of computer game design (Crawford 1984)	*p20 gameplay section*: Game play is a crucial element in any skill-and-action game. This term has been used for some years, but no clear consensus has arisen as to its meaning. Everyone agrees that good game play is essential to the success of a game, and that game play has something to do with the quality of the player's interaction with the game. Beyond that, nuances of meaning are as numerous as users of the phrase. The term is losing descriptive value because of its ambiguity. I therefore present here a more precise, more limited, and (I hope) more useful meaning for the term "game play". **I suggest that this elusive trait is derived from the combination of pace and cognitive effort required by the game.**
Basics of Game Design (Moore 2011)	*P4:* **The actions a player performs during a game constitute the game play**. Each game genre has its own set of actions, although many games share common action, such as moving objects around on the screen. Simples games have few actions for the player to perform while complex games can have many actions. In the classic arcade game pong, for example, the players only have to move a paddle up and down the screen to intercept a moving ball and send it flying back at, and hopefully by, the other player (see figure 1.1). In a first-person shooter, the primary focuses are on moving a character through the game world and shooting AI-controlled enemies - and sometimes other players in deathmatches). There might be several different kinds of movement - running, walking, jumping, learning, crouching, and so on. There are also a number of different weapons the player can collect and wield during play.

Game Architecture and Design - A New Edition (Rollings and Morris 2003)	*Dedicated section P59 (...*) Now, suppose the priest has two kinds of spells, each of which cost him the same number of magic points. One spell injures the enemy (we'll call those "E-Bolts"), and the other heals injuries to your own group (we'll call those "Band-Aids"). Which should he cast during a fight? (...) There's no easy answer. It depends on lots of things**. That makes it an interesting choice. And that's what gameplay is all about**. (…) Sid Meier said, "A game is a series of interesting choices." To be worthwhile, gameplay choices must be non-trivial.
fundamentals of game design - 2nd edition (Adams 2009)	*In part one chap 9 dedicated + in glossary p640:* gameplay The challenges presented to a player and the actions the player is permitted to take, both to overcome those challenges and to perform other enjoyable activities in the game world.
21st Century Game Design (Batemann and Boon) 2005	*In section Gameplay versus Toyplay, p54:* We would therefore choose to define a toy as a 'tool for entertainment', and a game as 'a toy with some degree of performance'. Every game that can be conceived will include some degree of performance, either in the form of victory conditions to be achieved, failure conditions to be avoided, or metrics to measure progress. This in turn leads to two useful definitions: **gameplay, defined as 'performance-oriented stimulation'** and *toyplay*, defined as 'unorganised stimulation'.
Players Making Decisions (Hiwiller 2015)	*p78: in* A NOTE ON "GAMEPLAY": I try to avoid using the word gameplay. **What is usually meant by the term is the experience of playing a game**. however, it is a milquetoast cop-out of a word that keeps the writer or designer from really explaining what he is talking about. When you say a game has "good gameplay," what does that even mean? that it controls fluidly? that it has interesting dynamics? that the rules make sense? that it is fun for its target players? that it meshes with its theme well? these are all more precise and useful descriptions.

Annexe 2: Side definitions and sample of the use of "Gameplay" in Game design books without formal definitions

Books without definitions	**Side definitions or meaningful samples using "Gameplay"**
Designing Games (Sylvester 2013)	*p332:* CORE GAMEPLAY is what emerges from the irreducible mechanics of a game at the bottom of its dependency stack. Remove everything that can be removed without making a game emotionally worthless, and what's left is core gameplay.
Game Design Workshop - 2nd Edition (Fullerton 2008)	*p209*: The core gameplay mechanism, or "core mechanic," can be defined as the actions that a player repeats most often while striving to achieve the game's overall goal.
Game Design - Second edition (Bates 2004)	*Section on concept document p205*: **Gameplay - Describe what the player will do while he's playing the game**. Emphasize any new twists to the genre that your game provides. *And p274 in Game proposal document template:* 3. Gameplay - A paragraph that describes what kinds of actions the player can perform during the game.
David Perry on Game Design (Perry and DeMaria, 2009)	*What is expected as gameplay in a pitch: P515:* It always surprises me that someone can work for hours, weeks, and even months on a game concept and not be able **to describe the gameplay to me. I ask them, "Can you describe in detail what the player will be doing when playing your game?"**
Games, Design and Play (Macklin and Sharp 2016)	*About the use of play/gameplay, epub p16*: One of the first things you will notice about this book is the emphasis on play and play experiences. In fact, throughout the book we use gameplay and play experience interchangeably. We do this to challenge our mind-set about games. Instead of focusing on the idea that we are designing games, we prefer to think about designing opportunities for play. **By play, we mean the thinking and actions that emerge when we engage with games.**
Chris Crawford on Game Design (Crawford 2003)	*Side definition in Chapter 6 on interactivity*, *section "History"*: Interactivity (sometimes called "gameplay") is the real schwerpunkt of games. *(schwerpunkt: center of gravity)* *Typical use of gameplay: Chapter 19 section Implementation Woes:* The gameplay was simple: The player would use a cursor to designate a person to be called. Pressing the button would select that person, whose telephone would ring with an appropriate jangling sound and the handset jiggling on the telephone base. The person called would pick up the handset with a simple three-step animation, hold it to his or her ear, and say something like "Air-oh?",(…)
A Game Design Vocabulary (Anthropy and Clark 2014)	*Sample of "gameplay" citation: P38, description of a group activity than:* Using one of the verbs that you just discussed, come up with an idea for a game that develops this verb. This could involve special objects that help develop the verb, such as an object that the player can jump on to change the direction of gravity or the entire view of the game world or multiple verbs in conjunction with each other, such as a gun that changes objects into jumping platforms**. Talk about what kind of gameplay might result from these combinations of verbs and objects.**
Game Design Foundations (Pedersen 2003)	*Describing gameplay of Delta Force Urban Warfare p76*: Diverse, intense gameplay includes wild shoot-outs combined with stealth tactics, close quarters combat (CQC) with strategic infiltration, time-sensitive ops, sniping, and demolition.
Game Design From Blue Sky To Green Light (Todd 2007)	*Describing game elements of halo3 p113*: And because environment artists are coming up with the themes for these crazy spaces, we're also figuring out if there are special gameplay elements, like force fields you can't shoot through, or jump pads that can get you places very quickly, and how those relate to the overall theme of the environment.
Game Feel (Swink 2008)	*About mario64 gameplay prototype, p269:* Anecdotally, the prototype form of Mario 64 was a "gameplay garden," a test level which included a near-final version of Mario, complete with animations and moves, and a wealth of different things for him to interact with. *p322:* In a video game, some obfuscation is necessary and desirable; if intent and action merge, there's no challenge and no learning, and much of the fundamental pleasure of gameplay is lost.
Theory of Fun for Game Design, 10th anniversary, 2nd edition (Koster 2013)	*Samples of "gameplay" citation: p70:* Early platform videogames followed a few basic gameplay paradigms: • "Get to the other side" games: Frogger, Donkey Kong, Kangaroo. These are not really very dissimilar. Some of these featured a time limit, some didn't. • "Visit every location" games: Probably the best-known early platformer like this was Miner 2049er.* Pac-Man and Q Bert also made use of this mechanic. p164: The core of gameplay may be about the emotion I am terming "fun," the emotion that is about learning puzzles and mastering responses to situations, but this doesn't mean that the other sorts of things we lump under fun do not contribute to the overall experience.
The Art of Game Design, Second Edition (Schell 2014)	*Samples of "gameplay" citation: space invader gameplay p53*: The gameplay mechanic of Space Invaders was new, which is always exciting. But more than that, it was interesting and well balanced. Not only does a player shoot at advancing aliens that shoot back at him, the player can hide behind shields that the aliens can destroy (or that the player can choose to destroy himself). Further, there is the possibility to earn bonus points by shooting a mysterious flying saucer.

	p167: Gameplaying is decision making. Decisions are made based on information. Deciding the different attributes, their states, and what changes them is core to the mechanics of your game.
Level Up! (Roger 2014)	*Samples of "gameplay" citation: p16:* Game genre describes the type of gameplay (…) The game genre describes the play, not the art or story (…) ■ Action - Action games rely on eye/hand coordination and skill to play. (…) ■ Augmented Reality—Augmented Reality (or AR games) incorporate peripheral devices like cameras and global positioning (GPS) into gameplay *p353*: Video game mechanics are objects that create gameplay when the player interacts with them. They can be jumped on, activated with a button press, or pushed around.

Annexe 3: *gameplay* citations per book

Books	Pages	gameplay (-er; -ing) citation	In Index	In Glossary
(Rollings and Adams 2003)	648	398	Yes	N/A
(Adams and Dormans 2012	360	233	Yes	N/A
(Bartle 2003)	768	93	Yes	N/A
(Oxland, 2004)	368	226	Yes	No
(Rouse 2004)	704	776	Yes	Yes
(Crawford 1984)	120	Game play 10, Game-play 7	N/A	N/A
(Moore 2011)	400	Game play 129	N/A	N/A
(Rollings and Morris 2003)	960	474	Yes	No
(Adams 2009)	700	517	Yes	Yes
(Batemann and Boon 2005)	332	213		No
(Hiwiller 2015)	480	14	Yes	N/A
(Sylvester 2013)	416	30	No	N/A
(Fullerton 2008)	496	366	Yes	N/A
(Bates 2004)	450	97	"gameplay elements"	No
(Perry and DeMaria, 2009)	1040	153	No	N/A
(Macklin and Sharp 2016)	288	53	No	No
(Crawford 2003)	504	53	N/A	No
(Anthropy and Clark 2014)	240	22	No	N/A
(Pedersen 2003)	384	55	No	N/A
(Todd 2007)	304	117	N/A	N/A
(Swink 2008)	376	30	No	N/A
(Koster 2013)	304	16	N/A	N/A
(Schell 2014)	600	152	N/A	N/A
(Roger 2014)	550	332	Yes	N/A

AFFECTIVE GAMES: ADAPTATION AND DESIGN

Salma Hamdy[1] and David King[2]
[1] Division of Cybersecutiy
[2] Division of Computing and Mathematics
SDI, Abertay University
DD1 1HG, Dundee,
Scotland
E-mail: {s.elsayed, d.king}@abertay.ac.uk

KEYWORDS
Biofeedback, affective adaptation, game design patterns.

ABSTRACT

Biofeedback applications, including games, implement affect recognition and adaptation modules. The adaptation mechanism in affective games associate the extracted emotion or affect variables with in-game features and events; mainly with game mechanics. This is used to procedurally create new content or modify existing content or game characters. Explicit or implicit feedback and control offer different interaction experiences through visible indicators of affect or back-scene subtle changes in gameplay. Several experiments in the literature attempted to formalise the adaption mechanism by deriving mathematical relations between game feature and user satisfactions or exploit machine learning to predict adjustment parameters and playstyles. Others propose affective game design patterns.
A game designer's main role is to create engaging mechanics that provide players with an immersive experience. Affective game design requires a thorough understanding of the role of emotions in games and the expertise of an interdisciplinary design team from computer science, psychology and physiology. It seems that research is approaching affective game design with over-simplified definitions, and mainly from an implementation perspective, rather than a psychophysiological one. Generally, a deep understanding of the affective quality of certain game features is lacking and there is rising need to study such association and investigate ways to guide game designers on the selection and operation features.

1. INTRODUCTION

Human Computer Interaction (HCI) has evolved from classical interfaces with mouse and keypads, through touchpads and haptics, into more advanced hands-free input devices and wearable technology. Computer games benefit from such advancements as users usually enjoy the "exploration" phase of a new controller, and the promised enhanced usability and maximum experience. Relevantly recently, another form of interaction came to focus; biofeedback applications. Such applications react to users' psychophysiological states with the aim of learning how to modulate their physiological activities to improve performance. Overall, HCI applications require a level of adaption (system features alteration) to maintain interaction, and hence the biofeedback versions ideally adapt according to the predicted user emotions in an affective loop. There are two types of affective feedback; direct body-activation measures (e.g. heart rate) obtained from biometric sensors, and indirect where emotions are inferred from indirect features (buttons pressure or body posture). The former suffers usability and measurements issues, while the later may be difficult and time-consuming to tune.

Affective feedback systems were often coupled with adaptive task automation and health improvement applications. However, this recently shifted to a broad range of other applications, including games. The inevitable fusion of biofeedback and games initially lead to games designed to encourage users towards some target to improve performance. This is often through decision-making, attention and arousal validation, supported by the game score (Jerčić and Sundstedt 2019). The literature shows such bio-serious games focussing on emotion regulation and game-based learning. Nevertheless, affective feedback design is now being used for both serious and entertainments games creating adaptive gameplay for increased immersion.

With respect to ludology frameworks and game design, mechanics, dynamics, and aesthetics are the sources of players' emotional reactions that produce gameplay and engagement. Mechanics construct the rules by which the player interacts with the game and how the game responses, and involve elements like gravity, jumping, dialog, timing, weapons, etc. In biofeedback games, mechanics activations can be tied to direct or indirect physiological inputs. (Nogueira et al. 2016) argue to additionally consider player perception of the mechanics and whether they should be aware of the resulting adaptation. Consequently, a direct/explicit feedback approach conveys to players their emotional/physiological state using visual indicators, while an indirect/implicit approach presents physiological information indirectly through subtle changes in gameplay (Kuikkaniemi et al. 2010). Regardless of categorisation, an affective game (AG), where certain elements dynamically change based on player's emotions, includes adaptation capabilities that varies (or should vary) according to different design parameters. In an affective loop, the output of the classification module is the player predicted emotion and a mapping is required to link this to the game context by finding a game state suitable for the recognised emotional state (Yannakakis et al. 2016). This process can alter game content through adaptive changes to different game elements and mechanics.

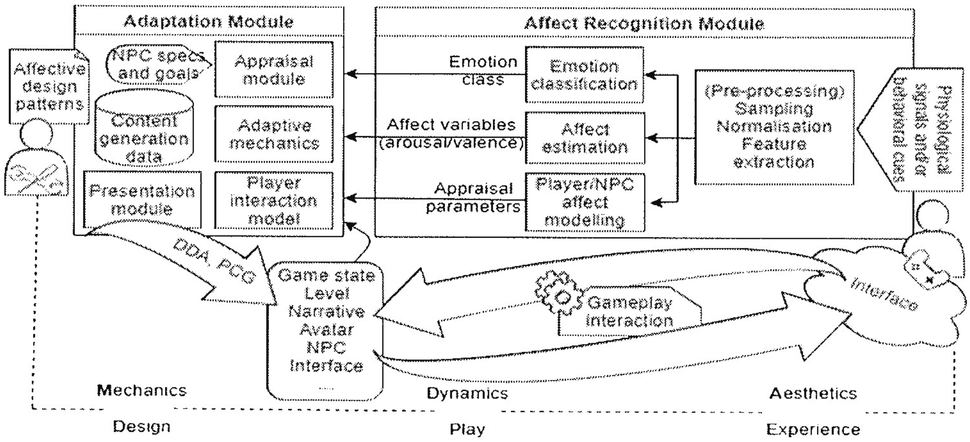

Figure 1. A framework for affective games.

This paper discusses the internal part of the AG loop as a design task and reviews the different adaptation mechanisms and design requirements suited for affective play experiences. In section 2, a number of dynamic adaptation elements are discussed along with examples from the literature that employ these in games. Section 3 analyses these sources of interaction in the overall design process and highlights relevant design issues. A generic framework for affective game design is presented. Conclusions are presented in section 4. In Figure 1, we propose a basic framework for affective games design that incorporates biofeedback and adaptive mechanics. The adaptation module is addressed in this paper. The affect recognition module was investigated in (Hamdy and King 2018).

2. DYNAMIC ADAPTATION

Several categories emerge of game components that adaptively change during gameplay (Bontchev 2016, De Byl 2015) but dynamic difficulty adjustment (DDA) seems to be the most exploited in the literature as it can be tied to several mechanics at once and most of the other components can be designed within this category. Using player performance, the difficulty of any of game task can be controlled through positive/negative feedback and hence the overall game difficulty is adjusted automatically.

2.1 Adaptation Targets

Three main directions for affective adaption in games can be identified:

Content generation is typically applied to platform games and uses methods like procedural content generation to automatically generate textual and multimedia game content algorithmically. A taxonomy of level generators in (Togelius and Yannakakis 2016) suggested a first-order player-centric approach where interaction with the level elicits affective responses (partly caused by design) that are used as input to a model to predict emotions. A second-order designer-centric approach highlights the affective goals of the designer and aspects of the internal player model captured to predict the emotions. Interaction patterns are analysed to infer the player's/designer's emotional state which is associated with level context variables. Furthermore, in a direct generator; the evaluation function of the level generation mechanism is built on a computational model of the player's affect as opposed to considering other aspects of the player experience beyond affect and emotion, such as behavioural traits and cognitive processes, in an indirect generator.

Content adjustment dynamically adapts elements or level of interaction of specific context/items according to player's skill acquisition. Content collectively falls under game narrative which consists of different descriptors that create the overall impression of "story" to the player. These include most visual elements of the game (text, cut-scenes, interface, world, audio-visual, plot, sounds, music, atmosphere, dialogues, player choices and gameplay etc.). Unlike other forms of entertainment, game narrative is not just the plot. It entails interactivity and player participation. (De Byl 2015) realise this in their conceptual model as interface and point out how game interfaces can be exploited to provide "seamless" interaction mechanisms with the game environment that encourage emotional behaviour. Adapting audio-visual content was first used in multimedia applications like music recommendation. One of the early attempts in games directly mapped physiological player responses to changes in game properties (Dekker and Champion 2007).

Character adaptation involves the design of adaptive intelligent agents to create emotionally believable behaviour, or to perform "proper" decision making based on their current emotional state. These include NPCs that react to players affect, use their emotional stat to decide their next behaviour, or forge a long-term companionship with players. Either way, the affective gameplay experience is improved through the player's emotional attachment to the NPC adaptability. This extend to the player's avatar whether central (emotionally embodied as themselves) or acentral (emotionally seen as a third person) that require a gradual bond to establish. Affective game characters and believability design was addressed comprehensively in (Hamdy and King, 2017).

Overall, most literature in adaptation focusses on game world, level or character. Despite research signifying that changes in musical expressions and sound effects can be reliably identified by electrodermal activity (EDA) levels (Garner 2016), a very limited number of studies addressed procedural music or game sounds that reflects, or changes to, the player's emotional state, and were mostly horror games targeting fear. Next, we present selected examples from the literature that adaptively exploit game mechanics within the above categories for both serious and entertainment affective scenarios.

2.2 Related Work

(Zafar et al. 2018) implemented three "casual" biofeedback games incorporating explicit and implicit features and providing visual feedback for players to control stress levels and regulate their breathing. Four components were dynamically altered: game environment, game-controlled elements, player-controlled elements, and object manipulation. Players showed better breathing control during gameplay and improved attentional-cognitive performance in subsequent tasks. The study concluded that game adaptation is more suitable for skill acquisition and transfer than visual biofeedback. However, evidence from other work suggest that combining both is more effective.

Another relaxation game in (Chittaro and Sioni 2014) used a multi-modal affective input to control the behaviour of a character inside a realistic virtual environment. Level progression was subject to avoiding stress sources and relaxing, allowing the character to complete a task. The study carried out three experiments for single EDA input, multi-modal, and a placebo condition. Interestingly, the multi-sensor technique did not show noticeable difference from the EDA-only one, which performed significantly better than the placebo condition. This can simplify an effective model if a single input is enough to gauge a change in the emotional state. Moreover, authors highlight the importance of a control condition in experiments design as comparing two non-placebo conditions may not reflect authentic results.

(Negini et al. 2014) developed a FPS zombie survival game where Galvanic skin response (GSR) controlled the difficulty by altering three mechanics: player (avatar speed and rate of grenade respawns), NPC (speed and number of zombies) and environment (density of fog and rate of health packs respawns). Authors suggest a formula for updating the game elements based on constants and the arousal variable inferred from the GSR. However, these constants, and the threshold resembling a normalised GSR, are adjusted manually based on design experience and play testing prior to the experiment. Although the experiments were limited to a small number of participants, result indicated that GSR was higher for the affective version of the game. They also reported that adapting the NPC resulted in less enjoyment and that player adaptations were the most noticed by participants.

(Nogueira et al. 2016) implemented biofeedback modulation through a horror game with four mechanics that could be influenced by players emotional state: character sanity (sane, scare, terrified, insane), creature AI (passive, passive-aggressive, and aggressive), character ability (sprint velocity, stamina and orientation) and evasion tunnels (spawned in strategic positions to reward quick-thinking). The game supports VR with Oculus Rift and is suitable for emotion regulation tasks as it records skin conductance (SC), heart rate (HR) and facial electromyography (EMG). Both the game's level layout and events can be generated at runtime. Results perceived the biofeedback to increase the gameplay depth. However, different participants had significantly different interpretations of some gameplay features, highlighting the need for design guidelines based on player type.

(Gilleade and Allanson 2003) incorporated three genres in their experiments with a hybrid action-sport-puzzle game. This paper concluded that a complex analysis of the physiological signals is not possible and opted to a class-based assessment; board, tired, content, excited and ecstatic. The tempo and environment elements changes according to player positive/negative response through HR. However, experiments were conducted on a very small sample with the electrocardiograph (ECG) employed to map a range of physiological state expected from the player. Also, the study assumed decreased and increased heart rates are always associated with negative and positive responses, respectively, which may not be a valid assumption.

(Nacke et al. 2011) used a standard controller in Xbox360 2D side-scrolling shooter game in addition to physiological input processed separately. The study addressed direct physiological signals (gaze, EMG, and respiration) and indirect ones (EDA, ECG, and temperature). The latter was implemented by blowing hot air on the sensor. These inputs were linked to variables that altered five mechanics in different ways: enemy target size, flamethrower length, speed/jump height, final boss speed and weather conditions, and Medusa's gaze to freeze enemies and platforms. The study tabulates the measured signals and what pre-processing applied to it from each sensor, which could be useful in deciding which mechanics suit which sensor, and how the variables are augmented into the game after pre-processing. Results showed the majority of participants preferring direct control as indirect ones are difficult to control separately and may not be suitable for fast-paced action games. Also, controls were perceived as best when they matched natural input (melting snow by blowing and freezing by gazing), a guideline observation for associating mechanics design with sensors and vice versa.

Authors is (Yannakakis and Hallam 2009) derived differentiable relations between game features (response time, pressure on tile, number of interactions) and entertainment value, quantified as a variable, in a physical Bug Smasher AR game. Machine learning models were used to predict entertainment preferences and gradient descent was applied to suggest direction/magnitude of parameter adjustments. Two mechanics were altered: challenge (speed with which bugs appear/disappear) changed adaptively by a simple set of rules, and curiosity (spatial diversity of the opponents appearing in the game space) adjusted through the partial derivative. Results reveal a preference for the adaptive Bug Smasher version over the static one. Although reported a few limitations and maladjustments, the study shows the possibility of quantifying fun, and that even simple methods can reasonably adjust game features to maximise player satisfaction.

(Balducci et al. 2017) introduced a set of formal design guidelines based on RPG features to describe two levels: boredom and flow. A support vector machine is used to classify five emotions detected by a wireless electroencephalography (EEG) headset recording brainwaves. Both levels implemented simple/dubbed/riddle dialog, single/group fight, chest open, skills upgrade, and stealing task mechanics. Event logs were crossed to match in-game data and external data from the headset and reported high accuracy for both game levels confirming the goodness of design and development method. Dialogs were classified best for both boredom and flow proving to be effective in manipulating player emotions positively and negatively in RPGs. Quantitative results show the boredom level being perceived as repetitive and tedious with almost similar accuracy across tasks, while rating varied for the flow level with gradual increase confirming progressive player engagement due to adaptation.

The study in (Nacke et al. 2014) investigated the accuracy of automatic in-game recognition of four playing styles based on facial expressions. A simple linear regression model was constructed from structured interviews with domain experts and applied to infer the styles. The game adapted shooting, discovering, puzzle solving and planning tasks to both player's skills and emotions. Machine learning was used to implicitly recognise playing style at run time and results were cross-matched to self-reported data to record an in-game recognition accuracy of around 71%. However, the experiments only used still facial expressions on a small number of participants and is based on certain learning theory and styles.

(Li et al. 2015) suggested a systematic model of designing affective VR games based on a dynamic graph structure and a closed-loop affective computing system. Skin conductance and heart rate variability are mapped to discrete emotional states. The graph-based approach includes different sub-graphs scenarios that induce different emotions (neutral, positive and negative). The system keeps track of the path a user takes to complete the game, including the order of nodes visited and each transition between sub-graphs. Authors proposed parameter-based formulas to measure the effectiveness of emotions (based on nodes in the subgraph), interaction (based on the transition between subgraphs) and game design (based on successful trials of completing the game). Although limited to the specific experiment, this seems a promising trial to quantify a qualitative experience and can be extended to other scenarios.

(Plass et al. 2019) investigated the effect of choosing design features for game characters. They altered four visual attributes: shape, colour, expression and size of the character. The study associated round shapes and warm colours with positive arousal, and used happy, sad and neutral expressions of a 3D game character to elicit positive/negative responses. Participants perceived expression and dimensionality with the strongest effect, while colour had a medium effect, and shape a small-medium effect. Of course, not all characters can be round, orange and smiley, but the above observations lend themselves to affective NPC design and help form some guidelines for creating appealing agents in certain situations.

(Lara-Cabrera and Camacho 2019) provides a taxonomy of affective games based on type of feedback, with a tabulated review of games that employs direct/indirect affective feedback. Reader may also refer to (Novak et al. 2012) for more examples on research-induced affective games.

It is undisputed that including a biofeedback mechanism in the system have positive impact on user engagement and overall experience. However, as appealing as it sounds, affective adaptations are difficult to implement, and it is often unclear whether the expected user state has been achieved can be reached more efficiently. The next section analyses affective games from a design perspective and points out issues that game designers are encouraged to consider early in the design process.

3. AFFECTIVE GAME DESIGN

3.1 Biofeedback System

An adaptation mechanism can be a system with no explicit data fusion where a variable is directly proportional to the physiological signal, a classifier followed by a simple decision-making with a pre-defined action for each recognised emotion class, or an emerging system that gradually adapts to users as they gain experience. Either way, the module is preceded by pre-processing the physiological inputs for noise removal and feature separation. The extracted components are then either used to identify the player's emotional state or streamed directly into the adaptation mechanism to adjust mechanics. (Novak et al. 2012) recommend using classification over estimation because discrete classes are easier to validate using questionnaires or independent observers than continuous values. However, for games, it seems that estimating values is sufficient when the system is designed for continuous physiological inputs, while a classifier is used when the adaptation operates on discrete values, typically the arousal and valence (Bontchev 2016). This is also believed to vary depending on the game requirements and mechanics to be adapted. As mentioned in section 1, quantifying player satisfaction and elicited emotions helps evaluate the overall experience on a parameter-based model. Finding a mathematical mapping between the extracted physiological variables and desired game features facilities the game design and implementation processes and offers more generic approaches to be followed.

Rather than explicitly identifying players' emotions, most experimental affective games tie the sensing device output(s) to a general emotion and employ variations of interaction-feedback mechanisms. Direct-explicit feedback is best for easy mechanics and simple gameplay and should be intuitively mapped to actions and mechanics of the game. Indirect feedback allows for more subtle adaptation and is more suited for slow-changing game world variables that do not directly influence the mechanics. It was also found that combining negative and positive feedback types contributes essentially for a higher player satisfaction (Bontchev 2016). Player's immersion is strongly increased by using direct control, but this can be exploited if players learn to intentionally manipulate physiological inputs to cheat (Nacke et al. 2011). Overall, it is best when control matches natural input (like flexing legs to jump) with more intuitive game interfaces, but it limits the flexibility and generality of the sensors.

Due to the diversity of emotions across people and events, it is important to find best suitable combinations of affective variables. to baseline the hardware. Most of the work in section 2 begin by normalising the collected physiological data though a neutral or a placebo scenario, where the game score and difficulty are independent of and not influenced by the biofeedback. Otherwise, generalised conclusions cannot be drawn. This is to calibrate the affect detection/recognition module and eliminate the variation of first exposure (Li et al. 2015). The quality of the raw data will impact on the adaptive gameplay and hence prep-processing is often applied. Nevertheless, significant increases in recognition accuracy may not result in significant changes to user experience (Guillotel et al. 2015). The work in (McCrea et al. 2017) concluded that classification accuracies below 70% are unacceptable to end users while increasing accuracy above 90% has only small benefits. They suggest an acceptable accuracy of about 70–80% for affective games. Furthermore, adaptation frequency was not explicitly tested in the literature. Ideally, it should depend on the game and mechanics, but (Yannakakis and Hallam 2009) reported that adjustments at 45 seconds have, on average, a lower impact on the entertainment value. Again, this perhaps is a player-dependent parameter, and it is noted that most of the studies admit to the insufficient number of participants and short playing time impacting results reliability.

3.2 Game Genre

Affective adaptation should not change the nature of the gaming interaction irrespective of the game type, otherwise, the game loses its appeal and falls out of its designated genre(s). (Gilleade and Allanson 2003) argue that only game genres with high level of interaction are suitable for affective inputs. Hence, they suggest action-sport-puzzle games as they require full player attention, high physical responses and fast-paced interaction. Contradictorily, (Zafar et al. 2018) claim that action, adventure, sport and fighting games are inappropriate for affective adaptation as they lack common appeal, higher stimulation and may contain violent content. Alternatively, they point out what they refer to as "casual games" to be the best choice; defined as fast-paced games that are easy to play for a short duration and still deliver a good experience. Strategy, RPG and simulation games are excluded as they require less interaction levels, longer to execute actions, and do no elicit a wider range of emotions, while quiz-based and board games do not possess enough dynamic content for adaptive gameplay.

Possible mechanics that can adaptively incorporate or convey affective information vary greatly between genres, as each genre features different elements. Action and adventure games are all about exploring worlds and solving puzzles and will almost certainly feature NPC interaction. An inclined adaptation would involve dynamic generation of the NPC emotions, behaviour and adaptive social interactions. It can also include manipulating objects and interactions with other NPCs based on extracted emotions. Fighting and FPS games involve direct combat with another player or NPC, hence can include similar mechanics, although with different expressions of emotions. RPGs are based on adventures with complex missions that involve a wider range of social interaction and deeper connection to the NPCs. In addition to dynamic generation of emotions and adaptive level content, dialog and quests can be adjusted to affective states of player or NPCs. It is worth noting that serious games are the fastest growing genre with respect to adapting affective inputs, where all the other types can be designed within for instructional, training or therapeutic purposes (Broekens et al. 2016).

3.3 Patterns

An interesting approach to affective game design that emerged recently is built on the assumption that player's emotional reactions to in-game events can be associated with certain patterns early in the design phase. Originally, the well-known game design patterns by Holopainen and Björk, (2004) exploits the repetitive nature of mechanics across games and genres to provide a communication tool during the design process. The template includes pattern name, problem, solution and consequences, and used to express gameplay to designers, gamers and domain experts.

(Gizvcka and Nalepa 2018) highlighted the lack of such consistent methodology for affective games and implies that some game design patterns evoke emotional responses by nature and that detecting these responses is possible. Hence, an arbitrary set of affective design patterns can be distinguished, and a framework is needed to somehow formalise the design process with a catalogue of game design patterns including affective ones. It is assumed that a correlation exists between game events designed with affective patterns and the resulting players' physiological responses patterns caused by these (those?) events.

One work in (Argasiński and Węgrzyn, 2019) suggested some design patterns for game-based learning and a list of mechanics that can be exploited. A novel framework is presented for the creation and evaluation of serious affective games. The model was tested with six design patterns in a hidden object puzzle adventure game designed for occupational safety and health. The study emphasised traditional game design frameworks "lacking insight on the role of affect in game systems" and concluded that some patterns do have a clear affective association that can be exploited to establish an affective loop within gameplay. Another in (Ng et al. 2018) studied 21 affective game design components collected form the literature in an experiment that involved the puzzle game LittleBigPlanet2. The participants group included designers as well, in an attempt to discover their thoughts and feelings on the game design as sometime, the designer's intended emotions may not be experienced by the player. The study concluded with 15 recommendations for affect user-centred game design grouped under user diversity, challenging and creative gameplay, and impressive visuals, all of which server the flow in the game.

On a relevant note, in analogy to traditional game design/development tools and classical physics engines, (Hudlicka 2009) presented the notion of affective game engine; a system capable of dynamically instantiating game elements accordingly to specific emotions and affective factors. Authors suggest the engine containing both a modelling module to track player's emotions and identify triggers leading to emotions, and provide real-time construction of an affective user model, and a manifestation module responsible for implementing emotional expression across modalities. The engine should be scalable to facilitate any necessary tuning and extendible to accommodate advances in affective sciences. This means the engine would minimally include features like type of emotion, intensity, decay, triggering conditions, association between modalities and channels, and behavioural choices, and must be able to encode this knowledge and data as they emerge during gameplay. Similarly (Broekens et al. 2016) discussed an emotional appraisal engine as a specialised game engine to support modelling of emotions in NPCs, in a manner that does not require a commitment to a particular NPC architecture. Their system GAMYGDALA (Popescu et al. 2013) is a plug-in module that is capable of generating emotions and dynamic relationships among NPCs, and integrate emotional appraisal over time, intensity and decay. The designer can define NPCs goals and how they are affected by particular game events which governs their emotions generation. Furthermore, relationships among NPCs can be configured, which forms the basis for social emotions towards other agents. This work demonstrated the emotional appraisal engine independence from the NPC AI and integration with several known game engines and development tools.

Frameworks and Data

(De Byl 2015) suggested a conceptual framework to position adaptation categories on a layered emotional system to serve as general affective game design guideline. Also, the work in (Balducci et al. 2017) can be useful to improve the guidelines in traditional game design featuring different gameplay and interaction modalities. A model-driven engineering (MDE) approach was followed by (Tang et al. 2013) to propose a framework for serious game software independent of hardware or platform specifications. Whether this applicable for entertainment games given the diversity of requirements and design features is still to be explored.

It is no surprise that with the emerge of affective games, new types of data will start to play a role in game design and analytics, where the player's actual bodily responses, not just game states, may need to be stored. Game telemetry databases are contentious with ethics and security concerns and having bio-data collected will certainly fuel the risk, especially if identifying features are included. These issues need to be addressed and resolved for affective games to commence on a production level. (Drachen et al. 2013) elaborates on game telemetry and lists useful gameplay metrics for different genres which can guide some decision in the design process.

4. CONCLUSION

According to how the adaptation module handles the physiological inputs, players are able to control the game either indirectly through content adaptation if their predicted emotion is coupled to a game element, like colour or sound, or directly if their affective variable is controlling events, like steering or shooting. Players often need to feel in control and having a direct adaptive system that keeps changing gameplay every time the player exhibits a spike in affect, may not always reflect a positive experience, especially for implicit feedback. Direct control is not suited for face-paced entertainments scenarios as it slows down the pace (Zafar et al. 2018) and is difficult to accustom to. It may be more suitable for emotion regulation and skill training (Nacke et al. 2011). Furthermore, the adaptation frequency and duration should be cautiously set to fit seamlessly within the game flow. Attributes could be adjusted on certain time windows, every new level or game, or set of critical actions (Yannakakis et al. 2016). Decision-making tasks (serious games) and quests/missions (entertainment) must be designed to ensure sufficient levels of arousal to avoid breaking the adaptation loop (Jerčić and Sundstedt 2019). It might be more effective to combine the affective module with a another "static" one, timed on other parameters like game state or player skill.

Overall, existing affective games operate on a predefined set of rules and lack the "ambitious" ability to create new reactions on their own (Lara-Cabrera and Camacho 2019). NPC adaptation rules are often based on OCC and PAD emotional models for their analytical and computational nature, while avatar and affective dialogs are not widely explored. Interactive narrative, specifically dialog, seems a promising venue for affective games. (O'Neill and Riedl 2016) explores the concept of dramatis; a computational model that calculates suspense levels over time allowing for increased emotional content. A narrative engine can provide an affective game with artificially generated stories and interactive narratives tailored to an individual player's emotions.

Hybrid game genres are rarely investigated, and action, adventure and puzzle games were designed with the claim that an affective game has to be high-paced and casual. However, we believe it is not possible to categorise games with respect to experienced emotions. Players need a broad spectrum of options and mechanics per genre and playstyles are heavily influenced by certain design affordances. Recommendations can be made within each genre, but still players react differently to the same genre and their responses vary with experience and time spent on the game.

Currently, game designer should focus on available consoles and controllers and device "affective scenarios" and methods exploiting those within a demographic game design approach in mind (Bontchev and Georgieva 2018). Moreover, commercialisation is reliant on sensors accessibility and usability in dynamics conditions, hence the suggestions of augmenting existing controllers with affective technology.

Decisions!

The game designer may be faced here with several decisions that are somewhat out of their traditional role. This is why affective applications design is a multi-disciplinary task that requires fellow members from psychology and sociology, something that seems lacking in the experiments from the literature.

Exisiting designs are often built around distinguishing boredom and engagement, which is perhaps sufficient for monitor flow and immersion, or coupling mechanics with basic classes of emotions or physiological levels based on the OCC appraisal model. It is understandable that this facilitates implementation, but for actual gaming environments, even small-scale commercial level, more complex creative gameplay is owed, and the design team should thoroughly investigate the range and depth of experienced emotions in the designated genre.

This, consequently, should give an insight on the choice and number (unimodal vs multimodal) of physiological inputs to harvest and how to include them in the game, what theory/model to adopt for modelling the predicted emotions, and what mechanics are best suited for each input and how these vary with respect to the changes in affective variables. It is difficult to say which impacts which, decision of mechanics or physiological inputs type; some experiments in the literature used existing games and hence adapted for existing gameplay, while others designed and created their own test games based on their choice of affective inputs. Also, with respect to hardware, either simple cheap devices, powerful multichannel commercial devices, or custom home-made ones were used approach in mind (Bontchev 2016). This, again, is a design decision that has to be carefully consider as it reflects on the capabilities of the system and impacts subsequent features.

Collectively, the game designer needs to consider affective adaptation that incorporates regular user interface for explicit feedback, provides a suitable control mechanism, preferably direct; making the players feel they still have the upper hand, while still applying implicit changes to game elements to achieve the desired adjustments, maximising the affective experience. Some more solid designs would include a player-centric model to tailor gameplay with specific playing styles and skills. All of this must be in a timely manner that achieves the affective experience, without breaking immersion or causing boredom.

Choices for the above design parameters will certainly vary with different design aspects like game genre, targeted audience, and mechanics implementation. Some intuition is required from the designer to associate available physiological signals to traditional game event or actions within an acceptable engaging narrative, and the possibilities are endless. However, the nature of biofeedback systems combined with the classical uncertainty of "good" game design, makes the task more of a constrained challenge. The more complex the game, the more creative the designer can be with affective inputs, going beyond the discussed categories, but this is likely be hindered by development and testing issues, in addition to usability and user satisfaction. After all, not all fabulously drafted ideas are produced well into computer games.

To conclude, the designer has to remember that the goal is improving players experience not necessarily their score, and it is better to think of the adaptation mechanics as an enhancement rather than fundamental. Furthermore, perfectionism limits creativity; even simple techniques, emotions, and mechanics can generate interesting gameplay with a clever enough design.

REFERENCES

Argasiński, J.K. and Węgrzyn, P. 2019. "Affective Patterns in Serious Games". Future Generation Computer Systems, vol. 92 (March), 526-538.

Balducci, F.; Grana, C.; and Cucchiara, R. 2017. "Affective Level Design for A Role-Playing Videogame Evaluated by a Brain–Computer Interface and Machine Learning Methods". The Visual Computer, vol. 33, No. 4 (April), 413-427.

Bontchev B. 2016. "Adaptation in Affective Video Games: A Literature Review". Cybernetics and Information Technologies vol. 16, No.3 (Sep), 3-34.

Bontchev, B. and Georgieva, O. 2018. "Playing Style Recognition Through an Adaptive Video Game". Computers in Human Behavior, vol. 82 (May), 136-147.

Broekens, J.; Hudlicka, E.; and Bidarra, R. 2016. "Emotional Appraisal Engines for Games". In Emotion in Games, Karpouzis K. and Yannakakis G.N. (Eds.). Springer, Switzerland, 215-232.

Chittaro, L. and Sioni, R. 2014. "Affective Computing vs. Affective Placebo: Study of a Biofeedback-Controlled Game for Relaxation Training". International Journal of Human-Computer Studies, vol. 72, No. 8-9 (Aug-Sep), 663-673.

De Byl, P. 2015. "A Conceptual Affective Design Framework for The Use of Emotions in Computer Game Design". Cyberpsychology: Journal of Psychosocial Research on Cyberspace, vol. 9, No. 3 (Oct), --.

Dekker, A. and Champion, E. 2007. "Please Biofeed the Zombies: Enhancing the Gameplay and Display of a Horror Game Using Biofeedback". In Situated Play: Proceedings of the DiGRA Conference (Tokyo, Sep.24-28). The University of Tokyo, 550–558.

Drachen, A.; El-Nasr, M.S.; and Canossa, A. 2013. "Game Analytics – The Basics". In Game Analytics: Maximizing the Value of Player Data, Drachen, A.; El-Nasr, M.S. and Canossa (Eds.), Springer, London, 13-40.

Garner, T.A. 2016. "From Sinewaves to Physiologically-adaptive Soundscapes: The Evolving Relationship between Sound and Emotion in Video Games". In Emotion in Games, Karpouzis K. and Yannakakis G.N. (Eds.). Springer, Switzerland, 197-214.

Gilleade, K. and Allanson, J. 2003. "A Toolkit for Exploring Affective Interface Adaptation in Videogames". In HCI International (Crete, June 22-27), 370-374.

Gizycka, B. and Nalepa, G.J. 2018. "Emotion in Models Meets Emotion in Design: Building True Affective Games". In IEEE Games, Entertainment, Media Conference (Galway, Ireland, Aug 15-17), IEEE, --, 1-5.

Guillotel, P.; Koenig, A.; Chanel, G.; and Novak, D. 2015. "Toward Commercial Applications of Affective Computing". IEEE Transactions on Affective Computing, vol. 8, No. 2 (April-June), 145 – 147.

Hamdy S. and D. King. 2017. "Affect and Believability in Game Characters – A Review of the use of Affective Computing in Games". In GameOn'17 – The Simulation and AI in Games Conference (Carlow, Ireland, Sep.6-8). Eurosis, 90-97.

Hamdy S. and D. King. 2018. "Affective Games: A Multimodal Classification System". In GameOn'18 – The Simulation and AI in Games Conference (Dundee, Scotland, Sep.18-20). Eurosis, 19-26.

Holopainen, J. and Björk, S. 2004. Patterns in Game Design. Charles River Media Inc., Hingham, Massachusetts.

Hudlicka, E. 2009. "Affective Game Engines: motivation and requirements". In Proceedings of the 4th International Conference on Foundations of Digital Games (Orlando, Florida, April 26-30). ACM, NY, USA, 299-306.

Jerčić, P. and Sundstedt, V. 2019. "Practicing Emotion–Regulation Through Biofeedback on the Decision–Making Performance in the Context of Serious Games: A Systematic Review". Entertainment Computing, vol. 29 (March), 75-86.

Kuikkaniemi, K.; Laitinen, T.; Turpeinen, M.; Saari, T.; Kosunen, I.; and Ravaja, N. 2010. "The Influence of Implicit and Explicit Biofeedback in First-Person Shooter Games". In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (Atlanta, Georgia, April 10 – 15). ACM, NY, USA, 859-868.

Lara-Cabrera, R. and Camacho, D. 2019. "A Taxonomy and State of the Art Revision on Affective Games". Future Generation Computer Systems, vol. 92 (March), 516-525.

Li, Y.; Elmaghraby A.S.; El-Baz, A.; and Sokhadze, E.M. 2015. "Using physiological signal analysis to design affective VR games". 2015 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT) (Abu Dhabi, UAE, Dec. 7-10), IEEE, --, 57-62.

McCrea, S.; Geršak, G.; and Novak, D. 2017. "Absolute and Relative user Perception of Classification Accuracy in an Affective Video Game". Interacting with Computers, vol. 29 No. 2, (March), 271-286.

Nacke, L.E.; Kalyn, M.; Lough, C.; and Mandryk, R.L. 2011. "Biofeedback Game Design: Using Direct and Indirect Physiological Control to Enhance Game Interaction". In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (Vancouver, May 07-12), ACM, NY USA, 103-112.

Nacke, L.E.; Bateman, C.; and Mandryk, R.L. 2014. "BrainHex: A Neurobiological Gamer Typology Survey". Entertainment Computing, vol. 5 No. 1 (Jan), 55-62.

Negini, F.; Mandryk, R.L. and Stanley, K.G. 2014. "Using Affective State to Adapt Characters, NPCs, and the Environment in a First-Person Shooter Game". IEEE Games Media Entertainment (Toronto, Canada Oct. 22-24), IEEE, --, 1-8.

Ng, Y.; Khong, C.W.; and Nathan, R.J. 2018. "Evaluating Affective User-Centered Design of Video Games using Qualitative Methods". International Journal of Computer Games Technology, vol. 2018 (June), 13 pages.

Nogueira, P.A.; Torres, V.; Rodrigues, R.; Oliveira, E.; and Nacke, L.E. 2016. "Vanishing Scares: Biofeedback Modulation of Affective Player Experiences in a Procedural Horror Game". Journal on Multimodal User Interfaces, vol. 10, No. 1 (March). 31-62.

Novak, D.; Mihelj, M.; and Munih, M. 2012. "A Survey of Methods for Data Fusion and System Adaptation Using Autonomic Nervous System Responses in Physiological Computing". Interacting with Computers, vol. 24, No. 3 (May), 154-172.

O'Neill, B. and Riedl, M. 2016. "Emotion-driven Narrative Generation". In Emotion in Games, Karpouzis K. and Yannakakis G.N. (Eds.). Springer, Switzerland , 167-180.

Plass, J.L.; Homer, B.D.; MacNamara, A.; Ober, T.; Rose, M.C.; Pawar, S.; Hovey, C.M.; and Olsen, A. 2019. "Emotional Design for Digital Games for Learning: The Effect of Expression, Color, Shape, and Dimensionality on the Affective Quality of Game Characters". Learning and Instruction. In press, https://doi.org/10.1016/j.learninstruc.2019.01.005

Popescu, A.; Broekens, J.; and Van Someren, M. 2013. "GAMYGDALA: An Emotion Engine for Games". IEEE Transactions on Affective Computing, vol. 5, No. 1 (Oct), 32-44.

Tang, S.; Hanneghan, M.; and Carter, C. 2013. "A Platform Independent Game Technology Model for Model Driven Serious Games Development". Electronic Journal of e-Learning, vol. 11 No. 1 (Feb), 61-79.

Togelius, J. and Yannakakis, G.N. 2016. "Emotion-driven Level Generation". In Emotion in Games, Karpouzis K. and Yannakakis G.N. (Eds.). Springer, Switzerland, 155-166.

Yannakakis, G.N. and Hallam, J. 2009. "Real-Time Game Adaptation for Optimizing Player Satisfaction". IEEE Transactions on Computational Intelligence and AI in Games, vol. 1, No. 2 (June), 121-133.

Yannakakis, G.N.; Martinez, H.P.; and Garbarino, M. 2016. "Psychophysiology in Games". In Emotion in Games, Karpouzis K. and Yannakakis G.N. (Eds.). Springer, Switzerland, 119-137.

Zafar, M.A.; Ahmed, B.; Al-Rihawi, R.; and Gutierrez-Osuna, R. 2018. "Gaming Away Stress: Using Biofeedback Games to Learn Paced Breathing". IEEE Transactions on Affective Computing, No.- (March), IEEE, 1.

AUTHORS' BIOGRAPHY

SALMA HAMDY obtained her degrees in computer science from Ain Shams University in Egypt. She is a lecturer at Abertay University in Scotland. Her research interests include computer vision, AI, and affective computing.
Email: s.elsayed@abertay.ac.uk.

DAVID KING is a lecturer in Maths and Artificial Intelligence at Abertay University teaching on the Computer Games Technology and Computer Games Application Development Programmes. Email: d.king@abertay.ac.uk.

Gender and Play in Goblin Dice

Hamna Aslam, Joseph Alexander Brown, Evgenii Nikolaev, Elizabeth Reading
Artificial Intelligence in Games Development Lab, Innopolis University, Russia, 420500
email: h.aslam@innopolis.ru, j.brown@innopolis.ru, e.nikolaev@innopolis.ru, beth_reading@hotmail.com

KEYWORDS
Board Game Design, Affordances, Gender and Game Design, Player Perceptions, User Testing

ABSTRACT

Ensuring gender equality in game design is important for creating equal opportunities for fun. This paper presents an investigation into which factors contribute to the gender-neutral game design. Goblin Dice has been examined by both male and female participants. The tested game has a racing theme and does not include either a soldier or a princess as the main character which is stereotyped to be more associated with either group of participants. The participants were asked to predict the rules of the game in a specified time without the rule book. The analysis of participants performance shows that the majority of the game's design was equally intuitive for both genders. While differences in the perception of male and female participants were observed towards the same object, females seeing it as cooperative, and males seeing it as a competitive aspect. Hence, gender bias in perceptions exists in games without explicit gender-based themes.

Introduction

The following is the reaction of a six-year-old girl who played the board game. *Guess Who*, in which there were five female characters and nineteen male characters:

> "It is not only boys who are important, girls are important too. If grown ups get into thinking that girls are not important, they won't give little girls much care. Also if girls want to be a girl in *Guess Who*, they will always lose against a boy and it will be harder for them to win. I am cross about that and if you don't fix it soon, my mum could throw, *Guess Who* out." The Huffington Post (2012)

The board game manufacturing company responded by assuring the little girl that they liked her suggestion and they will consider adding more female characters in the game. This motivates the demand for games with a *gender neutral design*- defined as:

> A design that ensures no bias towards any gender through number, color, appearance of objects as well as game mechanics.

Gender neutral design is an element of the human centered design process. The philosophy of human-centred design focuses on making the interactions between humans and objects as desirable as possible. A desirable interaction minimizes annoyance, frustration, and confusions, and leaves a positive impression. This research focuses on human centered design for games and considers gender explicitly.
While *gender aspect* is a broad term, our focus is to identify any patterns in design (if any) that either reduce or increase the desirability of the object interaction and to further investigate if such behaviour from a participant is gender-related or based upon individual preferences.
The tested game does not include objects and features in the design that have apparent inclination towards either gender. The game has been tested by two groups of participants: males and females. The intuitiveness of game design for both genders has been analyzed to: 1) Study the game design from a gender equality perspective; 2) Observe how and to which extent, male and female participants perceive the intuitiveness of game design and determine which factors, if any, make them understand the game mechanics correctly or incorrectly.
The remainder of the paper includes the game description that has been tested for the study, experimental setup discussion, results and analysis, the implication of research methodology and conclusions.

Context

Games are objects which have diversity in terms of their types, applications and design. Entertainment Standards Association (2014) shows demographics of computer and video gaming industry in 2014. The gaming industry has a diverse worldwide consumer base. Fifty-nine percent of Americans play video games with purchases of games divided fifty percent males and fifty percent females in the year 2014. In 2016 as presented by Entertainment Standards Association (2016), among the most frequent game purchasers, sixty percent were males, and forty percent were females. In 2017, the numbers changed to sixty-three percent males and thirty-seven percent females, Entertainment Standards Association (2017). One of the speculations for the decrease in the percentage of female purchasers from the year 2014 to 2017 is the failure to consider gender aspects in game designs. With the increasing popularity of gaming industry and a number of users, significant measures are required to make games serve the purpose they are intended to for all expected consumers irrespective of gender, age, or any other human factors. With the increased diversity of game players as shown in Entertainment Standards Association (2016), creating equal opportunities for all players to receive the intended benefits of the game has become an evolving issue, especially considering the gender element. This raises the question, if games are intended to cater both genders, are they being designed for both genders? Empirical testing of games is required to determine if the affordances are perceptible to both genders. The number of female characters in a game as compared to male characters is not the only concern that highlights the issue of gender equality in games. Laydehgad (2009) refers to the morally inappropriate representation of female characters in some video games. In contrast, Sue (2016) lists board games in which female characters are not over sensualized. Along with gender portrayal, gender aspect in game design also needs to be included, Stefansdottir and Gislason (2008) defines design process as placing and patterning of any act towards a desired goal and emphasizes on the inclusion of gender aspect in design innovation processes. Furthermore, Erb (2009) and Steiner et al. (2009) highlights the significance of the inclusion of gender-sensitive approach for designing educational games to ensure equal opportunities for learning for both genders. In this context, Steiner et al. (2009) presents a model that include factors (i.e. a reason to play, competition orientation, preferences, etc.) for the consideration of gender aspects in educational video game designs.

Gender studies are crucial to detect and understand the factors that can narrow gender bias and contribute to gender equality. In this regard, Holmlid et al. (2006) discusses the complexities and challenges that might arise during gender studies. While a simple task was given to two groups, one consisting of male participants and the other consisting of female participants, the difference in actions and priorities of participants has been observed. Also, the two groups showed different attitudes towards their instructors, where male participants showed refusal in following the instructions given by female instructors this could be anticipated as a gender reaction. Jenson and De Castell (2010) in their review of thirty years of research on gender and gameplay concludes that it is time to pursue gender research without making stereotypical assumptions in the beginning.

Methodology

Affordances

The concept of affordances is significant while studying any design. The term affordance was first introduced by Gibson in his book, *Ecological Approach to Visual Perception*, Gibson (2014). Gibson posits that affordances are part of the environment and are independent of the individual's perception. Hence, affordances are always in the environment to be perceived even when an individual cannot recognize them. Norman refers to affordance as a relationship between the properties of an object and the capabilities of the agent that determine just how the object could be used, Norman (2013). One such factor of the relationship is the gender of the agent. Browne (2015) discusses game design patterns and principles that can lead to games which are enjoyed by the players. The game objects should speak for themselves, which means to embed the rules in the game design itself. This encourages design elegance and clarity and hence gives the player an enjoyable experience. Some examples from games have also been demonstrated by Browne, where the game design has a "poke-yoke" effect, poke-yoke refers to reducing the player error by making a design that itself inhibit such errors. Identifying perceptions, Boschi et al. (2018) investigates peoples' perceptions of fairness about oddly shaped dice. Peoples' perceptions of fairness were identified to be influenced by their past usability experience. Yermolaieva and Brown (2017) showed that even with objects as simple as dice — the design could influence play in a game due to errors caused by the incorrect reading.

Playtesting

Game design exploration and investigation is an old practice. In the Persian *Book of Kings* (*Shāh-nāmeh*), the great poet Firdausi, gives an interesting account of how chess made its way from India to Persia, Warner and Warner (1909).

> As the story goes, in the sixth century, the Raja of India sent the shah a chess set made of ivory and teak, telling him only that the game was "an emblem of the art of war," challenging the shah's wise men to figure out the moves of the individual pieces. Of course, to the credit of the Persians (this being a Persian story), one of them was able to complete this seemingly impossible assignment. The Shah then bettered the raja by rapidly inventing the game of "nard" (a predecessor of backgammon), which he sent back to India with the same challenge. Despite its simplicity relative to chess, the intricacies of nard stumped the raja's men. This intellectual gambling proved to be extremely costly for the raja, who was obliged to pay a heavy toll: two thousand camels carrying "Gold, camphor, ambergris, and aloewood,/As well as raiment, silver, pearls, and gems,/With one year's tribute, and dispatched it all/From his court to the portal of the Shah,

Yalom (2004)."

Fifteen centuries later, Daviau (2011) used the same process by presenting a game to some participants without the rule book. The participants were asked to unfold the game rules by merely looking at the game objects. The focus was on investigating the intuitiveness of game design for the people who had never seen that game before. Daviau states:

> Rules should not explain a game; they should only confirm what the rest of the game tells you, Daviau (2011).

The game design investigation method with the removal of rule books has been demonstrated by Brown et al. (2019). It is observed that features such as game mechanics, game colours, and objects etc. contribute to developing a player's conceptual model about the gameplay. Moreover, Aslam et al. (2018) has applied the same methodology, for examining player's affordances and game intuitiveness for different age groups. The removal of rule books gave insight into player's affordances and perceptions about design elements. The identification of common interaction patterns of players within the same age group is advantageous for game designers to initiate an intuitive design process considering the player's age and preferences.

Tested Game

For the investigation of gender-sensitive design[1] in board games as well as analyzing the performance of male and female participants fairly, we had listed the following criteria about the game that is to be presented to the participants. We hypothesize, these criteria will provide an attempt at gender-neutral design.

1. The game objects should be gender neutral, at least in appearance.
2. The game should not include a dominant character object that can be classed as either a male (i.e. a soldier) or a female (i.e. a princess).
3. The game objects should have a mix of colours; not a predominately pink pallet, which is considered to be feminine, nor a blue pallet, stereotyped as masculine.
4. The game should preferably include popular board game objects such as tiles, a board, dice etc.
5. The basic theme of the game should be simple to understand, and does not require a prior game playing experience.

Based on the criteria mentioned above, we selected Goblin Dice by Bazylevich (2015), Figure 1. Goblin Dice is a board game that has twenty-two path tiles, one start and one finish tile, six goblins, one stone, one stone speed marker, and twelve dice. It is played between two to six players where goblins are racing across a path out running the bolder trying to crush them. The winner is either the first player to reach the final tile or in the event that the stone catches up to the goblins, sadly running them over and flattening them, the last goblin remaining.

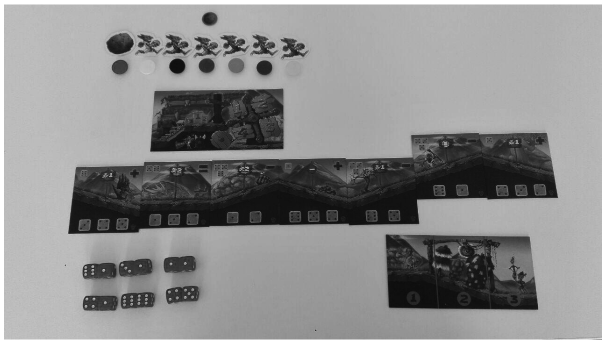

Figure 1: Game Setup

In the beginning, each player gets two dice. The dice are rolled into the middle of the board simultaneously. The dice can be used to perform one of two different actions. If the number on the die matches any of the number on the bottom of the path tile, a player can move a goblin forward. If the number on the die matches any number on the top of the path tile, a player can use special features of the tile. The game is played in rounds and continues till a goblin reaches the finish tile or all but one are flattened by the stone.

Experimental Design

The research methodology has been expanded from Daviau's anecdotal study setup, in which games were presented to participants without the rule books. In a specified amount of time, the participants had to guess the rules they could perceive from the game design and objects, Daviau (2011).

The study is approved by our institution's research ethics committee, has been conducted with eighty participants, thirty females and fifty males. The participants were the student of computer science bachelors and masters program. The average age of male and female students is twenty. The males have spent an average of approximately twelve hours playing board games and an average of approximately thirty five hours playing digital games in the past thirty days. The females have spent an average of approximately six hours playing board games and the same for digital games in the past thirty days.

The game with the rule book removed, was presented to each participant individually. The participants were instructed to analyze the game for fifteen minutes and

[1] This study assumes a gender binary (male/female); participants gave an anonymous and self-identified response to their gender

after that they had to fill a questionnaire The questionnaire examines the following: (1) Questions about their game play hours in the past thirty days (2) Game mechanics, game and winning condition etc.

Results and Analysis

The responses obtained from both male and female participants are summarized in Table 1.

Quantitative Analysis

In order to provide a quantitative evaluation of the rules, the elements which were highlighted by multiple participants were classified together, (in all) there were thirteen frequently made statements about the game on nine topics, see Table 1. In order to determine the significance of the response, several common responses were summed into binomial data; the user wrote this statement or did not write the statement. Statistical tests were undertaken with a two-tailed z-score test for a population proportion. The z-score test for two population proportions is used to determine whether two populations, i.e. males and females, differ significantly on some categorical characteristic, e.g. if they said *Goblin Dice* has six players, the Null Hypothesis is that no difference can be determined between the responses. We apply a Bonferroni correction to the tests in order to remove the likelihood of the family of comparisons being in error. Given that seven topic classes are examined in order to have a $\alpha < 0.05$ for all tests, then we must have a $\alpha < 0.007301$ for each test.

One of the factors is statistically significant; females stated that *Goblins are competing or playing a football game* ($p = 0.0010$), which is not correct to the actual game mechanics. While not statistically significant males were more likely to state that goblins were running away from the stone ($p = 0.2420$) and state that the stone will kill the goblin ($p = 0.0151$). Taken together these findings shows that females believe the stone to be in a more cooperative relationship with the Goblins. Whereas, males believed it to be a threat or a competitive element.

In this study, the methodology removed the context of the stone which is visible via the theme, i.e. box art showing the goblins running away. Hence, males and females have reverted into their preconceptions of the mental model in terms of interactions. The default conceptual model in this case for males depicts, favouring the competitiveness in-game aspects — stone as threat — and females favouring cooperation actions — stone as a football.

This leaves an open question as to how the designers of Goblin dice could have avoided this confusion. Note that the themes of the goblins are that of sporting competition, they wear gym outfits, which fits both of the narratives of a race and a football game. The stone is the factor of confusion, the small addition of a goblin perhaps flattened by the stone behind it, or just before it looking back in horror perhaps could have clarified the purpose.

Qualitative Analysis

The observational and qualitative analysis shows that male and female students have used the same approach while analyzing the game. The game objects were counted and similar objects were placed together.

The maximum number of players for the tested game, Goblin Dice, have been predicted by counting the goblins. However some of the participants also included the stone and the speed marker in the counting and assumed that there are maximum seven or eight players for Goblin Dice. Some participants also suggested three and four players because there are three dice symbols at the bottom of some path tiles, shown in Figure 1, and they thought that twelve dice are divided among players and each player gets three or four dice. The correlation between the dice and the movement of goblins was understood by most of the participants as fifteen out of thirty female students and fourteen out of fifty male students could predict that after throwing the dice, the numbers on the dice are compared to numbers on path tiles to decide further action.

The participants predicted Goblin Dice to be a competitive or a racing game in which a goblin has to reach the finish tile before others. However, a different idea suggested by six female participants was it to be a goblin football or a goblin soccer game in which the goblins are divided into three teams of two players and are trying to play football, rolling the stone to the finish tile. In contrast, the male participants mostly suggested that goblins are trying to run away from the stone.

Considering the stone being a football is Gibson's affordance which refers to all the action possibilities with an object. A stone can be kicked and rolled, however considering its physical properties, it is not suitable for soccer which is played using the ball. The female participants, in this case, are implying Gibson's idea of affordance as well as depicting their perceived affordance for the stone which is not an actual affordance. The goblins also look as if they are kicking a ball with one foot; this can also be the reason why females perceived it to be a football game. The gap between the perceived affordance and actual affordance highlights the need of a signifier to match the goal of the design. Furthermore, discussing the correct prediction and the accurate mapping of the object's intended functionality, three female and ten male participants mentioned that goblins are running away from the stone. These responses demonstrate that the stone's affordance was perceivable by these participants. The exclusion of the rule book played an essential role in this study, without which it would not have been possible to determine possible interpreta-

Responses (C-Correct to Game Rules; I - Incorrect to Game Rules)	Number of Responses from Female Participants (30)	Number of Responses from Male Participants (50)	p-value (two-tailed)
The game is played by max six players (C)	20	36	0.6171
The game is played by seven players (I)	0	4	0.1118
The game is played by max four players (I)	7	8	0.4179
The game is played by max three players (I)	3	2	0.2846
Each player should use two dice (C)	9	14	0.8493
Each player takes three dice (I)	5	12	0.4354
Signs on the tiles shows how to move (C)	7	13	0.7872
Numbers on the dice can be compared to numbers on tiles to decide action (C)	15	14	0.0477
Goblins are running away from a stone (C)	3	10	0.2420
Goblins are competing or playing a football game (I)	6	0	**0.0010**
Goal is not to die from a stone (C)	1	12	0.0151
Goal is to bring stone to the finish tile (I)	4	4	0.4413
Player who reaches the finish tile is the winner (C)	15	26	0.8630

Table 1: Comparison of Responses from Male and Female Participants with z-score Test for Two Population Proportions. Correct statements to the rules marked with (C) and incorrect statements marked with (I). Statistically significant with $p < 0.007301$, for seven classes, as per the Bonferroni correction highlighted in bold

tions of game objects which can significantly influence a player's experience. This can provide useful insight into the game design to add signifier and clues where needed and to remove unnecessary confusion.

The difference in male and female participants' thinking, such as stone is a destructive object or an enemy as anticipated by males and stone is a fun object and an ally or a reason to win, as predicted by females, can be attributed to gender rather than merely an individual opinion. This might support the gender stereotype as females being cooperative, thinking stone as a friend and males being competitive, thinking it as a threat. However, in order to concretely prove or disprove this point, an investigation with larger sample sizes and different age groups is required.

Based upon the observation during the test and qualitative analysis, we could not see any difference in the performance of participants which can show that either group was better than the other. The difference in gender, based on the performance of participants in the test is not apparent. The game is liked by both groups and they have expressed their desire to play the game after the test. Participants were also excited to know the actual rules and most of them requested the rule book from the observer to find out the actual rules.

Investigating the game design, results show that the design is intuitive to lead participants in understanding that the game involves the movement of goblins and the stone. The dice and the pictures of dice on the tiles show a relation between throwing the dice and comparing the outcomes with the dice pictures on the tiles to decide further actions. There are no game objects that have misled participants towards an exact inverted mapping of the actual game mechanic. The only exceptional case was when few participants considered taking the stone to the finish line instead of goblins, as they thought it to be a goblin football game. The study with the current sample size shows that game design is intuitive for both male and female participants and does not present any bias towards either gender.
Referring to the two questions in the Introduction section, an analysis of game design based upon participants responses suggest that there exist;

1. No bias towards any gender in-game objects to design and representation, other than the stone which was perceived as an enemy by male participants and a friendly object by female participants;
2. The game design was intuitive for most of the participants to catch the broader theme of the game mechanics. For example, participants understood that gameplay involves goblins mobility because of the path tiles and goblins picture over the tiles (goblins appear to be running). The association between dice rolls and number on path tiles was also perceivable by most of the players.

The comments from both male and female groups who did not find the game to be fun, show that it had to do with their individual preference for games and does not reflect any difference in liking or disliking of the game because of gender reasons.

Implications of Research

The research methodology adopted in this paper is a playtesting process with a change in the environment of the object under testing. The change in the environment of the game is facilitated by the removal of the rule book. It is necessary for a designer to determine the people's perception of the objects and play mechanics. The goal of the testing methodology is:

1. The identification of the maximum possible interpretations and perceptions about object usability. This ensures avoiding a design dimension that can lead to hurtful or unpleasant experiences on the user's side.
2. The selection of appropriate and recognizable signifiers/clues to ensure the correct usability such as letting users acquire the real affordances (usability intended by the designer) of the objects and designs.

Point (1), is referring to Gibson's notion of affordances, as to all possible usability manners, regardless of if they are the object's real affordances or not. By not having guidance through rule books, players could infer any possible interaction with the game objects, which is giving a more significant set of usability and interaction patterns. This also measures the change in the game environment based upon the existence and non-existence of rules. The goal is to analyze different interpretations to filter out any undesirable aspect in the design.

Since human beings are different, and they may associate different perceptions and interpretations for the same object usability based upon their culture, gender, age and any other human factor. Therefore, an approach towards understanding players is enabling them to play with the design with the freedom of making any move and interaction they deem possible. The demonstrated methodology is a non-competitive process. A non-competitive playtesting environment encourages players to try out various action possibilities. As there are no rules imposed, players are free to use an object in any way they think it can be used. This is a significant observation for the designers as they can see incorrect mappings of their designs and embed careful signifiers to avoid all incorrect mappings. Although, knowing all sorts of perceptions toward objects and their usability is complicated and unachievable. Through playtests, we can approximate common beliefs, perceptions and interpretations.

When people are interacting with the design, the interpretations of usability might differ based upon the gender of the individual. Designers can consider this for avoiding frustration and any bias towards either gender. Point (2) highlights Norman's goal for achieving a design that naturally leads users to correct usability. The results of the playtesting process help designers to make a comparison between user's perceived affordances about the object and real affordances, such as which part of the game objects have directed players toward the correct mapping of game mechanics and which objects have led player towards the incorrect or opposite mapping of actual game mechanics.

A study of game design from Gibson's point of view of affordances is significant to understand all possible actions associated with game objects and analysis of Norman's view of affordance helps in the recognition of player's level of understanding of game mechanics as well as their interpretations of game objects. Designing games which are enjoyable for all players irrespective of gender requires the recognition of all possible interpretations and associations among game objects, for example, while playing Guess Who, a little girl perceived that female characters would mostly lose because there are more male characters in the game to choose from. Game mechanics in Guess Who does not have a gender bias. The bias exists in a set which has one trait more determinant than the other. The little girl's strategy, in this case, is probably to only select a female character, and because of this set is smaller, it makes the final determination of the character selected a more accessible pathway. The mechanic is not biased to either gender, but the theming of the game is what has introduced gender issues for the girl. Guess Who in order to be fun must be discriminated because otherwise if it has an equal distribution of all traits, then the time to the solution will mathematically be a constant number of guesses. So, therefore, in order to be fun Guess Who has to be discriminatory by providing an unbalanced set of traits. At the same time, human factor's impact on the level of player's enjoyment should be investigated to take effective design decisions consequently, and game design can be modified according to the preferences and needs of the target audience.

Acknowledging that a single or a few players' interpretations of game design cannot speak for the whole game audience, still, recognition of such possible interpretations is significant for improvising game designs for clarity and exploring various play dimensions. This increases possibilities for making the accurate themes for the same game but for different audiences so that player feels that they are a part of the system and can immerse in the play.

Conclusions

The motivation for this study is to make the user experience better for games for all players. The study focuses on investigating the aspect of gender in gaming such as which factors in game design introduces or removes the gender bias in the game. The game selected for testing does not include any objects that are stereo-typically inclined towards either gender. The male and female groups have performed the same in guessing the broader theme of the game and game mechanics. However, a difference in performance was observed when males referred to an object in the game as a threat and females speculated it to be an ally. The difference highlights the fact that even with an attempted gender-neutral design, males and females demonstrated a different perception of the same object. This draws attention to the idea that for the inclusion of gender aspects in gaming to make them fun and create equal opportunities to acquire intended benefits of the game, it is not only object representation that is important, but how males and females perceive objects and associate meanings to them is a significant factor. The research methodology applied enables us to understand player perceptions about game objects and design. The game designers can apply this process to ensure its best suitability for the intended audience.

With the tested sample size, the study uncovers Goblin Dice to be intuitive for both male and female participants. We could see a difference in the perception of one of the objects of the game that can be speculated to be because of the gender. The playtesting method adopted for testing Goblin Dice assist game designers to test the clarity of game mechanics through game objects and to identify possible interpretations and play dimensions under the influence of human factors such as gender, age, etc. Though this does not mean, removing complexities and unpredictability that are sometimes crucial for an adventurous play journey but that designers must remove confusions that are unnecessary taking into account the player's needs, emotions and expectations. For future work, we will run the test with bigger sample sizes and with different age groups of participants. We will consider other games which have objects showing the dominance of a particular gender in representation or characteristics, to compare the participant's performance in an apparent gender neutral and biased game. Further, we want to examine more structured questioning methods, e.g. Brown (2019).

REFERENCES

Aslam H.; Brown J.A.; and Reading E., 2018. *Player Age and Affordance Theory in Game Design*. In *19th International Conference on Intelligent Games and Simulation, GAME-ON 2018*. Eurosis, 27–34.

Bazylevich M., 2015. Goblin Dice. Game [Table Top], Foxgames and GaGa Games. Poland.

Boschi F.; Aslam H.; and Brown J.A., 2018. *Player Perceptions of Fairness in Oddly Shaped Dice*. In *Proceedings of the 13th International Conference on the Foundations of Digital Games*. ACM, New York, NY, USA, FDG '18. ISBN 978-1-4503-6571-0, 58:1–58:5. doi:10.1145/3235765.3236389.

Brown J.A., 2019. *A Focused Conversational Model for Game Design and Play-Tests*. Manuscript Under Review.

Brown J.A.; Aslam H.; Makhmutov M.; and Succi G., 2019. *Intuitive Rules Design Evaluation Methods and Case Study*. In M. Guzdial; J.C. Osborn; and S. Snodgrass (Eds.), *Proceedings of the 2nd Workshop on Knowledge Extraction from Games co-located with 33rd AAAI Conference on Artificial Intelligence, KEG@AAAI 2019, Honolulu, Hawaii, January 27th, 2019*. CEUR-WS.org, *CEUR Workshop Proceedings*, vol. 2313, 35–42.

Browne C., 2015. *Embed the Rules*. *Game & Puzzle Design*, 1, no. 1, 60–70.

Daviau R., 2011. *Design Intuitively*. In *The Kobold Guide to Board game design*, Open Design, Kirkland, Washington. 42–49.

Entertainment Standards Association, 2014. *2014 Essential Facts about the Computer and Video Games Industry*. Tech. rep., ESA.

Entertainment Standards Association, 2016. *2016 Essential Facts about the Computer and Video Games Industry*. Tech. rep., ESA.

Entertainment Standards Association, 2017. *2017 Essential Facts about the Computer and Video Games Industry*. Tech. rep., ESA.

Erb U., 2009. *Reflections on Gender Aspects of Designing an Educational PC-Game exemplified by a Project in Cooperation with the German Maritime Museum Bremerhaven*. In *5th European Symposium on Gender & ICT, University of Bremen, German*.

Gibson J.J., 2014. *The ecological approach to visual perception: classic edition*. Psychology Press.

Holmlid S.; Montaño F.; Johansson K.; Datavetenskap N.D.O.I.; and Datavetenskap N.D.O.I., 2006. *Gender and design: Issues in design processes. Proceedings from Women in information technology, University of Salford*.

Jenson J. and De Castell S., 2010. *Gender, simulation, and gaming: Research review and redirections*. *Simulation & Gaming*, 41, no. 1, 51–71.

Laydehgad, 2009. *Gender in Game Design*. URL `https://laydehgad.wordpress.com/`. Online; accessed 2 April 2017.

Norman D.A., 2013. *The Design of Everyday Things*. Basic Books, expanded ed. ed.

Stefansdottir S. and Gislason H., 2008. *Design Innovation for Gender Equality*.

Steiner C.M.; Kickmeier-Rust M.D.; and Albert D., 2009. *Little big difference: Gender aspects and gender-based adaptation in educational games*. In *International Conference on Technologies for E-Learning and Digital Entertainment*. Springer, 150–161.

Sue T.M., 2016. *Climbing Chutes and Ladders, The Positive Representation of Women in Board Games*. URL `https://www.themarysue.com/women-in-board-games/`. Online; accessed 2 April 2017.

The Huffington Post, 2012. *Jennifer O'Connell, Mom, And 6-Year-Old Daughter Ask Hasbro About Gender Inequality In Guess Who?)*. URL `http://www.huffingtonpost.com/2012/11/21/jennifer-oconnell-hasbro-guess-who_n_2165482.html`. Online; accessed 2 April 2017.

Warner E. and Warner A.G., 1909. *Shahnama Of Firdausi, trans.* Kegan Paul, Trench, Trubner And Co., London.

Yalom M., 2004. *Birth of the chess queen: a history*. Pandora.

Yermolaieva S. and Brown J.A., 2017. *Dice Design Deserves Discourse*. *Game & Puzzle Design*, 3, no. 2, 64–70.

GAME
AI

USING A GENETIC ALGORITHM FOR THE PROCEDURAL GENERATION OF LAYERED MATERIALS FOR REAL-TIME RENDERING

Alessio Bernardi[†], Davide Gadia[*], Dario Maggiorini[*], Laura Anna Ripamonti[*]
Department of Computer Science, University of Milan
via Celoria 18, 20133 - Milan (Italy)
Email: [†]{firstname.lastname}@studenti.unimi.it, [*]{firstname.lastname}@unimi.it

KEYWORDS

Procedural Content Generation, Computer Graphics, Layered Material, Genetic Algorithm, Real-Time Rendering

ABSTRACT

The use of Procedural Content Generation techniques in the production of video games has seen a large diffusion in the last years. However, in the Computer Graphics field, very few works have addressed the procedural generation of Bidirectional Scattering Distribution Functions for complex materials. In this paper, we present a method for the automatic generation of realistic layered materials, based on the application of a Genetic Algorithm. We show that, with the proposed approach, it is possible to generate several instances of a target material, maintaining a desired level of closeness to the original simulated interaction between the light and the surface, but introducing also a controlled amount of differences in the final perceived appearance.

INTRODUCTION

Procedural Content Generation (PCG) techniques have seen an increasing application in the production of video games. Several works have shown the potentialities of PCG, combined also with AI and evolutionary techniques (Shaker et al., 2016), for the automatic creation of contents like e.g., game levels (Ripamonti et al., 2017b; Mazza et al., 2017; Mourato et al., 2011), the features of a character (Guarneri et al., 2013; Norton et al., 2017), or the impact of Non Player Characters (Ripamonti et al., 2017a; Yannakakis and Hallam, 2009).
In the Computer Graphics (CG) field, there is a relevant literature on the use of PCG for the creation of complex models like buildings (Schwarz and Müller, 2015) or cities (Scalabrin et al., 2016). However, with the exception of procedural texturing (Ebert et al., 2003), few works have been proposed on the procedural generation of models surface appearance. In this paper, we show how a Genetic Algorithm (GA) can be efficiently applied to an advanced model for real-time rendering of layered materials. The main idea is to generate several versions of a target material by evolving the parameters of the function calculating the interaction between the light and the surface layers, and evaluating their effectiveness by considering the perceptual differences with the original material. This approach can be used when several instances of a model are generated in a large virtual world. The material of each instance must share a common "physical" behaviour (for example, a rusty metal effect), but we want to introduce a certain amount of perceptual differences in each instance, in order to enhance the variety of the generated scene.
The paper is structured as follows. In the next two sections, we describe the mathematical functions used to represent a material in real-time rendering, and we provide an overview of PCG techniques for the generation of materials in CG. Then, we describe the proposed PCG method and its experimental evaluation. Finally, we draw conclusions and discuss future developments.

MATERIALS REFLECTANCE IN CG

The rendering process of a CG scene is based on the simulation of the physical interaction between light and the materials of the surfaces. For a uniform, not-emitting material, the light reflected by a surface in the direction of the virtual camera is given by the *rendering equation* (Pharr et al., 2016):

$$L_o(\vec{\omega_o}) = \int_{\vec{\omega_i} \in \Omega} f_r(\vec{\omega_i}, \vec{\omega_o}) L_i(\vec{\omega_i}) (\vec{\omega_i} \cdot \vec{n}) \partial \vec{\omega_i} \qquad (1)$$

where

- $\vec{n}$ is the surface normal
- $\vec{\omega_i}$ and $\vec{\omega_o}$ are, respectively, the directions of incoming and reflected light
- $L_i(\vec{\omega_i})$ and $L_o(\vec{\omega_o})$ are, respectively, the incoming and the reflected radiances
- $f_r(\vec{\omega_i}, \vec{\omega_o})$ is the *Bidirectional Reflectance Distribution Function (BRDF)*, which describes the amount of radiance reflected in the direction $\vec{\omega_o}$, given the radiance coming from direction $\vec{\omega_i}$

Among the possible approximations of the rendering equation, the *microfacets* approach (Akenine-Möller et al., 2018) is based on the idea that surfaces are composed by a large collection of microscopic facets bouncing light in different directions. Thus, the scattering properties of a material are described by the statistical distribution of the microfacets orientations. Following this approach, several methods for real-time rendering of complex materials use equation (2) to define the BRDF:

$$FGD = f_r(\vec{\omega_i}, \vec{\omega_o}) = \frac{F(\vec{\omega_i} \cdot \vec{h})G(\vec{\omega_i}, \vec{\omega_o})D(\vec{h})}{4(\vec{\omega_i} \cdot \vec{n})(\vec{\omega_o} \cdot \vec{n})} \quad (2)$$

where

- $\vec{h}$ is the *half vector* between $\vec{\omega_i}$ and $\vec{\omega_o}$, used to approximate the direction of specular reflection
- $F()$ is the *Fresnel reflectance term*, which describes the amount of reflection and trasmission of light
- $G()$ is the *Shadowing-Masking Function*, which accounts for microfacets self-occlusion effects
- $D()$ is the *Normal Distribution Function (NDF)*, which describes the statistical distribution of the microfacets normals

Among the several proposed NDFs, currently the most used is the *GGX distribution* (Walter et al., 2007):

$$D = GGX(\vec{n}, \vec{h}, \alpha) = \frac{\alpha^2}{\pi\Big((\vec{n} \cdot \vec{h})^2(\alpha^2 - 1) + 1\Big)^2} \quad (3)$$

where α controls the overall *roughness* of the surface. When both reflection and transmission are considered, then $f_r()$ in equation (2) is called *Bidirectional Scattering Distribution Function (BSDF)*, rather than *BRDF*, because the latter describes only the reflectance.

RELATED WORK

To our knowledge, PCG techniques for the generation of BRDFs or BSDFs have not been extensively investigated. Brady et al. (Brady et al., 2014) proposed a framework for learning new analytic BRDF models through Genetic Programming (GP). They use as initial population the set of expressions of different BRDFs, and as target a trained set of measured materials from a freely available database. During the pairing process, different symbolic transformations could be applied to the parameters and operations of the two parent BRDFs, in order to create a new, more complex, reflectance function. Then, the fitness of the generated formulations is evaluated considering an error function with respect to the training set. A GP approach is used also in the work by Sitthi-Amorn et al. (Sitthi-Amorn et al., 2011). In this work, the evolutionary method is applied to increasingly simplify the source code of a shader, in order to find the optimal compromise between rendering speed and accuracy of the light-material interaction. To evaluate the accuracy, a per-pixel color difference metric is applied between the images generated using the original shader, and the images created using the simplified offsprings. Masia et al. (Masia et al., 2009) have applied a GA to determine the reflectance characteristics of an object in an image. The parameters of two well-known BRDFs are used as chromosomes, and an initial population is created assigning random values to these parameters. At each iteration, fitness is evaluated by rendering an image for each chromosome created in each generation, and calculating a per-pixel difference between the target and the rendered image.

METHODS

In this paper, we apply a GA to evolve the parameters of a recent computational model for the rendering of *layered materials*. In this section we stepwise go through our approach, by describing the considered BSDF model, and by stating the principal parts of the GA: representation, algorithm and fitness functions.

The considered BSDF for layered materials

Layered materials are composed by different layers, each with peculiar reflectance characteristics. Examples of layered materials are car paint, rusty metal, etc. Techniques for the simulation of layered materials usually consider different BSDFs for each layer, and then apply some kind of blending to approximate the scattering of light between the layers (Akenine-Möller et al., 2018). Recently, a work by Belcour (Belcour, 2018) has proposed a novel framework for the simulation of light transport within layered materials. Starting from the GGX distribution (equation (3)), a set of atomic statistical operators are proposed to describe reflection, refraction, volume scattering and absorption, starting from the *energy*, *mean*, and *variance* of the BSDFs at each layer. The atomic operators for each layer are then combined, considering also the possible presence of a participating media among the layers. The result after the combination is used to instantiate a single BSDF approximating the complex light scatterings within the original layers. Table 1 summarizes the atomic operators. We refer to (Belcour, 2018) for a more detailed description of the method.

Representation

Belcour's method represents an excellent candidate for an evolutionary technique for the automatic generation of layered materials, due to its compact description, and limited and intuitive set of variables. In (Belcour, 2018), only a top and bottom layers are considered for real-time rendering , with the latter representing an approxi-

Table 1: Statistical atomic operators of Belcour's method (Belcour, 2018), used to approximate the outgoing energy e, mean μ, and variance σ given incident e_i, μ_i, and σ_i (the square on the variance is omitted for better readability). R and T are used for a reflected or transmitted parameter. α is the layer roughness, s is a roughness scaling factor for the trasmission, η_{12} is the ratio of the refractive indices, h is the depth of the layer, σ_t is the transmittance cross-section, σ_s is the scattering cross-section, and σ_g accounts for the increase in variance due to the width of the phase function.

	Reflection	**Refraction**	**Absorption**	**Scattering**
energy	$e^R = e_i \times FGD$	$e^T = e_i \times (1 - FGD)$	$e^T = e_i \cdot e^{-\frac{\sigma_t \cdot h}{\sqrt{1-\lvert\mu_i\rvert^2}}}$	$e^T = e_i \cdot \frac{\sigma_s \cdot h}{\sqrt{1-\lvert\mu_i\rvert^2}} \cdot e^{-\frac{\sigma_t \cdot h}{\sqrt{1-\lvert\mu_i\rvert^2}}}$
mean	$\mu^R = -\mu_i$	$\mu^T = -\eta_{12}\mu_i$	$\mu^T = -\mu_i$	$\mu^T = -\mu_i$
variance	$\sigma^R = \sigma_i + f(\alpha)$	$\sigma^T = \frac{\sigma_i}{\eta_{12}} + f(s \times \alpha)$	$\sigma^T = \sigma_i$	$\sigma^T = \sigma_i + \sigma_g$

mated combination of multiple other scattering phenomena. Moreover, an optional partecipating media acting as third layer can be considered. In our GA, we follow the same approach. As a consequence, the genotype is composed by 12 floating point chromosomes:

- η_1: refractive index of the top layer
- η_2: refractive index of the bottom layer
- α_1: roughness of the top layer
- α_2: roughness of the bottom layer
- h: the depth of the partecipating media layer between the top and bottom layers
- σ_s^R, σ_s^G, σ_s^B: scattering cross section values
- σ_a^R, σ_a^G, σ_a^B: absorption values (parameters for σ_t in Table 1)
- g: anisotropic factor (parameter for σ_g in Table 1)

If there is no participating media, then the final appearance is controlled only by η_1, η_2, α_1 and α_2 chromosomes. If a texture is applied in one of the layers, then the roughness value is extracted from the texture, and α_1 or α_2 are used as weights applied to this value, in order to control the overall roughness effect.
Even if, theoretically, the values of the parameters can be set without constraints, there are some rules to follow in order to create physically realistic materials. There are differences in the possible value ranges, on the basis of the *family* of the desired material. We call *roughcoat* a material with a smooth bottom layer (a metal or a plastic), covered by a rough top layer (rust, dirt, etc.). This kind of material has a high value of roughness in the top layer, and a lower one in the bottom. A *clearcoat* material is used to simulate materials like e.g., car paint, ceramic. In this case, the roughness value of the top layer is lower than the value of the bottom layer. Moreover, there are further differences, according if the bottom layer is a *conductor* (like e.g., metal) or a *dieletric* (like e.g., plastic) material. We have summarized these rules and constraints, determined empirically via pre-experimentation, in Table 2.

Algorithm

Initialization
The proposed GA follows a standard approach. Given a target material, an initial population of N individuals of the same material family is created. As shown in the *Experiments* section, an adequate value for N is between 50 and 100. The values of the chromosomes of each individual are generated randomly, but respecting the rules of Table 2. Each individual is then ranked using a fitness function (described in a following section).

Selection
We apply *tournament selection* with tournament size 4 and $p = 1$. A deterministic tournament leads to the selection of only the best individuals: however, considering only a limited set of participants allows to maintain a high level of variety in the selected materials.

Crossover
Each couple of parents has a 0.9 probability to generate offsprings. We set a very high probability in order to further enhance the variety of the population. If crossover is applied, we apply a *uniform crossover*: each chromosome has a 0.25 probability of being swapped among the two parents. With this value, at least one chromosome is almost always swapped, but it is highly improbable to have crossover applied to all the chromosomes at the same time. To avoid issues with offsprings not following the rules of Table 2, we apply two steps in the selection of individuals for the crossover operation. In the first step, for each individual we search in the population the first material with refractive indices suitable to

Table 2: Rules and constraints for the values of the material chromosomes.

	Bottom Layer: conductor	**Bottom Layer: dielectric**	**Layer with textured roughness**
roughcoat	$\alpha_1 \geq 0.1$ $\alpha_2 < 0.1$ $1.0 < \eta_1 \leq 2.0$ $\lvert\eta_1 - \eta_2\rvert \in (0.0, 1.5]$[a]	$\alpha_1 \geq 0.1$ $\alpha_2 < 0.01$ $1.0 < \eta_1 \leq 2.0$ $\lvert\eta_1 - \eta_2\rvert \in (0.0, 0.5]$[b] $\eta_2 \geq 1.0$	top layer
clearcoat	$\alpha_1 < 0.1$ $\alpha_2 \geq 0.1$ $1.0 < \eta_1 \leq 2.0$ $\lvert\eta_1 - \eta_2\rvert \in (0.0, 1.5]$[a]	$\alpha_1 < 0.01$ $\alpha_2 \geq 0.01$ $1.0 < \eta_1 \leq 3.0$ $\lvert\eta_1 - \eta_2\rvert \in (0.0, 3.0]$[b] $\eta_2 \geq 1.0$	bottom layer
roughcoat/clearcoat with participating media	$h \in [0.05, 40.0]$[c] $\sigma_s^R, \sigma_s^G, \sigma_s^B \in [0.0, 1.0]$ $\sigma_s^R, \sigma_s^G, \sigma_s^B \in [0.0, 1.0]$ $\eta_2 < 1.0$[d]	$h \in [0.05, 40.0]$[c] $\sigma_s^R, \sigma_s^G, \sigma_s^B \in [0.0, 1.0]$ $\sigma_s^R, \sigma_s^G, \sigma_s^B \in [0.0, 1.0]$	—

[a]we set η_2 slightly less than η_1 to favor the internal reflection, without the dependance from the attenuation factor of the refractive index.
[b]we set η_2 greater than η_1 to favor light transmission for subsurface scattering.
[c] $h \in [0.0, 1.0]$ favors the color given by scattering cross section (with less absorption). $h \in [2.0, 40.0]$ favors the color given by transmission cross section. With $h = 40.0$: full light absorption: only the top layer is visible.
[d] with $\eta_2 < 1.0$ the effect of the partecipating media on the color is more visible. But increasing the value of η_1, and lowering the value of η_2 such as $\eta_2 << 1.0$, the original color of the metal progressively disappears.

be swapped (i.e., respecting the constraint on $|\eta_1 - \eta_2|$). If this individual is found, crossover is applied. In the second step, a matching process is applied sequentially to create couples from the individuals excluded from the first step. For these couples, we apply uniform crossover excluding from the process the chromosomes related to the refractive indices.

Mutation
Mutation on the generated offsprings can occur with a 0.4 probability . If activated, each chromosome has a 0.1 probability to mutate. If a chromosome is subject to mutation, a new value is created randomly, but respecting the rules of Table 2.

Termination
Considering the goal of the proposed GA, we have set a fixed number of generations as the termination condition of the evolutionary process.

Fitness functions

The GA approach can help to speed up the generation process of the most appropriate material instances. Simpler methods based e.g., on a random perturbation of the BSDF parameters, may generate a higher number of unwanted/inadequate samples, requiring thus a longer time to converge to the desired result. Considering the final goal, the fitness function must be chosen in order to measure in a simple but effective way the distance between the target and generated genotypes, yet allowing to adapt the effect of the proposed method to different situations and applications (e.g., applying different weights to specific subsets of chromosomes). Thus, we have decided to consider two well-known vector distances applied between the target material and the generated individual, the *Chebyshev distance*

$$d_{ch}(g, tm) = \max_i \{|g[i] - tm[i]|\} \tag{4}$$

and the *Euclidean distance*

$$d_{eu}(g, tm) = \sqrt{\sum_i (g[i] - tm[i])^2} \tag{5}$$

where g is an individual, tm is the target material, and i represents each of the chromosomes.

In the next section, we present an experimental evaluation of the effect of the two distances on the final perceptual differences of the generated materials.

EXPERIMENTS

In order to evaluate the effect of the considered fitness functions in the selection of the best individuals in the population, we have set up a test scene consisting of a sphere, illuminated by a single light. We have then selected two target materials:

- **validation metal**, a clearcoat metal with no texturing. The values for the chromosomes are $\eta_1 = 1.2$, $\eta_2 = 0.8$, $\alpha_1 = 0.03$, $\alpha_2 = 0.1$
- **validation painted metal**, a roughcoat metal with a roughness texture applied in the top layer. The values for the chromosomes are $\eta_1 = 1.2$, $\eta_2 = 0.8$, $\alpha_1 = 2.5$, $\alpha_2 = 0.099$

We have decided to not consider the presence of a participating media in this experimental setup, because it usually has a relevant effect on the final color of the material, thus providing a more evident variety in the generated materials. Without the participating media, any perceptual difference between the original and generated materials is given only by the core chromosomes of the top and bottom layers, which represents a more tricky situation to manage. Figures 1 and 2 show the test scene with the two target materials.

For both the target materials, we have generated a set of new materials using the proposed GA, applying both the fitness functions. For each combination of target material and applied distance, we have performed 36 executions of the GA with a population of 50 individuals and termination after 5 generations. For each execution, we have then selected the material with the best fitness (i.e., with minimum distance to the target material). Figures 5, 6, 7 and 8 show some examples of these generated materials.

To evaluate the differences among the generated and target materials, we have applied the *CIEDE2000* $\Delta E_0^* 0$ difference (Mokrzycki and Tatol, 2011), a measure proposed in colorimetry for the perceptual difference among two colors. $\Delta E_0^* 0$ value ranges from 0 and 100.

- $\Delta E_0^* 0 \leq 1.0$: color difference is not perceptible by human eyes
- $\Delta E_0^* 0 \in [1.0, 2.0]$: color difference is perceptible through close observation
- $\Delta E_0^* 0 \in [2.0, 10.0]$: color difference is perceptible at a glance
- $\Delta E_0^* 0 \in [10.0, 49.0]$: colors are more similar than opposite
- $\Delta E_0^* 0 \in [49.0, 100.0]$: colors are exact opposite

We have applied $\Delta E_0^* 0$ between each pixel of the image rendered using the target material, and the corresponding pixels in each of the images created using the

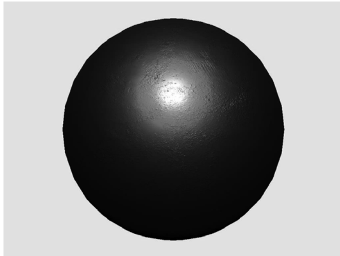

Figure 1: Test scene with **validation metal** material. Chromosomes: $\eta_1 = 1.2$, $\eta_2 = 0.8$, $\alpha_1 = 0.03$, $\alpha_2 = 0.1$.

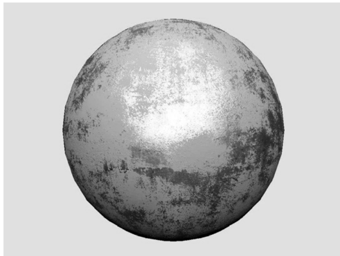

Figure 2: Test scene with **validation painted metal** material. Chromosomes: $\eta_1 = 1.2$, $\eta_2 = 0.8$, $\alpha_1 = 2.5$, $\alpha_2 = 0.099$. In this case, α_1 is multiplied to the value read from the roughness texture.

generated materials parameters. Then, for each image couple, we have calculated the mean $\Delta E_0^* 0$ value by averaging the difference values on the single pixels. To focus only on the colors generated on the sphere, we have set the background of the scene as full transparent. If this is acceptable for a numeric measure, the choice of the background has been proven to be a delicate choice for color perception tests with human subjects (Rizzi et al., 2013).

From the plots in Figures 3 and 4, it can be noticed how the applied parameters are able to produce at least half of the materials with evident perceptual differences with the target one, and an adequate number of materials with more subtle differences. This is in line with the intended behaviour: the generation of new materials with a strong bond with the reference, but presenting a range of perceptual differences ranging from low to moderate. Moreover, comparing the two considered fitness functions, we can conclude that the Chebyshev distance introduces more differences in the generated individuals than the Euclidean distance. In Figures 3 and 4 we show also the resulting mean $\Delta E_0^* 0$ values repeating the whole experimental setup with increasing values of population size and number of generations. The data confirms that the generation using the Euclidean distance as fitness function converges faster to materials perceptually indistinguishable from the original. In any case, with a population larger that 100 individuals, and with a number of generations higher than 10, the evolu-

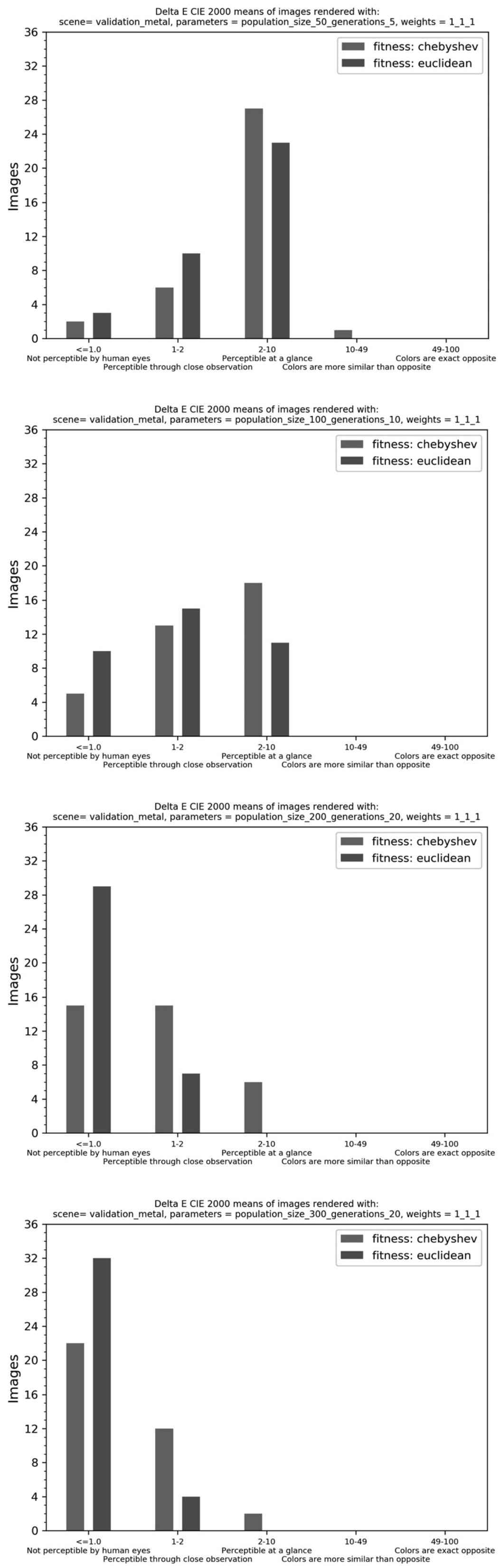

Figure 3: Average ΔE^*_{00} using the two distances, between the test scene rendered with **validation metal**, and the images with the generated materials, increasing the population size and the number of generations.

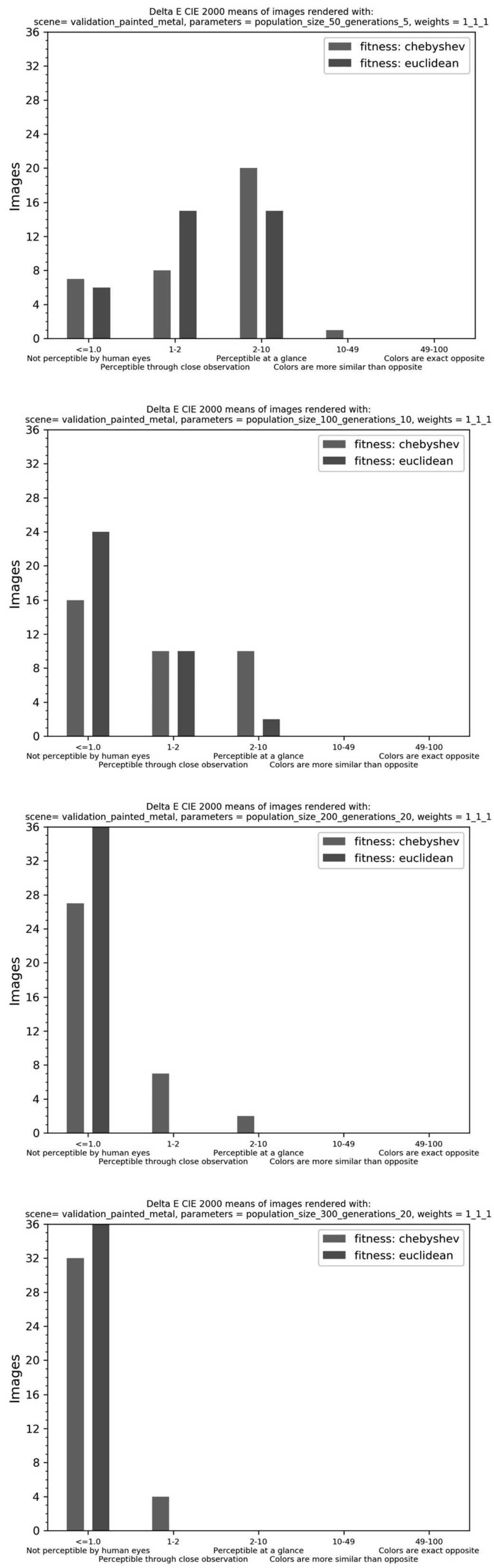

Figure 4: Average ΔE^*_{00} using the two distances, between the test scene rendered with **validation painted metal**, and the images with the generated materials, increasing the population size and the number of generations.

Figure 5: A subset of the generated materials (population 50, generations 5), with target material **validation metal** and Chebyshev distance.

Figure 6: A subset of the generated materials (population 50, generations 5), with target material **validation metal** and Euclidean distance.

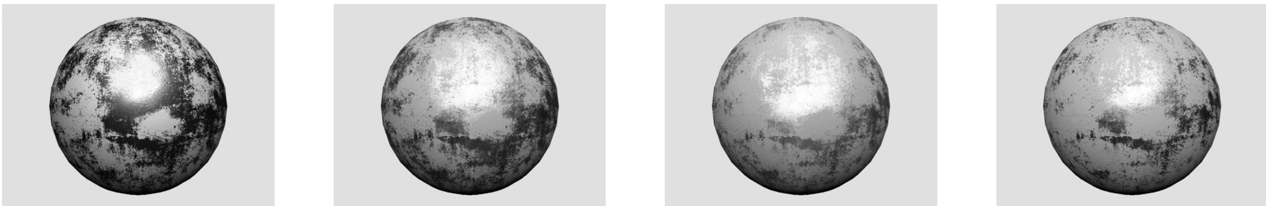

Figure 7: A subset of the generated materials (population 50, generations 5), with target material **validation painted metal** and Chebyshev distance.

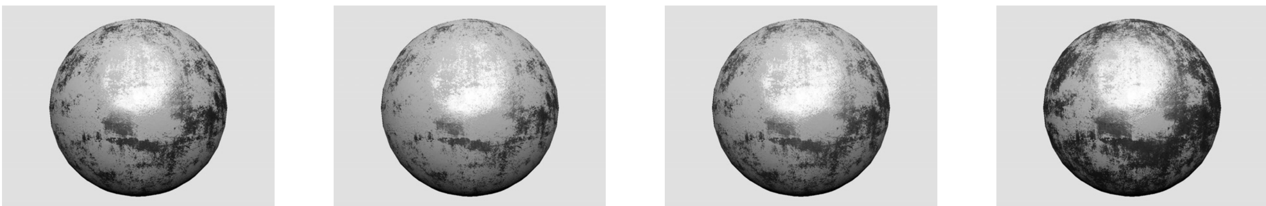

Figure 8: A subset of the generated materials (population 50, generations 5), with target material **validation painted metal** and Euclidean distance.

tionary approach generates materials mainly identical to the reference, which is not the desired result.

CONCLUSION AND FUTURE WORK

In this paper, we have shown how a GA can be applied efficiently to evolve the parameters of a BSDF, in order to generate different versions of a target material presenting a moderate amount of perceptual differences.

The presented approach is computationally efficient. The GA and the BSDF computations are independent: the first is intended to be executed offline on the CPU, during the loading of a scene, or following the procedural generation of several istances of a reference model, while the second is executed as a shader on the GPU, during the actual real-time rendering, using the results of the GA computation as parameters. On a Intel Quad-Core i7-6700HQ machine with 8 GB DDR4 RAM, equipped with a NVIDIA GeForce GTX 970M with 3GB GDDR5 VRAM, the GA took around 30 ms to generate a material using a population of 100 individuals and 10 generations, while the performances of the shader implementing the Belcour's method are in line with the one stated in the original paper (Belcour, 2018).

The choice of the fitness function leads to different behaviours in the overall GA computation. However, this

can be seen as a control parameter for the final user, who can decide about the generation of materials more or less close to the target one, by selecting a different distance function.
In future research, we will evaluate the effect of other distance functions, considering also other target materials with the presence of a participating media. Moreover, we will test the effect of different weights applied to the chromosomes in the fitness function, in order to allow the final user to select the more relevant features to consider in the selection of the best individuals.

REFERENCES

Akenine-Möller T.; Haines E.; Hoffman N.; Pesce A.; Iwanicki M.; and Hillaire S., 2018. *Real-Time Rendering 4th Edition*. A K Peters/CRC Press. ISBN 978-1-13862-700-0.

Belcour L., 2018. *Efficient Rendering of Layered Materials using an Atomic Decomposition with Statistical Operators. ACM Trans on Graph*, 37, no. 4, 1.

Brady A.; Lawrence J.; Peers P.; and Weimer W., 2014. *genBRDF: Discovering New Analytic BRDFs with Genetic Programming. ACM Trans Graph*, 33, no. 4, 114:1–114:11.

Ebert D.S.; Musgrave F.K.; Peachey D.; Perlin K.; and Worley S., 2003. *Texturing & Modeling: a procedural approach*. Morgan Kaufmann.

Guarneri A.; Maggiorini D.; Ripamonti L.A.; and Trubian M., 2013. *GOLEM: Generator Of Life Embedded into MMOs*. In *Proceedings of the Twelfth European Conference on the Synthesis and Simulation of Living Systems: Advances in Artificial Life, ECAL*. 585–592.

Masia B.; Munoz A.; Tolosa A.; Anson O.; Lopez-Moreno J.; Jimenez J.; and Gutierrez D., 2009. *Genetic algorithms for estimation of reflectance parameters*. In *Proceedings of the 2009 Spring Conference on Computer Graphics (SCCG09)*.

Mazza C.; Ripamonti L.A.; Maggiorini D.; and Gadia D., 2017. *FUN PLEdGE 2.0: A FUNny Platformers LEvels GEnerator (Rhythm Based)*. In *Proceedings of the 12th Biannual Conference on Italian SIGCHI Chapter (CHItaly '17)*. ACM, 22:1–22:9.

Mokrzycki W. and Tatol M., 2011. *Colour difference Delta E - A survey. Machine Graphics and Vision*, 20, no. 4, 383–411.

Mourato F.; dos Santos M.P.; and Birra F., 2011. *Automatic Level Generation for Platform Videogames Using Genetic Algorithms*. In *Proceedings of the 8th International Conference on Advances in Computer Entertainment Technology*. ACM, ACE '11, 8:1–8:8.

Norton D.; Ripamonti L.A.; Ornaghi M.; Gadia D.; and Maggiorini D., 2017. *Monsters of Darwin: A Strategic Game Based on Artificial Intelligence and Genetic Algorithms*. In *Proceedings of the 1st Workshop on Games-Human Interaction (GHItaly 2017)*. No. 1956 in CEUR Workshop Proceedings.

Pharr M.; Wenzel J.; and Humphreys G., 2016. *Physically Based Rendering: From Theory to Implementation 3th Edition*. Morgan Kaufmann.

Ripamonti L.A.; Gratani S.; Maggiorini D.; Gadia D.; and Bujari A., 2017a. *Believable group behaviours for NPCs in FPS games*. In *Proceedings of IEEE Digital Entertainment, Networked Virtual Environments, and Creative Technology Workshop (DENVECT 2017)*.

Ripamonti L.A.; Mannalà M.; Gadia D.; and Maggiorini D., 2017b. *Procedural content generation for platformers: designing and testing FUN PLEdGE. Multimedia Tools and Applications*, 76, no. 4, 5001–5050.

Rizzi A.; Bonanomi C.; Gadia D.; and Riopi G., 2013. *YACCD2: yet another color constancy database updated*. In *Color Imaging XVIII: Displaying, Processing, Hardcopy, and Applications, Proc. of IS&T/SPIE's Symposium on Electronic Imaging*. SPIE, 86520A–86520A–10.

Scalabrin M.; Ripamonti L.A.; Maggiorini D.; and Gadia D., 2016. *Stereoscopy-based procedural generation of virtual environments*. In *Proceedings of IS&T's Stereoscopic Displays and Applications XXVII (28th Symposium on Electronic Imaging : Science and Technology)*. 5, 1–7.

Schwarz M. and Müller P., 2015. *Advanced Procedural Modeling of Architecture. ACM Trans Graph*, 34, no. 4, 107:1–107:12.

Shaker N.; Togelius J.; and Nelson M.J., 2016. *Procedural Content Generation in Games: A Textbook and an Overview of Current Research*. Springer.

Sitthi-Amorn P.; Modly N.; Weimer W.; and Lawrence J., 2011. *Genetic Programming for Shader Simplification. ACM Trans Graph*, 30, no. 6, 152:1–152:12.

Walter B.; Marschner S.R.; Li H.; and Torrance K.E., 2007. *Microfacet Models for Refraction Through Rough Surfaces*. In *Proceedings of the 18th Eurographics Conference on Rendering Techniques*. EGSR'07, 195–206.

Yannakakis G.N. and Hallam J., 2009. *Real-Time Game Adaptation for Optimizing Player Satisfaction. IEEE Trans Comput Intellig and AI in Games*, 1, no. 2, 121–133.

N-Layered Feudal Network in an RTS Game Environment

Benjamin Bugeja
Department of Computer Science
University of Malta, Malta
benjamin.bugeja.13@um.edu.mt

Jean-Paul Ebejer
Centre for Molecular Medicine and Biobanking
University of Malta, Malta
jean.p.ebejer@um.edu.mt

Sandro Spina
Department of Computer Science
University of Malta, Malta
sandro.spina@um.edu.mt

KEYWORDS

Neural networks, reinforcement learning, game AI, real-time strategy, hierarchical.

ABSTRACT

In an RTS, players act simultaneously in adversarial conditions. The agent must gather and remember information about the map and the opponent, while making decisions with long term consequences. FeUdal networks (FuN) tackle the domain using hierarchical reinforcement learning. A Manager sets goals for a Worker, which is intrinsically rewarded for accomplishing the goal. To explore the effectiveness of FuN, we propose an N-Layered FuN (NL-FuN), which generalises the Manager to fit an arbitrary number of tiers (3 tiers in this case). This requires adapting FuN to the StarCraft 2 domain. In addition, Atari-Net, FullyConv, and FullyConvLSTM are recreated and used as a baseline. Agents are implemented using PySC2 and TensorFlow. The scenarios used for training are: MoveToBeacon, DefeatSingleZealot, and BuildMarines. BuildMarines is the most complex of the three, having sparse rewards and requiring long-term planning. NL-FuN performs poorly in MoveToBeacon and similarly to FullyConv and FullyConvLSTM in DefeatSingleZealot but obtains a higher maximum reward than the baseline agents trained by DeepMind in BuildMarines in less time steps (540,000 vs 600,000,000).

INTRODUCTION

Reinforcement learning (RL) consists of learning systems with their own goals, that depend on an environment, and are capable of adapting their behaviour to maximise a reward signal. RL problems are traditionally framed as a finite Markov Decision Process (MDP). The *agent* represents the decision maker, such as a chess player. The *environment* is everything the agent interacts with, such as the board and the pieces. Interaction occurs at discrete time steps. The environment provides the agent with a representation of its *state*, such as the positions of the pieces on the board. The agent performs actions, such as moving a piece, using a *policy*, which maps states to action probabilities. The environment returns a reward for the action taken, such as the cost of captured pieces, and the next state. Negative rewards penalise agent, such as losing pieces. It should be noted that the reward is not *how* the goal should be reached, but rather *what* the goal is. Otherwise, the agent might find ways to obtain rewards for its sub-goals and still fail at the primary goal. Using the chess example, rewarding the agent for capturing pieces might result in prioritizing captures over checkmating. A more accurate reward is whether the player has won, lost, or drawn. The agent's goal is to maximise the total cumulative reward, or the expected return. The agent can estimate the expected return, given a state or a state-action pair, using a *value* function (Lanctot et al., 2017, Sutton and Barto, 2017).

Rewards obtained across long intervals—sparse rewards—make planning harder, as the agent needs to search further for the next reward. Actions can be abstracted to encapsulate a sequence of lower level decisions, reducing the number of actions required to get to the next reward, thus simplifying the search space. FeUdal Networks (FuNs) can model this by taking advantage of hierarchical RL to represent higher level actions (Vezhnevets et al., 2017). FuNs have been applied to Atari games, but real-time strategy (RTS) games represents a deeper challenge, with more states, more actions, and potentially sparser rewards. Additional hierarchical tiers to model more abstract goals are suggested by A. S. Vezhnevets et al.

Aims and Objectives

The aim of this paper is to adapt and extend Feudal Networks to fit StarCraft 2's (SC2) environment. For evaluation purposes, three baseline agents are replicated: Atari-net (Mnih et al., 2015), FullyConv, and FullyConvLSTM (Vinyals et al., 2017). The original baselines were trained on resources which were not feasible to obtain for this paper, therefore replication using available hardware makes for a fairer comparison to other locally trained agents. The FeUdal Network presented by A. S. Vezhnevets et al. (2017), originally used on a suite of Atari games, is adapted to SC2. This network is then extended to use more hierarchical tiers. Hardware limitations force a number of concessions in feature complexity, emphasising the need for optimisation. A benchmarking scenario is developed using SC2's built-in map editor. This helps to provide insight into how reward shaping affects the learning process.

BACKGROUND

The RTS genre is of interest to AI research (Churchill, 2017, Ontanon et al., 2013, Vinyals et al., 2017) due to the complexity of the RTS domain. For any game, given a branching factor b and depth d, a game's complexity is given by b^d (Synnaeve, 2012). In games such as Go, $b \approx 30$ to 300 and $d \approx 150$ to 200. For StarCraft (SC), the game's complexity is even higher, where $b \in [10^{50}, 10^{200}]$ and $d \approx 36000$ (Ontanon et al., 2013). The RTS environment provides additional challenges over traditional boardgames such as Go. A real-time environment places a restriction on computation time as the players act simultaneously. Secondly, it is a game of imperfect information, since the agent observes the game through a local camera which can be moved over a global space. 'Fog-of-war' further compounds imperfect information, as unvisited regions of the map are hidden to the agent, requiring active scouting in order to gather information about their opponent. Many RL algorithms also require a state or action value estimation. The complexity of SC makes this a non-trivial task (Erickson and Buro, 2014). Finally, each game lasts for thousands of frames. Players must be able to plan ahead, making decisions which may only benefit them deep into the game.

The StarCraft 2 Learning Environment

The SC2 API provided by the SC2 Learning Environment (SC2LE) exposes access to SC2's interface. The PySC2 API wraps the SC2 API and restricts information to a set of observations. These observations consist of spatial feature layers and structured features. In addition, the action space is simplified. Each action is made up of a function call and its arguments, and a single action may generalize more specific functions. For example, *cancel building* and *cancel unit* are grouped under the same function, as their availability is unique given the context. A more detailed description of the exposed features can be found on the github page for PySC2.
Using the PySC2 API, DeepMind created three RL baseline agents: Atari-Net, FullyConv, and FullyConvLSTM. Atari-Net is adapted from the agent of the same name used in the Arcade Learning Environment (Mnih et al., 2013, 2015). The FullyConv baseline agent uses padded convolutions with stride 1 to retain spatial information. The FullyConvLSTM agent adds a convolutional LSTM after the state representation is constructed (Vinyals et al., 2017).

FeUdal Networks

A. S. Vezhnevets et al. (2017) propose FeUdal Networks (FuN), inspired by Feudal Reinforcement Learning originally proposed by P. Dayan and G. E. Hinton (1993). FuN consists of a high-level Manager and a low-level Worker. The model is constructed from the following equations:

$$z_t = f^{percept}(x_t) \tag{1}$$

$$s_t = f^{Mspace}(z_t) \tag{2}$$

$$h_t^M, \hat{g}_t = f^{Mrnn}(s_t, h_{t-1}^M); g_t = \hat{g}_t / \|\hat{g}_t\|; \tag{3}$$

$$w_t = \phi\left(\sum_{i=t-c}^{t} g_i\right) \tag{4}$$

$$h^W, U_t = f^{Wrnn}(z_t, h_{t-1}^W) \tag{5}$$

$$\pi_t = \text{SoftMax}(U_t w_t) \tag{6}$$

Here, z_t corresponds to the processed observations and counts as the perception space of the Worker. The Manager perceives the environment through s_t, a function of z_t. h^M and h^W represent the internal states of the Manager and the Worker respectively. These are returned by the corresponding RNN. ϕ is a linear transform that maps the Manager's goal, g_t, into an embedding vector w_t. This embedding vector is combined with the Worker's output, U_t, with a product. This produces the policy π, the probability distribution of primitive actions.

Goal Embedding
The goal, g, affects the Worker's actions by being embedded into a lower dimensional space, R^k, where k is much smaller than the Manager's space, d. To do this, the Worker must first produce an embedding vector, where each action is represented by a row in the matrix $U \in R^{|a| \times k}$. The Manager's last c goals are summed together and then embedded into vector $w \in R^k$ by applying the linear projection ϕ. No biases are used, and gradients are obtained from the Worker's actions. A matrix-vector product combines the embedding matrix U with the goal embedding w.

Learning
The agent's goal is to maximise the discounted return, $R_t = \sum_{k=0}^{\infty} \gamma^k r_{t+k+1}$, where γ is the discount factor and r is the reward, by optimising a stochastic action-selection policy, π. The discount factor is a value ranging between $(0, 1]$, used to scale the importance of each reward. While the system is trainable end-to-end, the goals of the Manager would simply become another internal variable. A. S. Vezhnevets et al. propose training the Manager and the Worker independently. The Manager predicts advantageous transitions in its state space, and the Worker is intrinsically rewarded for following these directions. The Manager's update rule is as follows:

$$\nabla g_t = A_t^M \nabla_\theta d_{cos}(s_{t+c} - s_t, g_t(\theta)), \tag{7}$$

where $A_t^M = R_t - V_t^M(x_t, \theta)$ is the Manager's advantage function computed using the value function estimate, $V_t^M(x_t, \theta)$. $d_{cos}(\alpha, \beta) = \alpha^T \beta / (|\alpha||\beta|)$ is the cosine similarity function between two vectors, and c is the horizon, defining the temporal resolution of the Manager by specifying the number of time steps which must be taken into account when choosing a new goal. A higher c means the Manager has a lower resolution, that is, a longer horizon, over which to obtain the next goal.
On the other hand, the Worker's intrinsic reward is as follows:

$$r_t^I = 1/c \sum_{i=1}^{c} d_{cos}(s_t - s_{t-i}, g_{t-i}), \tag{8}$$

The superscript I denotes that this is an *intrinsic* reward. The difference between the two states, $s_t - s_{t-i}$, represents the Worker's trajectory, while g_{t-i} represents the Manager's suggested trajectory. In addition, the Worker also retains the environment's reward as well. As a result, the Worker is trained to maximise the sum of the environmental reward and the weighted intrinsic reward: $R_t + \alpha R_t^I$ where α is used to weigh the intrinsic reward. The policy π is then trained to maximise the reward using A3C (Mnih et al., 2016). The Worker and Manager are allowed to have different discount factors. This can be used to force the Worker to focus on more immediate results while the Manager sets long-term goals.

Transition Policy Gradients

A. S. Vezhnevets et al. (2017) propose an update rule for the Manager in *transition policy gradients*. Given a high-level policy $o_t = \mu(s_t, \theta)$ that selects among sub-policies, assume that the sub-policies last a fixed amount of c time steps. Each sub-policy corresponds to a transition distribution, $p(s_{t+c}|s_t, o_t)$. This describes the distribution of states, s_{t+c}, found at the end of each sub-policy, given a start state, s_t, and the sub-policy selected by o_t. A transition policy π^{TP} can be obtained by composing the high-level policy with the transition distribution:

$$\pi^{TP}(s_{t+c}|s_t) = p(s_{t+c}|s_t, \mu(s_t, \theta)) \tag{9}$$

This describes the distribution over end states, s_{t+c}, given the start states, s_t. The original MDP is isomorphic to the new MDP with policy π^{TP}, and $s_{t+c} = \pi^{TP}(s_t)$, such that the state always transitions to the end state picked by the transition policy. As a result, the policy gradient theorem can also be applied to the transition policy π^{TP}, such that:

$$\nabla_\theta \pi_t^{TP} = \mathbb{E}[(R_t - V(s_t))\nabla_\theta \log p(s_{t+c}|s_t, \mu(s_t, \theta))] \tag{10}$$

FuN assumes that the direction in state-space, $s_{t+c} - s_t$ follows a von Mises-Fisher distribution. such that:

$$p(s_{t+c}|s_t, o_t) \propto e^{d_{cos}(s_{t+c} - s_t, g_t)} \tag{11}$$

where g_t represents the mean direction of the von Mises-Fisher distribution. A von Mises-Fisher distribution is a probability distribution on a (p-1)-dimensional sphere. This explains the form of Eq. 7 and Eq. 8. Since the Worker is learning to achieve the Manager's direction, it should follow a distribution around the given direction, hence reinforcing the assumption for transition policy gradients (Vezhnevets et al., 2017).

RELATED WORK

Q-Learning is a staple of RL. It is an off-policy approach, where the target policy is learned independently from the behaviour policy used to explore the action space. Q-Learning was made feasible for use in more complex environments with the proposal of Deep-Q-Networks (DQN) (Mnih et al., 2013), which integrated neural networks. This lead to a number of DQN extensions which improve on various parts of vanilla DQN (Li, 2017, Sutton and Barto, 2017). This can be used as an alternative to the on-policy approach used in this study.

Multi-agent systems present a distributed way of tackling complex scenarios, most notably, resource allocation tasks similar to those found in RTS games. Through this divide-and-conquer approach, responsibility is distributed among the agents in order to achieve a common goal (Busoniu et al., 2008). A multi-agent system may still contain hierarchical properties either in the relationship between the agents or within the structure of the agents themselves.

The Options framework is another approach to hierarchical RL. Sutton et al. generalise previous contributions to establish a simpler framework with less deviation from the existing RL architecture by exploring the relationship between Semi-MDPs (SMDPs) and MDPs (Sutton et al., 1999). P.L. Bacon et al. extend this to produce the Option-Critic architecture, a system capable of autonomously learning sub-goals as opposed to having them explicitly provided (Bacon et al., 2017).

N-LAYERED FEUDAL NETWORK

The N-Layered FeUdal Network (NL-FuN) extends FeUdal Networks by implementing a Super Manager, a Manager, and a Worker. Additional Managers can be added to the hierarchy to compute goals over longer periods of time. This is useful in environments where rewards are so sparsely located that a single tiered architecture with long rollouts would gather too much noise and hinder learning.

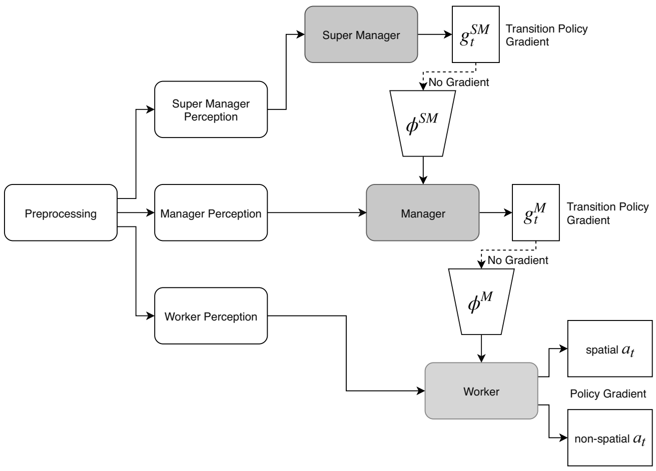

Fig. 1. 3-Layered FeUdal Network

Processing a rollout at multiple levels of temporal resolution allows each tier to sample a direction in its own latent space with respect to a super manager's goal (or the environment's reward in the case of the Super Manager), but defers achieving that goal to a sub-manager operating at a more granular level. Representing goals as directions being shifted to optimize the reward is preferred over representation as an arbitrary location as it is more feasible for the Worker to reliably cause directional shifts in the latent space rather than take us to arbitrary new locations (Vezhnevets et al., 2017).

Adapting FeUdal Networks to StarCraft II

The Arcade Learning Environment takes one input at a time (an action from a predetermined set), while PySC2 requires two inputs. This is addressed by splitting the action matrix, U, into two separate matrices—one for spatial actions, U^s, and one for non-spatial actions, U^{ns}. Spatial actions consist of picking coordinates on the screen. Non-spatial actions consist of picking a particular action such as 'Stop' or 'Attack'. If the chosen non-spatial action does not take coordinates as arguments, the spatial action is ignored. The spatial action matrix, U^s, is obtained by passing the ConvLSTM output through two 2D convolutional layers with a 1×1 kernel and k filters, where k is the dimensionality of the embedding vector, w. Similarly, U^{ns} is obtained by passing the convolutional LSTM output through two dense layers with a ReLU function, having $k \times |a|$ units, where $|a|$ is the total number of non-spatial actions. Finally, the embedding vector, w, is obtained from the Manager and both U^s and U^{ns} are multiplied with the vector separately. The policy is obtained with a softmax over the resulting tensors. Given that the Manager has its own state representation and only outputs goal directions, no changes were made to its structure.

To construct the shared perception, z, the preprocessed features representing the state are passed through a ConvLSTM, while a state preprocessed by a fully-connected layer is passed to the Manager. This follows the same procedure used in FullyConvLSTM to obtain both a state representation and a value function for the Worker, while following the FeUdal Network's setup of using the Worker's state to compute the Manager's internal state representation.

Network Blocks

Before obtaining the latent state space representation for the Manager and Super Manager, a higher level state representation is obtained as per the FuN schematics. To obtain the shared perception, the preprocessed features, x_t, are passed through a fully connected ReLU layer with d outputs, where d represents the dimensionality of the shared state space.

The **Super Manager** feeds the latent state space to a dilated LSTM, which outputs, $\hat{g}$. The final output, g_t^{SM}, is obtained by taking $\frac{\hat{g}}{||\hat{g}||}$. The linear transform ϕ is a tensor of size (d, k), where d is the dimensionality of the goal, g_t^{SM}, and k is the dimensionality of the embedding vector, w. The embedding vector itself is obtained by multiplying the summation of the past c goals with ϕ.

The **Manager** has the output of the dilated LSTM multiplied by the embedding vector provided by the Super Manager. The Manager also creates its own embedding vector in the same way that the Super Manager does.

The **Worker** model is similar to the FullyConvLSTM baseline agent. As input, the ConvLSTM takes the state and the step size. Convolution is performed using a 1×1 kernel and 1 filter reducing the input dimensionality. To obtain the spatial action, the output of the ConvLSTM is passed through another 1×1 convolutional layer with k filters. This matches the dimensionality of the embedding vector. To obtain the non-spatial action, the output of the ConvLSTM is passed through a fully connected layer with $|a| \times k$ units, where $|a|$ is the number of possible non-spatial actions. This allows the output to be reshaped and multiplied with the embedding vector, similarly to the spatial action. Finally, the value function also uses a fully connected layer, but with the dimensionality of the state as the number of outputs. All fully connected layers use a ReLU activation.

The **value** is computed by passing the state representation through a fully connected layer with 1 output.

Learning

Each tier is trained independently, where managers predict advantageous transitions in their state space, and reward their subordinates for following these directions. By rewarding the subordinate for fulfilling its manager's goal, a successful subordinate would shift the resulting trajectories in advantageous directions as well. This is formalised in Equation (7).

Managers with a Super Manager also have their own intrinsic reward. This changes their advantage function to be equivalent to the Worker's advantage function:

$$A_t^M = R_t + \alpha R_t^I - V_t^M(x_t, \theta) \tag{12}$$

where α is a hyper-parameter weighting the intrinsic reward. Note that the reward obtained from the environment is still included, as specified in the Manager-Worker FuN specification (Vezhnevets et al., 2017). The intrinsic reward is obtained from processing the Super Manager's state differences and goals, similarly to how the Worker's intrinsic reward is obtained from its own Manager, as in Equation (8). The Manager is still trained using a transition policy gradient, as its output, i.e. the goals, are still assumed to be a trajectory. Random goals are emitted with probability ϵ at each step for each tier to encourage exploration.

EVALUATION

The agents are trained on MoveToBeacon, DefeatSingleZealot, and BuildMarines. In all scenarios, fog of war is disabled and no camera movement is required. Some features are excluded in order to reduce the computational complexity. Reducing the number of features also reduces the amount of memory required for each agent, making it possible to run more agents in parallel. The chosen features must also be relevant to the agent's goal.

The **MoveToBeacon** scenario consists of a single Marine and a single Beacon. The agent earns rewards by moving the Marine toward the Beacon. Whenever the Marine reaches the Beacon, the Beacon is moved to a random location that is at least 5 units away from the Marine. The scenario lasts for 120 seconds and the episode ends when the timer expires. The agent must be able to select the Marine and move it on the Beacon as efficiently as possible. Features were selected as follows:

- **Screen features:** *player_id*, *player_relative*, *unit_type*, *selected*, *unit_density*
- **Minimap features:** *player_id*, *selected*
- **Non spatial features:** *available_actions*, *single_select*

The **DefeatSingleZealot** scenario consists of a single Stalker and a single Zealot. A Stalker is a ranged unit with fast movement speed. A Zealot is a melee unit with slower movement speed than the Stalker, but with more hit points. In addition, the Zealot's melee attack is stronger than the Stalker's ranged attack in a straight-up fight. Therefore, the agent's objective is to manoeuvre the Stalker in such a way that it uses its speed and ranged attacks to deal damage from a distance without being hit by the Zealot's strong attack. Whenever the Stalker deals damage to the Zealot, the agent is given a reward of 1. Whenever the Stalker kills the Zealot, the agent is given a reward of 5. If the Zealot damages or kills the Stalker, the agent is penalised with -1 and -5 respectively. Whenever a unit dies, the remaining unit is teleported to one side of the map while the dead unit is respawned on the other side. The scenario lasts for 120 seconds and ends when the timer expires. Features were selected as follows:

- **Screen features:** *player_id*, *player_relative*, *unit_type*, *selected*, *unit_density* *unit_hit_points_ratio*, *unit_shields_ratio*
- **Minimap features:** *player_id*, *selected*
- **Non spatial features:** *available_actions*, *single_select*

The **BuildMarines** scenario starts the agent off with 8 Mineral Fields (resources), a Command Center, and 12 SCVs beside the Command Center. SCVs, also known as workers, are units which can be used to gather minerals and construct buildings. The Command Center is a structure which can produce more SCVs and serves as the drop-off point for gathered resources. The agent must earn rewards by building as many Marines as possible using these tools. In order to build Marines, the agent must construct a Barracks to produce the Marines, which in turn requires a Supply Depot. Units and structures cost Minerals, therefore the agent must also learn to gather Minerals from Mineral Fields using SCVs. Marines also cost Supply, which is increased by building more Supply Depots. The final reward is equal to the number of Marines built. The scenario ends after 900 seconds. Actions which are not required to construct additional Marines are excluded. Features were selected as follows:

- **Screen features:** *player_id*, *player_relative*, *unit_type*, *selected*, *unit_density*
- **Minimap features:** *player_id*, *selected*
- **Non spatial features:** *player*, *available_actions*, *single_select*, *build_queue*

Preprocessing

Preprocessing is the same for all agents. Feature layers containing categorical values are converted to a one-hot encoding in the channel dimension followed by a 1×1 convolution. This is equivalent to mapping categorical values into a continuous space. The 1×1 convolution maintains the low dimensionality in the preprocessing while still extracting meaningful statistics. Numerical features are scaled using a logarithmic transformation for numerical stability. This matches the preprocessing done on the DeepMind baseline agents (Vinyals et al., 2017). Scaling numerical features is required as some features, such as hit points or number of resources, might attain large values which hinder the learning process.

The screen and minimap resolution were set to 32×32, as opposed to 64×64. This reduction is done in favour of training speed and lower memory usage over accuracy. Empirical observations found this value to be enough to visually distinguish individual units.

Baseline Agents

The reconstructed agents were trained under an ϵ-greedy policy. The initial value for ϵ was set to 0.5 and linearly annealed down to 0.1 throughout the run. This is to obtain better short-term results within the short training period. The maximum number of steps per rollout, K, is set to 40, as specified in the DeepMind paper (Vinyals et al., 2017). The learning rate is randomly sampled from a log-uniform distribution in the range of $[10^{-4}, 10^{-3}]$ and annealed to half the original value, then kept constant for the rest of the run. The entropy regularization weight is sampled from a log-uniform distribution in the range of $[10^{-4}, 10^{-2}]$. The weight of the value loss is sampled from a uniform distribution in the range of $(0, 1]$. The discount rate is set to 0.99. This discount rate follows the rate used by A. S. Vezhnevets et al. (2017).

FeUdal Network

For the MoveToBeacon and DefeatSingleZealot scenarios, K is set to 240, which is the length of a single episode. For the BuildMarines scenario, K is set to 600. The horizon, c_m, of the Manager is set to 40 in all scenarios, as per the specifications of A. S. Vezhnevets et al. (2017) (Vezhnevets et al., 2017). The learning rate is sampled from a log-uniform distribution in the range of $[10^{-6}, 10^{-4}]$ and annealed to half the original value by half the training run, then kept constant for the remaining episodes. As with the baseline agents, the entropy regularization weight is sampled from a log-uniform distribution in the range of $[10^{-4}, 10^{-2}]$. The weights of the value loss for the Manager and the Worker are sampled from a uniform distribution in the range of $(0, 1]$. The discount rates of the Manager and Worker are set to 0.99, and 0.95 respectively. This allows the Worker to focus on more short term goals while the managers prioritize more long term goals. The goal dimension for the Manager is randomly chosen from the set $\{64, 128, 256\}$. The embedding vector's dimensionality is randomly chosen from the set $\{8, 12, 16\}$. The complexity of the goal and the embedding vector may impact the time it takes to converge. While a higher dimensionality allows for more complex representation and possibly better results in the long term, a simpler representation uses less computation and memory, and may also converge to a better optimum in the short term. In this case, 'short term' and 'long term' aren't clearly defined and cannot be guaranteed to fall within a set number of steps. As such, the dimensionality is randomised in order to cast a wider net on all possible cases.

N-Layered FeUdal Network

The hyper-parameters for the Worker and Manager are the same as those of the FeUdal Network. The horizon, c_{sm}, of the Super Manager is set to 120 in the MoveToBeacon and DefeatSingleZealot scenarios, and set to 200 in the BuildMarines scenario. The weight of the value loss for the Super Manager is sampled from a uniform distribution in the range of $(0, 1]$. The discount rate of the Super Manager is set to 0.999. This follows the pattern set by the Worker and the Manager in providing the Super Manager with more foresight than its subordinate. The state and goal dimensions for the Super Manager are randomly chosen from the set $\{64, 128, 256\}$. The embedding vector's dimensionality is the same for all levels of the hierarchy, and is randomly chosen from the set $\{8, 12, 16\}$.

Optimisation

Training was performed using A3C. A parameter server maintains weights for a global network while worker networks collect samples from the environment and calculate losses. Each worker returns a rollout after either K forward passes or after a terminal signal is received. The global network is trained asynchronously using gradients obtained from each worker. The workers are then updated with new weights from the global network. Optimization consisted of running 8 asynchronous workers using a shared RMSProp (Hinton, 2016). Each new tier requires a rollout that operates at the new horizon. Each rollout feeds inputs to the respective tier's training operation at the required resolution. This had the Manager attempting to train from rollouts with different batch lengths. The issue was resolved by padding the shorter batch with either the last inputs from the previous batch, or with the same adjacent values given no previous batch. Truncating was considered and implemented, but padding was chosen in favour of retaining information.

Hardware

Agents are trained on two separate machines. One has an i7-6700K, 16GB of DDR4 RAM, and two GTX 1070 GPUs with 8GB of VRAM each. The other is a virtual machine set up on the Google Cloud Platform, with 8 vCPU's, 40GB of RAM, and 1 K80 GPU with 11GB of VRAM.

Evaluation Criteria

Each training scenario rewards agents with a score depending on their performance. The agent with the best maximum and mean rewards in a particular scenario is deemed to be the best performing agent. The mean rewards are taken to be the mean score recorded at the end of all episodes during training. The maximum rewards are taken to be the highest achieved score throughout training. The maximum and mean rewards are also used by *DeepMind* to evaluate the baselines, therefore the same metrics are used in order to make meaningful performance comparisons.
The time taken to accomplish each run is not taken into consideration. Multiple runs were done in parallel on two machines with different hardware specifications. As a result, the time taken for each run cannot be directly compared and used for evaluation. In addition, some training runs on the virtual machine were done on CPUs instead due to GPU memory restrictions, this also affects the amount of time a single run takes.

Results

Each agent was trained for 240,000 steps in the MoveToBeacon and DefeatSingleZealot scenarios, and for 540,000 steps in the BuildMarines scenario. Hyper-parameters were randomly sampled over 10 runs per agent per scenario. Random sampling was chosen over other methods due to its simplicity in terms of implementation as well as its effectiveness compared to grid search and manual tuning in high dimensional hyper-parameter spaces (Bergstra and Bengio, 2012).
Table I contains the best mean and maximum rewards obtained by the agents.

TABLE I
AGGREGATED RESULTS FOR MAX AND BEST MEAN

Agent	Metric	MtB	DSZ	BM
Human Player	Mean	28	70	138
	Max	28	71	142
Atari-Net	Best Mean	<1	**-61**	5
	Max	5	0	39
FullyConv	Best Mean	**16**	-78	2
	Max	**27**	-45	16
FullyConvLSTM	Best Mean	10	-85	2
	Max	25	-48	24
FuN	Best Mean	5	-72	2
	Max	14	0	22
NL-FuN	Best Mean	<1	-81	**8**
	Max	6	**2**	**59**

Baselines
The FullyConv and FullyConvLSTM agents performed as expected on MoveToBeacon. The Atari-Net agent performed worse than expected, being unable to obtain a mean reward of more than 1 reward over any of its runs. As per Figure 2, the baseline agents, with the exception of Atari-Net, excel at this task, almost immediately obtaining a score that is superior to the hierarchical agents. DefeatSingleZealot offered a greater challenge than expected. All agents learned rudimentary evasive movements, but the most successful agents were those which avoided the enemy entirely. The scenario had dense rewards but required precise timing. While the Atari-Net agent outperformed all other agents in this scenario, no agent managed to obtain a positive mean score over any of its runs. The Atari-Net agent's performance is in line with the

original baseline agent's performance on other similar combat scenarios tested by DeepMind (Vinyals et al., 2017), where Atari-Net also performed relatively well. BuildMarines was expected to offer the hardest challenge to the agents. As seen in Figure 4, Atari-Net diverges from its initial policy, leading to worse rewards over time. FullyConv and FullyConvLSTM performed as expected, being unable to learn a consistent policy to build Marines.

Hierarchical Agents
On the MoveToBeacon scenario, the hierarchical agents performed worse than expected. FuN performed better than the Atari-Net agent. Due to the dense reward found in DefeatSingleZealot, both hierarchical agents were expected to perform similarly to the baseline agents. NL-FuN obtained the highest maximum reward of any agent, but failed to learn a consistent policy. FuN performed slightly better, obtaining the second best mean reward of all agents and almost matching the performance of Atari-Net. The hierarchical agents were expected to perform best on BuildMarines, as their long-term goal trajectories are suited for scenarios with sparse rewards. While the FuN agent failed to outperform FullyConvLSTM and Atari-Net, NL-FuN outperformed all agents. Figure 4 shows the best mean rewards for each agent in this scenario. NL-FuN has multiple reward spikes throughout its run. This is a sign of instability which can be amended with a smaller learning rate or more asynchronous threads. The rewards for FuN change more aggressively compared to the other agents. This is likely due to the randomly emitted Manager goals exploring a trajectory toward a more advantageous direction, and the Worker catching up to that direction.

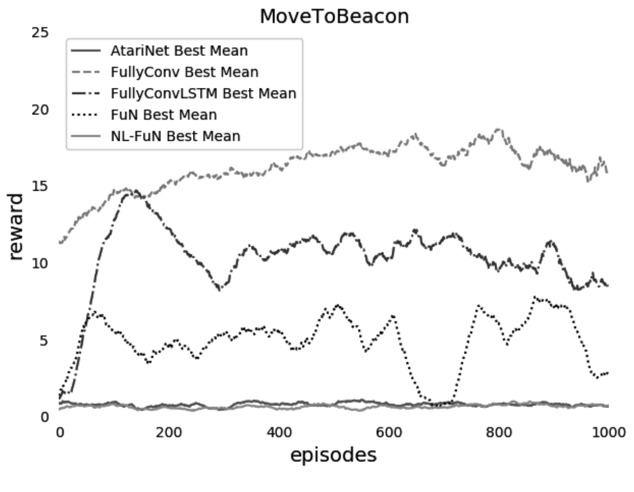

Fig. 2. Best MoveToBeacon Runs

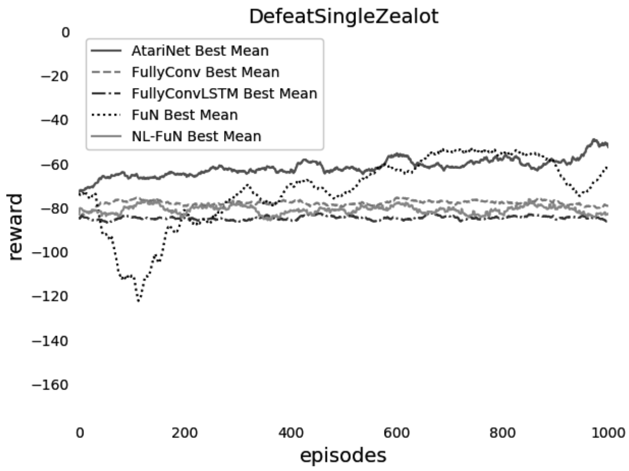

Fig. 3. Best DefeatSingleZealot Runs

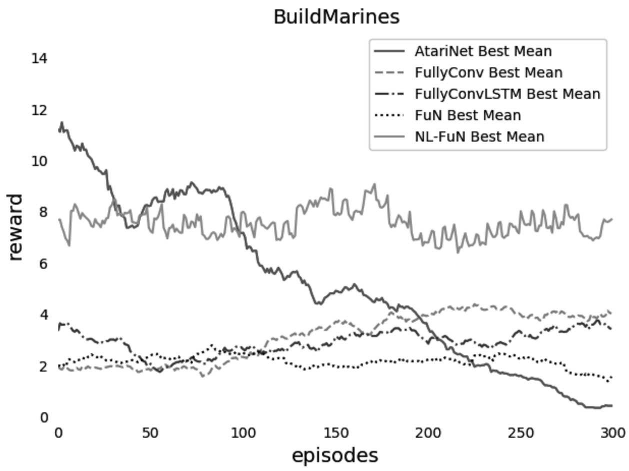

Fig. 4. Best BuildMarines Runs

CONCLUSION

This paper presents NL-FuN, a hierarchical network generalising the FuN architecture to three hierarchical tiers consisting of a Super Manager, a generalised Manager, and a Worker. Managers share attributes with both the Super Manager and the Worker, obtaining goal trajectories in their own latent state space with respect to the goal embedded by their own manager. This allows an arbitrary amount of additional Managers to be introduced to the system.

NL-FuN has trouble performing precise actions over short periods of time, but performs well over long term tasks. The problem might be that the assumption that goals follow a von Mises-Fisher distribution does not hold as strongly when additional Managers are added. Additional work is required to obtain insight into this interaction.

Contributions

Four contributions are outlined in this paper: Replicating three baseline agents developed by DeepMind (Vinyals et al., 2017); adapting FuN (Vezhnevets et al., 2017) to the SC2 domain; constructing NL-FuN, an extension of FuN, supporting at least three hierarchical tiers, with a generalised Manager that allows an arbitrary number of tiers to be added; DefeatSingleZealot, a new scenario to benchmark the agents.

NL-FuN is designed for long-term planning in environments with sparse rewards. In the BuildMarines scenario, NL-FuN obtained a better result than the other AI agents, including the original DeepMind baselines, in less time steps (540,000 vs 600,000,000). The BuildMarines scenario is particularly challenging due to its large state space and sparse rewards, requiring agents to make moves hundreds of steps in advance of the reward.

FUTURE WORK

NL-FuN's performance in MoveToBeacon and DefeatSingleZealot calls for further improvement. With better hardware, one can use a deeper convolutional network to extract more meaningful spatial features from the state representation (LeCun et al., 2015), resulting in better choices for spatial actions. Implementing batch normalization would allow more aggressive learning rates, speeding up convergence (Ioffe and Szegedy, 2015). Sampling efficiency can be improved with Proximal Policy Optimization (PPO) (Schulman et al., 2017). Better hyper-parameters can be obtained with Population Based Training (PBT) (Jaderberg et al., 2017).

Another avenue for future work is a more rigorous exploration of the assumption that sampled state space directions follow a von Mises-Fisher distribution, given by Equation (11). In particular, an exploration of whether this assumption still holds when more than two hierarchical tiers are introduced. This affects each hierarchical tier's ability to converge to its manager's goal at different time scales.

The agent's weights can be initialised using supervised learning from replays of other agents. After initialisation, RL can be used to further strengthen its performance. This was used in the first public iteration of AlphaGo to initialise the value network (Silver et al., 2016).

Self-play allows the agent to experience a gradual increase in difficulty as it trains against itself. This method has been used to great success in other agents such as TD-Gammon (Tesauro, 2002) and AlphaGo (Silver et al., 2016).

REFERENCES

Pierre-Luc Bacon, Jean Harb, and Doina Precup. The option-critic architecture. In *AAAI*, pages 1726–1734, 2017.

James Bergstra and Yoshua Bengio. Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, 13(Feb):281–305, 2012.

Lucian Busoniu, Robert Babuska, and Bart De Schutter. A comprehensive survey of multiagent reinforcement learning. *IEEE Transactions on Systems, Man, And Cybernetics-Part C: Applications and Reviews, 38 (2), 2008*, 2008.

David Churchill. The current state of starcraft ai competitions and bots. 2017.

Graham Kurtis Stephen Erickson and Michael Buro. Global state evaluation in starcraft. In *AIIDE*, 2014.

Geoffrey E Hinton. Neural networks for machine learning - lecture 6e: rmsprop: Divide the gradient by a running average of its recent magnitude. 2016. URL `http://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf`.

Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *ICML*, 2015.

Max Jaderberg, Valentin Dalibard, Simon Osindero, Wojciech M Czarnecki, Jeff Donahue, Ali Razavi, Oriol Vinyals, Tim Green, Iain Dunning, Karen Simonyan, et al. Population based training of neural networks. *arXiv preprint arXiv:1711.09846*, 2017.

Marc Lanctot, Vinícius Flores Zambaldi, Audrunas Gruslys, Angeliki Lazaridou, Karl Tuyls, Julien Pérolat, David Silver, and Thore Graepel. A unified game-theoretic approach to multiagent reinforcement learning. In *NIPS*, 2017.

Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.

Yuxi Li. Deep reinforcement learning: An overview. *arXiv preprint arXiv:1701.07274*, 2017.

Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin Riedmiller. Playing atari with deep reinforcement learning. *arXiv preprint arXiv:1312.5602*, 2013.

Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, et al. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533, 2015.

Volodymyr Mnih, Adria Puigdomenech Badia, Mehdi Mirza, Alex Graves, Timothy Lillicrap, Tim Harley, David Silver, and Koray Kavukcuoglu. Asynchronous methods for deep reinforcement learning. In *International Conference on Machine Learning*, pages 1928–1937, 2016.

S. Ontanon, G. Synnaeve, A. Uriarte, F. Richoux, D. Churchill, and M. Preuss. A Survey of Real-Time Strategy Game AI Research and Competition in *StarCraft*. *IEEE*, 5(4), 2013.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

David Silver, Aja Huang, Chris J Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, et al. Mastering the game of go with deep neural networks and tree search. *Nature*, 529(7587):484–489, 2016.

Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*, volume 1. 2nd edition, 2017.

Richard S Sutton, Doina Precup, and Satinder Singh. Between mdps and semi-mdps: A framework for temporal abstraction in reinforcement learning. *Artificial intelligence*, 112(1-2):181–211, 1999.

Gabriel Synnaeve. Bayesian programming and learning for multi-player video games. *Disertacnı práce*, 2012.

Gerald Tesauro. Programming backgammon using self-teaching neural nets. *Artificial Intelligence*, 134(1-2):181–199, 2002.

Alexander Sasha Vezhnevets, Simon Osindero, Tom Schaul, Nicolas Heess, Max Jaderberg, David Silver, and Koray Kavukcuoglu. Feudal networks for hierarchical reinforcement learning. In *ICML*, 2017.

Oriol Vinyals, Timo Ewalds, Sergey Bartunov, Petko Georgiev, Alexander Sasha Vezhnevets, Michelle Yeo, Alireza Makhzani, Heinrich Küttler, John Agapiou, Julian Schrittwieser, et al. Starcraft ii: a new challenge for reinforcement learning. *arXiv preprint arXiv:1708.04782*, 2017.

ADDING VARIETY IN NPCS BEHAVIOUR USING EMOTIONAL STATES AND GENETIC ALGORITHMS: THE GENIE PROJECT

Federica Agliata†, Marcello Bertoli†,
Laura Anna Ripamonti*, Dario Maggiorini* and Davide Gadia*
Department of Computer Science, University of Milan
via Celoria 18, 20133 - Milan (Italy)
E-mail: †{firstname.lastname}@studenti.unimi.it, *{firstname.lastname}@unimi.it

KEYWORDS
Games, Artificial Intelligence, NPC Behaviour, Genetic Algorithms

ABSTRACT

In recent years we have been observing an increasing adoption of artificial intelligence in video games. With the increasing availability of powerful hardware and advanced algorithms, we can now scale up the quality of the AI available to every Non Playing Character (NPC). Despite this increased quality, every NPC can get predictable with time and game designers are struggling to provide variety in games where many NPCs are present. Designing a specific, and unique, AI for every NPC can be a very time and resource consuming task. In this paper we propose GENIE (Genes Driven Decision Tree) as a tool to support game designers in the creation of a wide variety of behaviours. With GENIE, it is possible to define an NPC behaviour in term of internal parameters representing its state.

INTRODUCTION

As of today, we can observe an increasing adoption of artificial intelligence in video games. In modern video games, artificial intelligence techniques are used in a wide range of activities; from pathfinding to procedural content generation, to machine learning. Among all the possible tasks, providing complex and believable behaviour to *Non Playing Characters* (NPCs) is strategic to convey a rich and compelling player experience.

Despite the possibility to describe very articulated NPCs behaviour, every NPC will get predictable over time. NPCs are getting predictable to players whenever the same behaviour and/or decisions pattern is proposed over a long time: a human counterpart is not going to find this amusing or challenging. Another problem is about providing variety when a population of NPCs is involved: as a matter of fact, having all the NPCs behaving in exactly the same way will provide a poor player experience.

To support game designers in providing behavioural variety on a population of NPCs we propose here a system based on genetic algorithms. We baptised our system GENIE (*Genes Driven Decision Tree*). GENIE allows a game designer to define a set of possible states for every NPC and represent them in term of internal parameters. These internal states can be seen as emotional states for the NPC; in the same vein as human behaviour is affected by emotions felt while taking decisions, the behaviour of an NPC is affected by its current state. From a more technical standpoint, internal parameters can be used to drive a decision tree representing the NPC tactical and/or strategical behaviour. GENIE uses a genetic algorithm to generate internal states and provide multiple, and changing, behaviours to a population of NPCs. The fitness function of the genetic algorithm can be tuned to follow player's reactions and adapt game difficulty for an optimal user experience.

In this paper we are going to describe in detail how GENIE works and, to prove the effectiveness of our solution, we will present results obtained by testing GENIE on four games belonging to different genres.

RELATED WORK

The first historical example of AI applied to games where the player was supposed to confront NPCs exposing different behaviours is the game *Pac-Man* in 1980. In late 90's, agents in games started using information from the surroundings to influence decision making such as in the case of *GoldenEye 007*, *Thief: The Dark Project*, and *Metal Gear Solid* where allies' status was taken into account. Also in the late 90's, the newborn genre of *Real-Time Strategy* (RTS) introduced the adoption of a very large number of NPCs on the playfield. With RTS games, NPCs started using interaction with one another to implement strategies.

Moving now to a more scientific ground, we can find a large number of contributions addressing the problem of changing behaviours using an algorithmic approach. To the best of our knowledge, despite this wide literature, only a subset seems to be actually related to gaming.

Generative approaches are usually applied to games with the aim to develop a human-like behaviour for NPCs. As an interesting application, (Arrabales et al., 2012) uses cognitive architectures to address the design of believable bots for *First Person Shooter* (FPS) games. A more general approach is discussed in (Asensio et al., 2014), where the problem of believability is also taken into account trough Turing-like tests performed during live gameplay.

Despite their applicability, generative approaches do not provide a solution to the problem addressed in this paper: a new generated behaviour might be too different from the previous one and break continuity in use experience. A better solution could be to generate an initial (believable) behaviour and then evolve it. In (Lim et al., 2010), behaviour trees are evolved and recombined to raise the competition level of an NPC, while in (Schrum et al., 2012) a neural network is used to boost performances of an agent playing an FPS game. In (Floreano and Keller, 2010), a robot behaviour is evolved in order to improve survival probability, while in (Vaccaro and Guest, 2005) evolutionary computation is used to find optimal end moves for the tabletop game *Risk*. In all examples above, evolution is intended as a way to improve performances through generations and outperform a human player. While this is reasonable, to some degree, in games such as racing or chess, the resulting player experience will be poor in plot-driven games. NPCs must not be unbeatable: they are supposed to provide a reasonable challenge to the player and accompany her through the skill progression during the game.

The use of *Genetic Algorithms* (GAs) proved to be an interesting approach to the purpose of this paper. Since a genetic algorithm evolves a population by breeding *eligible* subjects over time, we should not observe sudden and abrupt behaviour changes between generations. Moreover, eligibility to reproduction can be tuned for optimal player experience.

Genetic algorithms have already been used to evolve soccer players (Whiteson et al., 2005) for the *RoboCup* competition, robots to be trained for space battles (Stanley et al., 2005), opponents in RTS games (Louis and Miles, 2005), tuning of FPS bots (Cole et al., 2004), and designing chess platers (Hauptman and Sipper, 2005).

Unfortunately, all the aforementioned applications of GAs to games are still targeting the evolution of the best possible player. To the best of our knowledge, there are no contributions about genetic evolution in games with the purpose to evolve NPCs which is deemed optimal for player experience. In our vision, evolution should not be leading to a specific target but provide continuous changes to follow the player skill and keep her in the game flow.

GENIE

GENIE (*Genes Driven Decision Tree*) is a software tool to be used within the Unity game engine. The purpose of this tool is to ease the design of multiple, variated, behaviours for NPCs in a video game. Branching conditions for the generated tree will be triggered by variables defining the internal (emotional) state of each NPC. This way, each NPC will offer a slightly different behaviour depending on its state, starting from the template.

In particular, the internal state of each NPC is defined by a set of possible emotion that the NPC can "feel". To each emotion, we associate a floating point value in the range $[0, 1]$. In this scale, the value 1 means maximum intensity for an emotion while 0 means that the emotion is absent in the NPC. The combined values of all emotions represent the emotional state of the NPC.

Inside the decision tree, decision nodes can define a threshold for each emotion. Emotions are not the only elements in play when selecting a branch: decision nodes may also need to check the surrounding environment, depending on the game.

While the emotional state is fixed inside each NPC, changes will take place when breeding new generations using a GA. The adoption of a GA grants a smooth and uniform transition between generations and allows random, unpredictable, changes to be added thanks to mutations. The intensity of each emotion is used as a chromosome for the evolution. We implemented the genetic crossover using the *single-point crossover* algorithm and mutation by selecting a random chromosome (a random emotion) and setting it to a random valuer. The adoption of this lightweight algorithm is also helpful to improve scalability in games where thousands of NPCs are required, such as RTS games. While the purpose of the selection phase is very simple: picking genomes (NPCs) eligible for breeding, this operation is always tightly coupled with game mechanics and player experience. As an example, an FPS game might want to select NPCs with a long lifespan while avoiding those who already killed the player; this might be reasonable to make the population challenging for the player, but not too powerful. For these reasons, GENIE is not implementing any fitness function by itself but is delegating the evaluation to the developer.

While implementing GENIE, a number of choices have been made to find an acceptable compromise between usability and complexity. The compromise we found is to have a general purpose tool but, to increase usability, we limited it to four mainstream game genres. The four game genres we selected for this implementation are: FPS, stealth, role-playing, and roguelike games. For each genre in this list, we are proposing (after thoughtful discussion with game designers) a reduced set of emotions deemed useful to evolve the specific NPCs. The emotions selected of each game genre are reported in Tab. 1. Anyway, given the object-oriented nature of Unity, it is possible for the final user to easily extend these sets and support new game genres. Moreover, to make de designing phase easier, we also decided to implement binary decision trees. This is not going to be an actual limitation because it has been demonstrated that

Table 1: Emotions Associated to Genres Inside GENIE

FPS	Stealth	Role-playing	Roguelike
Afraid	Bold	Anxious	Angry
Angry	Forgetful	Cautious	Coward
Bold	Paranoid	Considerate	Greedy
Tactical	Strategic	Panicked	
Yielding		Self-Assured	
		Shy	

the expressivity of binary decision trees is not a subset of the generic multi-branched version.

EXPERIMENTAL EVALUATION

To evaluate the actual effectiveness of GENIE, we performed experiments with games implemented using our tool. A game for every supported genre has been implemented and tested with actual players.

For the evaluation, we engaged a group of 20 volunteers. This group was made of Computer Science students with an age between 23 and 28 years. In the group, we had 18 males and 2 females. All the subjects declared to be active players and to be familiar with all the proposed genres.

We asked every volunteer to have a play session with each game and then fill in a feedback form. With this feedback form we aimed to understand if the player was actually perceiving a difference in the behaviours of the NPCs while playing. During analysis, for each experiment (genre), we classified the couples $[action, emotion]$ in three groups based on the share of players that perceived them: 66% or more, between 33% and 65%, and less than 33%.

In order to be able to compare results, demo levels have been designed following a common structure. For FPS and stealth games, where players are usually required to follow a given path, we adopted a linear level structure where three gameplay events (missions) are proposed in sequence, as depicted in Fig. 1. To complete the level, the player must survive all the events. For role-playing

Figure 1: Linear Level Organisation for FPS and Stealth Experiments

and roguelike levels, where players have more freedom to roam the map, we adopted a non-linear approach as reported in Fig. 2. In this non-linear approach there are three events available in the first part of the level. After surviving all the events in any order, the player can access the second part of the level through a bottleneck section. In the second part, the same pattern encountered before the bottleneck is proposed again in order to beat the level.

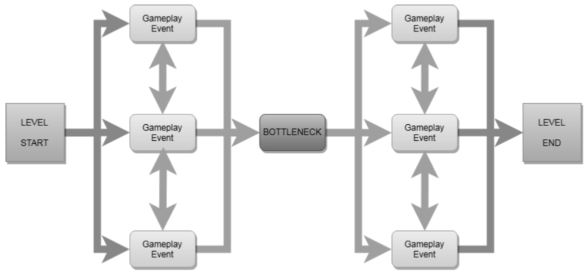

Figure 2: Non-Linear Level Organisation for Role-Playing and Roguelike Experiments

In all the games used for testing, fitness functions have been implemented using a scoring system. A score is associated to each NPC action: when the fitness function is run, the NPC history is evaluated and an integer number is returned. NPCs reporting the highest score are selected for breeding.

First Person Shooter Experiment

In an FPS, the player is engaged in a weapon-based combat simulation using a first-person perspective. During the game, a sequence of missions must be completed while fighting waves of NPCs. The artificial intelligence driving the skills of the NPCs is significative of the level of difficulty offered by the game. In this kind of games, NPCs should usually vary the attack style, how they move around, and how they take cover.

For this experiment we implemented a level with a combat zone in a harbour. On the map, players and NPCs can find environmental elements offering shelter. Some of this elements are containers providing ammunitions or safe passage between zones. During gameplay, a player must confront with waves composed by five NPCs.

In the decision tree of the NPCs, a lot of conditions must be evaluated. First, we have to check if the player is in the range of visibility; then, considerations about the current weapon and other equipment available on the fields are drawn. Just to give a couple of examples, taking cover takes precedence if the NPC is afraid to fight while being bold is pushing to engage fighting even if the NPC have a melee weapon (the player has always a rifle). A screenshot of the game when dealing with two tactical-inclined NPCs (pink particle effect) is shown in Fig. 3.

Stealth Experiment

In a stealth-based game, the player is required to move while staying undetected across an area guarded by NPCs. The player is usually subject to a swift death when confronting an NPC directly. In this kind of games, the difficulty offered to the player depends on the

Figure 3: Screenshot During the FPS Experiment

movement pattern, lever of awareness, and sensing sensibility of the NPCs. To put variety in a stealth game, NPCs should vary their pause/movement pattern as well as their policy about looking for (or chasing) the player. Our stealth game is set inside a small museum. In this museum, guards are deployed to protect the artworks. The player must traverse the map to steal a treasure and then leave the premise. Along the way, the player needs to retrieve two keys: one to access the treasure and one to open the exit door. While doing this, the player must remain unseen from the NPCs and avoid generating sounds by bumping into obstacles. When the player is detected, all NPCs will converge to the point where something has been spotter or a sound heard.

During gameplay, NPCs can be static or patrolling an area using a pre-determined path. The player is equipped with a torch to increase environment visibility and a crowbar to stun guards from the back; both, when used, increases the chances to be detected. The game ends whenever the player exits the building or is caught by a guard.

The decision tree for the stealth experiment is simpler than in the previous case but must include the possibility to coordinate with other NPCs. When the player is detected, the number of nearby NPCs comes into play and the guard might start chasing the player or move away and call for backup.

Role-Playing Experiment

A *Role-Playing Game* (RPG) is a game in which players assume the roles of characters in a fictional setting. A character must overcome a sequence of challenges in order to progress in experience and become more powerful. In RPG games, each NPC usually falls in a specific category and has its own statistics. Artificial intelligence must be specialised for each category, which should also evolve independently.

The RPG game we implemented is located in a dungeon made of rooms connected by corridors. This dungeonis populated by fancy creatures. In this game we implemented four different NPCs: *Minion*, *Melee*, *Mage*, and *Healer*. Minion and Melee are close-combat unit with different attack power, while Mage is a ranged combat unit, and the Healer will support other NPCs in the area.

In this experiment we deal with multiple decision trees: one for each NPC class. Even if the emotional traits are shared, each class will behave in its own way.

Roguelike Experiment

The roguelike genre is a sub-category of the RPG genre. A roguelike game requires the player to crawl through a dungeon to kill monsters (NPCs), collect treasures, and interact with the environment. In roguelike games, artificial intelligente of NPCs is typically limited to movement and attack strategies. Differently from the RPG experiment, the player has to follow a specific path to reach the final treasure and it is required to overcome all the enemies along the way. The player is equipped with a ranged magical weapon and can collect power ups as loot from slayed NPCs. A screenshot during gameplay is proposed in Fig. 4.

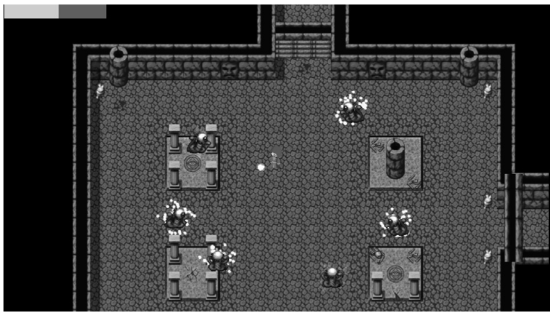

Figure 4: Screenshot During the Roguelike Experiment

Like in the previous case, in this experiment we are dealing with multiple decision trees. In particular, we defined seven different classes of NPCs, each one with different special abilities.

Experimental Results

As already mentioned, we classified each action associated to a specific emotion in three groups based on the share of players that perceived them. The outcome from the feedback forms is summarised in Tab. 2. As we can see in the table, a vast majority of the actions relative to each emotion was apparent to more than two thirds of the players. From these numbers, it seems that the strategy adopted by GENIE is successful in producing different behaviours.

In order to cross-check this result, we ran another set of experiments with a different feedback form. In this second set, we aimed to understand if the player could actually tell apart the different emotional states of the NPCs. Volunteers have been instructed in advance about the different emotions available in the game. Then, after

Table 2: Summary of Perceived Actions During Gameplay

Share of players	Experiment			
	FPS	Stealth	Role-play	Roguelike
less than 33%	1	0	0	0
between 33% and 66%	2	1	3	0
more than 66%	12	6	11	8

each play session, they reported how well the emotion was represented by the actions of each NPC.

Results show that there is actually a perceivable link between emotions and actions since the majority of feedbacks reported a good in-game representation. Anyway, we observed one exception when analysing the FPS experiment, as reported in Tag. 3. As we can see, the yielding emotion seems to be the only one not well represented, trending to a *fair* rating. To understand this phenomenon, we did some oral interviews with the volunteers. After the interviews, our hypothesis is that, even if yielding is important for a game designer to characterise an NPC, the player is not expecting to actually see it on the battlefield; therefore, attention for that kind of behaviour was low during the experiment.

Table 3: Aggregated Feedback About Emotions Representation for the FPS Experiment

Emotion	Perception				
	Poor	Fair	Good	Very good	Perfect
Afraid	1	1	5	5	1
Angry	0	1	3	8	1
Bold	0	0	8	4	1
Tactical	1	1	7	2	2
Yielding	1	6	5	1	2

CONCLUSION AND FUTURE WORK

In this paper we presented GENIE: a tool to support game designers in the creation of a wide variety of behaviours. GENIE is based on discrete representation of emotional states used as genomes for a generic algorithm. Emotional states, used to drive decision trees, are evolved to adapt NPCs to the player's needs and to enrich player experience by offering variety during gameplay. Experimental evaluation on four games provided promising preliminary results supporting the validity of our approach.

Possible extensions of the current work are the inclusion for validation of other game genres and a study about how it could be possible to evolve the emotional state together with the NPC skills and features, like in (Guarneri et al., 2013).

REFERENCES

Arrabales R.; Muñoz J.; Ledezma A.; Gutierrez G.; and Sanchis A., 2012. *A Machine Consciousness Approach to the Design of Human-Like Bots*, Springer Berlin Heidelberg, Berlin, Heidelberg. ISBN 978-3-642-32323-2, 171–191.

Asensio J.M.L.; Peralta J.; Arrabales R.; Bedia M.G.; Cortez P.; and Pea A.L., 2014. *Artificial Intelligence approaches for the generation and assessment of believable human-like behaviour in virtual characters. Expert Systems with Applications*, 41, no. 16, 7281 – 7290. ISSN 0957-4174.

Cole N.; Louis S.J.; and Miles C., 2004. *Using a genetic algorithm to tune first-person shooter bots.* In *Proceedings of the 2004 Congress on Evolutionary Computation (IEEE Cat. No.04TH8753).* vol. 1, 139–145 Vol.1.

Floreano D. and Keller L., 2010. *Evolution of Adaptive Behaviour in Robots by Means of Darwinian Selection.* In *PLoS biology.*

Guarneri A.; Maggiorini D.; Ripamonti L.; and Trubian M., 2013. *GOLEM: Generator Of Life Embedded into MMOs. The 2018 Conference on Artificial Life*, 585–592.

Hauptman A. and Sipper M., 2005. *GP-EndChess: Using Genetic Programming to Evolve Chess Endgame Players.* In *Genetic Programming.* Springer Berlin Heidelberg, Berlin, Heidelberg. ISBN 978-3-540-31989-4, 120–131.

Lim C.U.; Baumgarten R.; and Colton S., 2010. *Evolving Behaviour Trees for the Commercial Game DEFCON.* In *Applications of Evolutionary Computation.* Springer Berlin Heidelberg, Berlin, Heidelberg. ISBN 978-3-642-12239-2, 100–110.

Louis S.J. and Miles C., 2005. *Playing to learn: case-injected genetic algorithms for learning to play computer games. IEEE Transactions on Evolutionary Computation*, 9, no. 6, 669–681. ISSN 1089-778X.

Schrum J.; Karpov I.V.; and Miikkulainen R., 2012. *Humanlike Combat Behavior via Multiobjective Neuroevolution*, Springer Berlin Heidelberg. 119–150. URL `http://nn.cs.utexas.edu/?schrum:believablebots12`.

Stanley K.O.; Bryant B.D.; and Miikkulainen R., 2005. *Real-time neuroevolution in the NERO video game. IEEE Transactions on Evolutionary Computation*, 9, no. 6, 653–668. ISSN 1089-778X.

Vaccaro J.M. and Guest C.C., 2005. *Planning an endgame move set for the game RISK: a comparison of search algorithms. IEEE Transactions on Evolutionary Computation*, 9, no. 6, 641–652. ISSN 1089-778X.

Whiteson S.; Kohl N.; Miikkulainen R.; and Stone P., 2005. *Evolving Soccer Keepaway Players Through Task Decomposition. Machine Learning*, 59, no. 1, 5–30. ISSN 1573-0565.

DATA ANALYTICS AND PLAYER BEHAVIOURAL ANALYSIS

The ACE2 Model: Refining Bartle's Player Taxonomy for Creation Play

Thomas van Dam
University of Amsterdam
thomasvdam@me.com

Sander Bakkes
Utrecht University
s.c.j.bakkes@uu.nl

KEYWORDS

Player motivation, player taxonomy, Bartle, creation play

ABSTRACT

This paper proposes a new taxonomy of video-game players: the *ACE2 model*. Building upon related work in player taxonomies, the ACE2 model sets out to refine the established Bartle's taxonomy of player types by incorporating the aspect of *creation play* (e.g., as exhibited in modern games such as *Minecraft*), thereby rendering the refined model more generally applicable to present-day video games. The paper considers the model part of an ongoing investigation into the relationship between aesthetics and mechanics in games. As such, the contribution of this paper lies not in proposing a definitive answer to taxonomic demarcation, it foremost attempts to highlight a creative play dimension that could be considered under-explored in classic player taxonomies. A model-validation method to this end, is to allow human participants to identify the subjective demarcation of creation play in a user study, in which the aesthetic / mechanics expressiveness of games is assessed by participants. The paper reports on the results of a first user study, set to obtain an early indication of the model's validity, prior to extensive validation experiments. These first studies that compare Bartle's model with the ACE2 model indicate that (1) the ACE2 model allows for a more articulate labelling of single-player video games, and that (2) even though creation play does not feature often, when it does it is a defining feature in modern games. As such, the paper concludes by suggesting that (a) the descriptive expressiveness of the ACE2 model provides a substantial and functional refinement of Bartle's taxonomy of player types, and (b) further investigation of the interplay of aesthetics and mechanics – as experienced by game players – may yield important insight in (the taxonomic understanding of) creation play in games.

1 INTRODUCTION

Player modelling is a research area in game playing that is gaining attention from both game researchers and game developers. It concerns generating models of player behaviour and exploiting the models in actual play. The general goal of player modelling often is to steer the game towards a predictably high player satisfaction [35] on the basis of modelled behaviour of the human player (i.e., in-game and/or real-world behaviour). Moreover, next to being useful for entertainment augmentation, player models are useful (among others) for game design purposes (e.g., analysing whether the design leads to gameplay as envisioned by the designers), for simulation purposes (e.g., simulating stories or evaluating game maps), and for serious game applications such as education (e.g., tailoring the game to a player's model for reaching particular learning objectives) or health (e.g., personalizing games for rehabilitation of elderly patients).

Indeed, player modelling is of increasing importance in modern video games [16]. The main reason is that player modelling is almost a necessity when the purpose of AI is 'entertaining the human player' [35], with the human player and his/her affective response to a designed experience being largely unknown. One common method for player modelling, is to build on the established taxonomy of players by Bartle [5]. In general terms, the taxonomy demarcates between players being achievers, explorers, socializers, and killers. While the taxonomy is tailored to multi-user dungeon games (MUDs), the simplicity (and perhaps elegance) of the model render is somewhat suitable for application to modern video games as well [33].

However, there has been a fair amount of criticism on Bartle's model, noteworthy also by Bartle himself, who states that his taxonomy might be incomplete for games other than multi-user dungeon games [7]. Indeed, games have evolved substantially since 1996, with new manners of behaviour being exhibited which are not encapsulated in Bartle's taxonomy of player types. While numerous alternative models are investigated (e.g., Yee's seminal work on MMPORPG's [37], the Four Keirsy Temperaments [21]; the Demographic Game Design model [9]; and the Unified Model [33]) – as discussed further in the related work section – we observe that the alternatives do not explicitly consider a vital aspect of numerous modern video games, namely the aspect of *creation play*. Here, we consider creation play to be exhibited play behavior with no explicit purpose other than to build or create whatever the player desires – and will further demarcate the term in the next section. Broadly formulated, creation play is play behaviour beyond the traditional explorer type – that is also (a) interacting (b) with the game world – but done so for (often purposely) exploratory or goal-directed reasons, while creation play can be consider play behaviour without explicit purpose.

Indeed, the popularity of sandbox games such as *Minecraft* reveals that there is a strong desire for games that allow such expression. Furthermore, related work reveals that so-called sandbox players are motivated by a unique set of motivators that are not reflected in any existing player model [13, 36]. Throughout the course of the paper – and building upon a user-study – we will therefore advocate that the creation aspect of games should be seen as its own distinct category.

As such, this paper contributes a new taxonomy of video game players: *the ACE2 model*. Founded on related work in player taxonomies, the ACE2 model refines the established Bartle's model by incorporating the aspect of creation play, thereby rendering the refined model more generically suitable for present day video games.

2 RELATED WORK

The relationship between aesthetics and mechanics may be considered a foundational theme of game studies. It has already been discussed widely, for example, in terms of core and shell [26], aesthetic qualities and formal structures [29], visual appearance and procedural rhetoric [11]. One may correctly note that the exact nature of this relationship has been discussed in term of "tight coupling" [10], "seeing past fiction" [22], or a relationship in which a fictional surface layer helps the player understand the game's goals, and then fades to the back of the mind [20].

This paper attempts to tread carefully on these complexities, as they indeed cannot simply be reduced to an either/or: the player can care for both aesthetics and mechanics. A tinkering creator can care for mechanics; building something can be both an "aesthetic" act as it can be a "mechanical" act.

As such, the contribution of this paper lies not in proposing a definitive answer to taxonomic demarcation, it foremost attempts to highlight a creative play dimension that could be considered under-explored in classic player taxonomies. A model-validation method to this end, is to allow human participants to identify the subjective demarcation of creation play in a user study, in which the aesthetic / mechanic expressiveness of games is assessed by the participants. Thereby, one may indeed highlight the nature of creation play, which may arguably be a unique play dimension between aesthetics and mechanics.

To provide further context for the paper, we will go further into (1) Bartle's taxonomy of player types, (2) alternative player models, and (3) will provide concise context on the topix of player modelling.

2.1 Bartle's taxonomy of player types

Barte's taxonomy of player types was derived from the author's investigation into why people play MUDs. That is, when summarising the contents of his investigation Bartle saw a pattern emerging; most reasons for playing could be grouped up in four distinct player categories [5], illustrated in Figure 1. Bartle constructed two axes to map the four categories, based on the sources of interest that each player category has in the game. On the x-axis there is a focus on players on the left, versus a focus on the game world on the right. The y-axis goes from a focus on acting at the top, to a focus on interacting on the bottom. The player types are situated in the quadrants associated with their interests. An informal description of the categories is as follows.

Achievers focus on acting on the game world, which boils down to doing things in the game. They care little about the other players in the game, or about the intricacies of the game if it does not result in them gaining more points. *Explorers* are interested in interacting with the game world, always looking for new things in the game. They thrive on being surprised by the game, but not so much by other players. *Socialisers* focus on interacting with other players. They want to get to know new players and engage in social activity with them. For them, the game world is mostly a backdrop to their social engagements. *Killers* are looking to impose themselves on others, acting on players rather than the game world. They thrive on demonstrating how superior they are to other individuals.

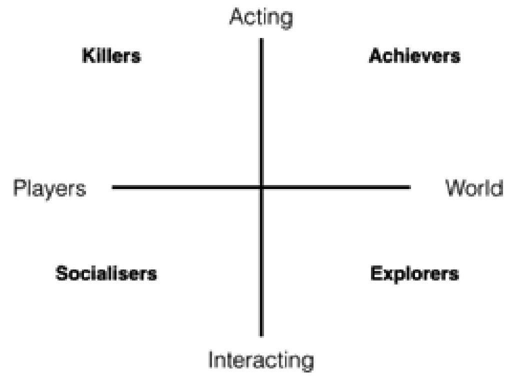

Figure 1: Bartle's taxonomy of player types [23].

Strengths of Bartle's Taxonomy. Perhaps one of the biggest strengths of Bartle's model is its simplicity. With just four player types, divided over two distinct axes it is easy to comprehend and intuitive to use. Additionally, the use of a scale allows for player models to have varying degrees of interest in the aspects of the game. A player is usually not limited to one style of play, and can dabble in other styles from time to time. Bartle's model can account for this by assigning values to each of the axes for a player, creating a multi-dimensional model rather than just a single player type. The fact that classifications similar to that of Bartle are widespread also adds merit to the quality of this type of classification. As Stewart notes, a great deal of player models are very similar to Bartle, and thus to one another [33]. Further on in this paper we will take a closer look at these other models. In addition to scientific player models, there are also industry examples of companies that use a classification which shares similarities with Bartle's model.[1]

Shortcomings of Bartle's Taxonomy. The main shortcoming of Bartle's Taxonomy of Player is that it is tailored for multi-user dungeon games (MUDs), and not present-day video games. This has made it difficult to use the model in games that are distinct from its original design-purpose, even Massive Multiplayer Online Role-Playing Games, which share many similarities with MUDs [7]. Indeed, applicability of the model is further reduced by the fact that MUDs (and Massively Multiplayer Online games in general) are steadily declining in popularity [8]. Pigeonholing Bartle's model even further is the fact that it was developed based on an online multiplayer game. This means that all games which focus more on delivering a *single player experience* are hard to classify using Bartle's model.

2.2 Alternative Player Models

Indeed, numerous other models exist that aim to categorise players by their playing style.

[1]Noteworthy is the model employed by Wizards of the Coast in their design of new cards for Magic: The Gathering [28]. They use a cast of three player types: Timmy, Johnny, and Spike, which roughly correspond to Bartle's Socialisers, Explorers, and Achievers. In addition, they also allow for players to associate with multiple playing styles in varying degrees of intensity. A possible reason for not having a Killer equivalent in the model Wizards of the Coast employ might be that the multiplayer aspect of the game is in most cases mutual. Players agree to play a game with each other, whereas in MUDs the players are placed in a game with random other players.

Yee's seminal work on MMORPG demographics, motivations and experiences [37] relates to the present research too. That is, an exploratory factor analysis revealed a five factor model of user motivations for MMORPG game – achievement, relationship, immersion, escapism and manipulation – illustrating the multi-faceted appeal of these online environments [37]. Indeed, the multi-faceted appeal of games may be particularly present in single player games too, and may not be appropriately captured by Bartle's model.

Tuunanen and Hamari's work [17] – while not directly focused on the descriptive expressiveness of a model, but on how players have been categorized in game research literature – also provides relevant input to our investigation. Their study suggests that player typologies in previous literature can be synthesized into seven key dimensions: skill, achievement, exploration, sociability, killer, immersion and in-game demographics [17]. These additional dimensions of player categorisation indicate, as we also do in the present paper, that important dimensions of player expressiveness (and thereby, player-driven game categorizations) are not fully addressed in established player taxonomies.[2]

Also, a particularly interesting model is the *Four Keirsy Temperaments* [21], which uses a categorisation very similar to Bartle's. These were not derived from people playing games, but rather a pattern Keirsey observed from the sixteen types of the Myers-Briggs personality model. These four categories are high level constructs of personality traits, which can be seen as a superset of Bartle's player types [33]. Even though Keirsey's Temperaments are not specifically tailored to games, they do allow for categorisation based on the type of behaviour a person exhibits in the world, or in a game world [33].

Another four type model is the model constructed by Bateman, the *Demographic Game Design model* (DGD1) [9]. Through observation of video games Bateman came to four player types that are all slightly different from the four Bartle types. However, as Stewart notes, it is possible to construe the types of the DGD1 model as hybrids of the Bartle types [33]. By elaborating on the Hardcore and Casual modes described by Bateman [9], Stewart [33] created six types that function as all possible hybrid combinations of the Bartle types.

Finally, an interesting model is the *Unified Model*, by Stewart [33]. This model incorporates the different player models that we already touched upon in the previous paragraphs. He shows that a number of the most well-known player models as well as game design models share so many conceptual elements, that – conceptually – it is possible to combine them all in a single model [33].

However, in all the different aforementioned models we observed that most did not explicitly deal with the *creation play* aspect that some players enjoy in video games.[3] The popularity of sandbox games such as *Minecraft* indicates that there is a desire for games with no explicit purpose other than to build or create whatever the player desires. Most models regard building as a component of simulation, where the player wants to copy something from the real world. While the unified model does consider creative building, it is shoehorned into Bartle's explorer category [33]. Indeed, research has shown that sandbox players are motivated by a unique set of motivators that are not reflected in any existing player model [13, 36]. As such, we would like to argue that the creation aspect of games should be seen as its own separate category.

[2] Tuunanen and Hamari's go so far as suggesting the self-fulfilling and self-validating nature of the current player taxonomies, because their relatively high use in game design practices – as well as discusses – the role of game design in segmentation of players [17].

[3] On a historic note, one may observe that Caillois already showed awareness of the category of "construction games", which he subsumed under mimicry [12]. The historically interested reader may also appreciate Liboriussen's application of craft theory to game studies [25].

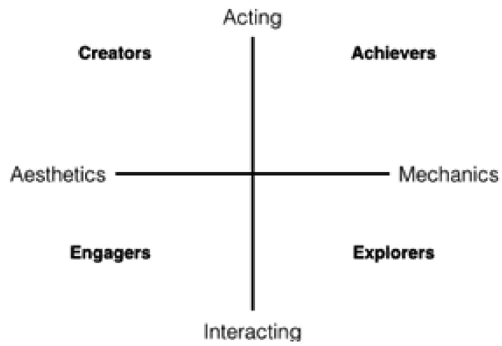

Figure 2: ACE2 taxonomy of player types.

2.3 Player Modelling

Player modelling is a research area that focuses on analysing how players go about in playing the games that they play, and then using this information for various ends [14, 18, 27, 34]. In this context, player modelling is generally concerned with four goals, namely (1) providing an interesting or effective experience on the basis of player models, (2) creating a basis for game developers to personalize gameplay as a whole, (3) creating new user-driven game mechanics, and (4) allowing the game developer to analyse how distinct groups of players interact with and respond to the game design [4]. In this paper, we will exclusively investigate the construction of models based on behaviour that is exhibited by a player within a game environment.

Noteworthy of the present paper, is that it analyses which games facilitate which specific play behaviours, according to both Bartle's reference model, as the new ACE2 model. Such an analysis reveals both the expressiveness of the investigated models, and indicates which player styles are facilitated within a games. Particularly this later aspect of player models, makes them applicable for use within the game development process [27, 34] and for game analysis [5, 33].

3 THE ACE2 MODEL

We propose a new taxonomy of video game players: the ACE2 model. Building upon previous work in player taxonomies, ACE2 refines the established Bartle's model by incorporating the aspect of creation play, thereby rendering the refined model more generically applicable to present day video games. Figure 2 illustrates the axes and player types in the ACE2 model. At surface level, one observes that it is reminiscent of Bartle's taxonomy of player types, at least with respect to it utilising two axes and four player types. Below we discuss the motivation and design choices for the axes and player types.

3.1 Horizontal axis

As observed earlier, part of the weakness of Bartle's model lies in the fact that it is geared towards a very specific kind of game: MUDs. Since we wanted to create a model that was applicable to a wider variety of games we took a more abstract approach to games. However, we quickly observed that the multiplayer aspect of games adds numerous intricacies to the kinds of behaviour that players display, that we decided to restrict the model to single-player games. Indeed, this is a design choice that allows for a greater balance between model simplicity and model articulation than would have been possible had we included all kinds of games.

As Bartle's x-axis dealt with the distinction between the virtual world and its player inhabitants, we were no longer able to incorporate this axis. Instead we consider the axis to deal with different ways of players enjoying games. Indeed, there are numerous reasons why players enjoy playing games [2, 24], and these can reasonably be abstracted into two main categories which we labelled Aesthetics and Mechanics. Whilst the term Aesthetics is also used in the MDA model [19], here, we consider aesthetics to be the aesthetic elements of the game that do not belong to the gameplay. That is, e.g., the narrative of a game, its visual style (or lack thereof) [31], the soundtrack, etc. etc. On the other side of the axis we place the Mechanics, which are the elements of the game that comprise the gameplay of a game, such as the actions that the player can perform in the game world, or the interaction between game elements.

3.2 Vertical axis

The vertical axis is exactly the same as it is in Bartle's model, since we observed that the distinction Bartle [5] makes between acting on the game world and interacting with the game world is explicitly (and particularly) present in single player games.

3.3 ACE2 Types

We will now describe all four player types of the ACE2 model, of which the model derives its name (Achievers, Creators, Explorers, Engagers).

3.3.1 Achievers. . The achievers in this model are closest to their Bartle counterpart, since they focus on acting on the game mechanics, which is similar in spirit to Bartle's achievers, who act on the game world. ACE2 achievers enjoy winning and gaining points like Bartle's achievers, but also enjoy obtaining mastery over the mechanics of the game. An example of mastering mechanics would be the ability to flawlessly execute complex combos in a fighting game, or perfectly time a jumping sequence in an action game. This way of enjoying games is not touched upon by Bartle.

3.3.2 Explorers. . The explorers closely resemble Bartle's explorers. They also seek to learn about the game's intricacies and quirks, but are more focused on the gameplay itself. Exploring terrain is not as interesting to them as it is to Bartle's explorers. They will often look for interesting interactions in games, such as unique combo's in deck building games, such as Hearthstone, or novel use of game mechanics. An example of the latter is 'snaking' in Mario Kart DS, a technique that uses the drifting mechanic, which was intended for taking corners, to increase the speed of the vehicle on straight sections of the track as well.

3.3.3 Engagers. . Engagers are the first completely new type, and focus on interacting with the aesthetics of the game. They are more interested in the story or views a game provides, and not so much the gameplay. They will often look for games that trigger an emotional response, or that allows them to form an emotional bond with the characters in the game. Interactive novels are an example of games that resonate with this player type, as these often provide minimal gameplay but instead deliver a rich aesthetic experience.

3.3.4 Creators. . Creators are the final player type in this model, and are also the type that sets the ACE2 model apart from most other models. While this kind of behaviour is often a minor part of a different category, or even completely disregarded, here it has its own player type. While these may appear counter intuitive, creators – like engagers – are drawn towards the aesthetics of a game, but seek to act on them rather than interact with them. This manifests as creating structures or visuals within the game, effectively using the game as a creative outlet. Creators can also use the game to create their own aesthetic experience as to trigger an emotional response in others who experience their work.

4 EVALUATION OF THE ACE2 MODEL

In order to analyse the conceptual refinement offered by the ACE2 model, we perform a user study in which the model is compared to Bartle's taxonomy of player types. The user study consists of a series of questionnaires in which participants were asked to rate how strong the focus on a particular kind of behaviour was in selected games. By looking at how the focuses are divided for both models we were able to compare the descriptive expressiveness of the models on the selected games. Here, we will fist describe (1) which modern video games were include in the study, (2) discuss the design of the questionnaire, and (3) present the experimental procedure.

4.1 Investigated Video Games

To ensure the inclusion of a wide variety of modern video games, we created a list of well-known games from many distinct genres. The existence of strictly-defined game genres is an ongoing topic of debate in the scientific community, despite the fact that the notion has been around for many years now [1]. For the purpose of the current investigation, we adopt the following commonly accepted game genres Action, Adventure, Role-playing, Simulation, Strategy (cf. [3, 15, 30, 32]), and Sandbox (cf. [36]). Indeed, a sandbox game is unique game genres in which the goals are set by the players themselves, which is why a pure sandbox game attracts a specific kind of player [36].

For each genre we selected three games in an attempt to cover as many of the sub-genres as possible. Some of the selected games were part of a series in which multiple games were nearly identical in terms of the gameplay they provided. In such cases all these games were grouped under the series. Table 1 shows a comprehensive list of all games considered for this study.

Table 1: List of the videos games that were included in the user study.

Genre	Archetypal game series	Matching inclusion criteria
Action	Super Mario Bros.	Super Mario Bros., Super Mario Bros. 2, Super Mario Bros. 3, New Super Mario Bros, New Super Mario Bros. 2, New Super Mario Bros. Wii, New Super Mario Bros. U.
	Street Fighter	Street Fighter IV
	Halo	Halo III and Halo IV
Adventure	Sam & Max series	Sam & Max Save the World, Sam & Max Beyond Time and Space, Sam & Max: The Devil's Playhouse
	Tales of Monkey Island	Launch of the Screaming Narwhal, The Siege of Spinner Cay, Lair of the Leviathan, The Trial and Execution of Guybrush Threepwood, Rise of the Pirate God
	The Walking Dead	Season 1: A New Day, Starved for Help, Long Road Ahead, Around Every Corner, No Time Left, 400 Days. Season 2: All That Remains, A House Divided, In Harm's Way, Amid the Ruins, No Going Back.
Role-playing	Baldur's Gate	Baldur's Gate, Baldur's Gate II, or their Enhanced editions.
	PokÃĬmon	Red, Blue, Yellow, Gold, Silver, Crystal, Ruby, Sapphire, Emerald, FireRed, LeafGreen, Diamond, Pearl, Platinum, HeartGold, Soulsilver, Black, White, Black 2, White 2, X, Y, Omega Ruby, Omega Sapphire.
	Final Fantasy	VII, VIII, IX, X, X-2, XII, XIII, XIII-2, Lightning Returns: Final Fantasy XII
Simulation	Sim City	Sim City 2000 and Sim City 3000
	Euro Truck Simulator	Euro Truck Simulator and Euro Truck Simulator 2
	Nintendogs	Nintendogs: Dachshund & Friends, Lab & Friends, Chihuahua & Friends. Nintendogs: Best Friends, Dalmatian & Friends. Nintendogs + Cats: French Bulldog & New Friends, Golden Retriever & New Friends, Toy Poodle & New Friends.
Strategy	Civilization	Civilization IV, Civilization V
	StarCraft	StarCraft, with or without the expansion Brood War, StarCraft II: Wings of Liberty, and StarCraft II: Heart of the Swarm
	Portal	Portal, Portal 2
Sandbox	Minecraft	Minecraft
	Garry's Mod	Garry's Mod
	Terraria	Terraria

4.2 Investigated Facets

All items in the questionnaire took the form of a question about how strong – according to the participant – the focus was in the game in question (e.g., "How strong is the focus on beating levels or opponents in the game"). The participant could answer on a five point Likert scale ranging from "Very Strong" to "Barely There". In addition, participants could also answer "Not Applicable" should they feel the item was not relevant to the game in question, or "Can't Remember" should they be unable to remember whether said element was present in the game or not. A list of facets investigated in the questoinnaire is provided in Table 2.[4]

Investigated facets applicable to Bartle's model are: Achievers (A1, A2), Explorers (A4, A5), Socialisers (A8, A9), Killers (A10, A11). Investigated facets applicable to the ACE2 model are: Achievers (A1, A2, A3), Explorers (A4, A6, A7), Engagers (A12, A13, A14), Creators (A15, A16, A17). Since we constructed our own items for this questionnaire we were very mindful of the fact that we could influence the results favourably for ACE2 just by how we chose the items. To prevent this we took special care to solely focus on the actual behaviours we observed in commonly-available gameplay footage, rather than on what would best differentiate the new model from Bartle's model.

[4]In addition, the Appendixes – available online at http://bit.ly/1nNDH0N – provide a full overview of the investigated questions and the accompanying results.

4.3 Questionnaire Procedure

Upon loading up the questionnaire the participant was greeted with an introduction screen where the goal of the questionnaire was briefly explained, as well as explaining what was expected of the participant in their answering of the questions (Appendix B). When starting the questionnaire, the participant was presented with a screen in which one could select the games with which they felt comfortable enough to answer questions about (Appendix C). For every selected game the participant was asked to fill in the questionnaire investigating applicable facets (Appendix A). In addition to the questions, the screen also showed the games in question, and a small reminder on how to judge certain questions.

For this first user study, set to obtain an early indication of the model's validity, 43 game players participated. Selection of the participants took place via convenience sampling of subject who

Table 2: List of facets investigated in the questionnaire.

Facet	Type	Category	Description
A1	Achievers	Winning	Beating levels or opponents in the game
A2		Gaining points	Increasing a value, be it experience points, gold, achievement points, or anything similar
A3		Mastering the game	Getting better and better at the game. The learning curve is a large part of the game
A4	Explorers	Finding interaction between game elements.	Discovering how game elements interact with each other, finding the limits of the game engine
A5		Finding unexplored territories	Discovering areas in the game that few other players have been to
A6		Finding alternate strategies	Beating levels in different ways than what is most obvious; finding new ways to accomplish something
A7		Finding the optimal solution or setup	Finding the optimal solution for a puzzle, or finding equipment/weapon combination that provide the best boosts
A8	Socialisers	Getting to know new players	Meeting new players and communicating with them to get to know them better
A9		Improving your social status in the community	Getting more players to know you and see you in a positive light
A10	Killers	Causing distress in other players	Interacting with other players in the game world as to ruin their day. Often by killing their in game character
A11		Imposing yourself on other players	(Forcefully) Interacting with other players in the game world
A12	Engagers	Experiencing the narrative of the game	The game features an extensive story
A13		Experiencing the visuals of the game	The game provides stunning views, or features a particular art style
A14		Interacting with the Non-player Characters of the game	Engaging in dialogue with computer controlled characters, or in other ways interacting with them
A15	Creators	Creating new levels.	Constructing new levels that are playable by others
A16		Creating your own structures, landscapes, or visuals	Using the game as a creative outlet. An example of visuals would be pixel art
A17		Creating your own narrative	Creating your own story for a custom campaign, or using the game to create a movie (machinima)

fit the following criteria (1) the subject plays games more than zero hours per week, (2) the subject at least a moderate game literacy, in being knowledgable and having personal experience with numerous classic game (e.g., Super Mario). The average age of the participants was 23 years. For this preliminary study no data on gender was collected for analysis.

4.4 Questionnaire Analysis

When analysing the results we transformed the answers given by the participants into their assigned ranks, which were averaged over the collected entries for the specific game / category. In the case a participant answered "Can't Remember" we did not take this answer into account in the average. This gave us a score for every category for both models, which we mapped on the plots shown below. The scores range from 0 to 5, where 0 means that this player type is not represented in the game at all according to the participants, and 5 that this is one of the main foci of the game. The reasoning behind this is that – while not a marker for model evaluation – it allows for a game to be visually identifiable through their shape on the plot, as well as making for an easier visual comparison of differences between the models in the results.

5 RESULTS

First we will discuss how the models compare over all games, looking at an overall analysis of the data. Second, we will take a closer look at each of the genres and how well the models are able to categorise their expressiveness.

5.1 Global analysis

By calculating the average for all player types among all games for both models we were able to create the plot that can be observed in Figure 3. The socialisers and engagers, as well the killers and

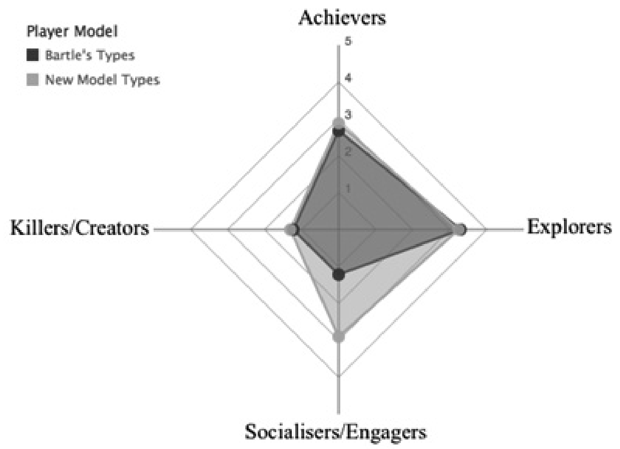

Figure 3: Global visualisation over all investigated games.

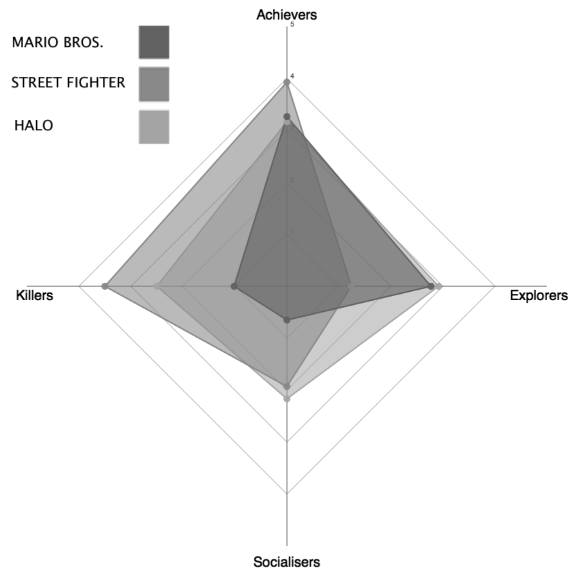

Figure 4: Action games in Bartle's model. The three profiles have very distinct shapes, which makes it difficult to create an abstraction for action games using Bartle.

creators have been put on the same ends of the axes in order to make comparison easier. While the two shapes are similar, the ACE2 model has three directions in which it expands, whereas Bartle's model only expands in two directions substantially. This indicates that participants were able to categorise with a higher degree of articulation in ACE2, since more relevant options were available to them. Appendix G1 plots the results for all games individually using Bartle's model. We observe that for Bartle's model that the killer and socialiser axes are sparsely populated with medium to low scores. Appendix G2 also plots the data for the individual games, but using ACE2 instead. We see that the achiever, explorer, and engager axes are densely populated with high scores for the ACE2 model. While the creator axis is sparsely populated, a select number of values scores quite high, which suggests that for the games in which the *creator* aspect was relevant, it was highly relevant according to the participants. This results suggest that the descriptive expressiveness of the ACE2 model substantially outperforms that of Bartle's taxonomy of player types.

5.2 Genre-specific analysis

Action games. Appendix H1 plots the averages of the data for action games for both models. We observe a slight difference between the two models. Overall, the Bartle killers are more relevant for action games than the ACE2 creators, but not significantly so; $p < 0.07$. However, when observing action games individually it becomes clear that ACE2 allows for a better abstraction of action games, since all three games share a similar profile. This is unlike Bartle's model, of which the results can be observed in Figure 4.

Adventure games. Both models generate unique profiles for adventure games, as is illustrated in Figure 5. However, Bartle's model only utilises two of the four axes, whereas ACE2 uses three. This allows for a higher degree of articulation in the categorisation of adventure games in ACE2.

Role-playing games. Like adventure games, role-playing games all have similar shapes and are thus close to their average for both

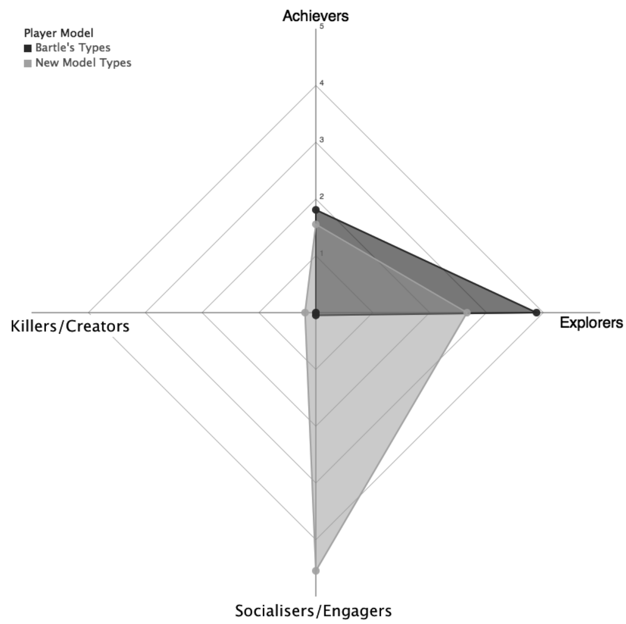

Figure 5: Adventure games in both models. Bartle's model utilizes only two axes, whereas ACE2 utilizes three, allowing for more articulation.

models. Bartle mainly utilises two of the axes, but it does not completely ignore the other two axes. The ACE2 model is again capable of showing more articulation by using three axes, but the creator axis is almost completely ignored, as is illustrated in Figure 6.

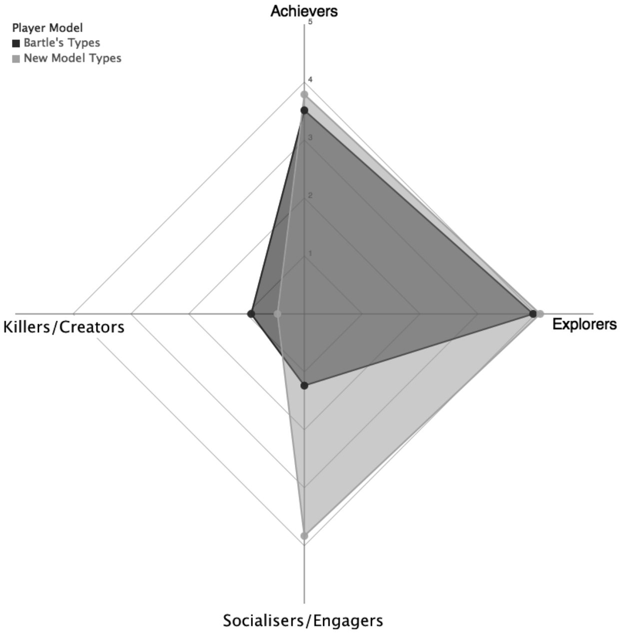

Figure 6: Role-playing games in both models. Bartle's model utilizes only two axes, whereas ACE2 utilizes three, allowing for more articulation.

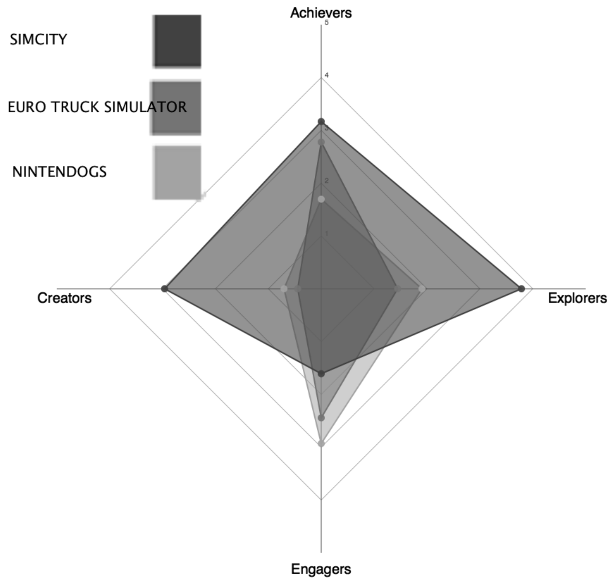

Figure 7: Individual simulation games in ACE2. Sim City clearly differs substantially from the other two games.

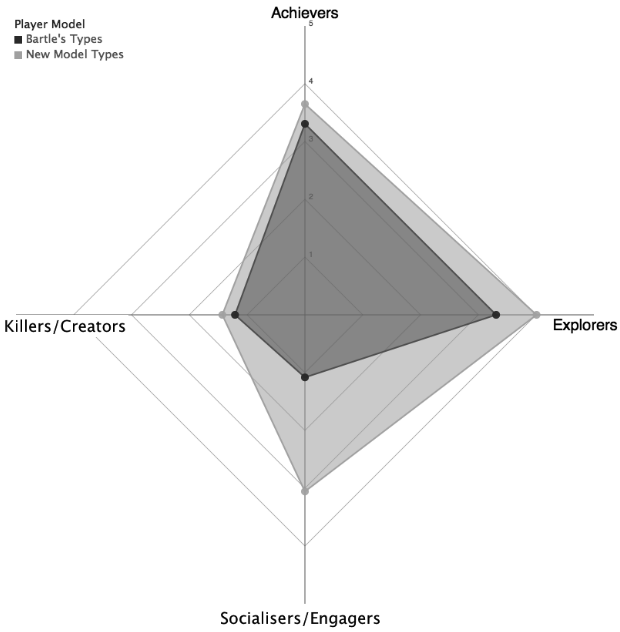

Figure 8: Strategy games in both models. Bartle's profiles is quite similar to the profiles for other genres.

Simulation games. Simulation games feature quite distinct shapes in both models, although all scores across both models are on the lower side. It seems that simulation games do not fit either model quite as well as other genres. When considering the games individually a clear outlier can be observed in *Sim City* in the ACE2 model. This is illustrated in Figure 7. We will investigate this in more detail in the discussion section.

Strategy games. In Bartle's model strategy games have a profile that is quite similar to other genre profiles, whereas ACE2 produces a more unique profile. Figure 8 shows both profiles, and when compared to Figures 3 and 6 it is clear that Bartle's model is not well suited for creating abstract genre profiles.

Sandbox games. Sandbox games generate distinct patterns in both models (Figure 9), making them easily identifiable. The Bartle model shows a little more variance in the individual games than ACE2. When looking at the creators axis in the individual games (Appendices H3-M3), we can see that with a single exception all high scores are in the sandbox genre. The one exception is in simulation games, where the city builder Sim City also scores high on the creators axis. The difference in scores on the creators axis for sandbox game and any other genre is significant, with an unpaired t test – and initial data indicating a Gaussian distribution – yielding a value of $p < 0.04$ for sandbox versus simulation, and $p < 0.003$ for sandbox versus other genres.

6 DISCUSSION

To conclude the present study, we feel it is important to discuss several limitations of the investigation, as they link to interesting future work (6.1), and wish to discuss several general observations that support the intuition that creation play is an important aspect in recent video games, and as such should by design be incorporated in player taxonomies.

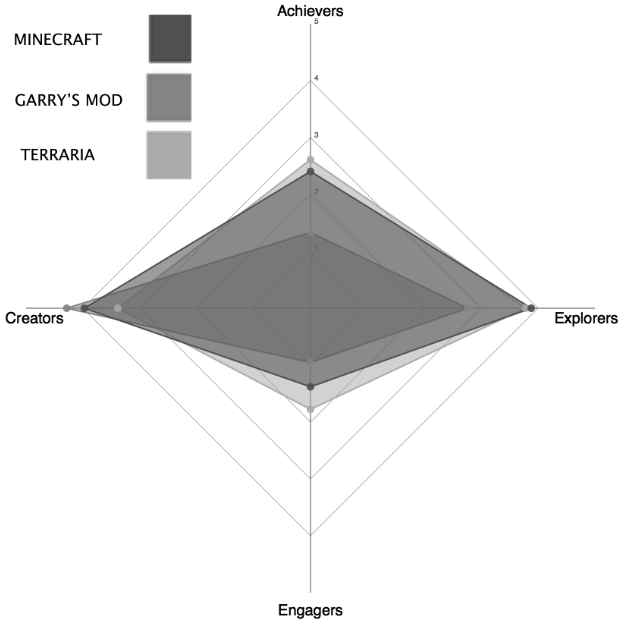

Figure 9: Sandbox games in ACE2. All three games score highly on the creators axis.

6.1 Limitations

In the present paper, Bartle's player taxonomy is purposely leveraged to provide a means for initial comparison of a revised model that intentionally – and by design – incorporates a 'creation play' aspect within its taxonomy. While Bartle's taxonomy of player types as a starting point indeed does not provide the basis that more recent models offer in terms of scientific embedding in personality theory (see Section 2.2), Bartle's taxonomy of player types however still provides a solid means for comparative analysis of conceptual revisions; the comparative analysis can thereby be focused not so much on model validation, but on what we are interested in foremost: the subjectively-experienced (creative) expressive range of video games as a factor of distinct player types.

We must also consider that Bartle himself proposed a so-called 'hacker' player type in his later work [6]. While to some degree this player type tinkers with available game mechanics, a hacker player does so purposely, as compared to play with no explicit purpose other than to build or create whatever the player desires.

Finally, we acknowledge that while gender information is not encapsulated in this preliminary work – which focused on obtaining an early understanding of creation play as an important facet of a game's expressive range – analysis along gender lines is certainly a point of interest in subsequent investigations.

6.2 General observations

When comparing the various axes with unpaired t-tests across various genres we found very little significant differences, even though by observing the graphs there seems to be a substantial difference. An explanation for this is that it seems not all participants understood that the questionnaire was focused exclusively on single-player games and thus still used Bartle's killers and socialisers, whilst one would assume that these play no role in single-player games. However, the creators type forms the exception to this, showing an overall highly significant ($p < 0.04$) difference between the sandbox genre and others. This supports our hypothesis that the creative player is a unique kind of player that should be considered separately from other player types.

Focusing on the creative type, we want to briefly reflect on the ACE2 outlier in the simulation games, *Sim City*. Due to their nature, simulation games will often borrow elements from other game genres in order to create the best simulation. In the case of Sim City, which is a city builder type game, it is no surprise that the creator player type is strongly represented whereas it is not in the other simulation games. The answers for the items pertaining to the creator type for *Sim City* differ significantly from those for the other two simulation games, *Euro Truck Simulator* and *Nintendogs*, with $p < 0.0133$. This strengthens our hypothesis that creative gameplay is worth considering separately even further.

Lastly, while none of the results were significant, we did find that the ACE2 model made it easier to differentiate between genres by eye. When looking at Figure 10, the three shapes in Bartle's model are nearly identical even though they belong to very different genres. This is more accurately reflected in Figure 11, where the three games feature have distinct shapes, allowing for simple and intuitive identification when observing the data. This shows us that while the differences might not appear substantial on the surface, the models do offer clear use in creating intuitive comparisons that can help people in finding similarities and differences between games.

7 CONCLUSION AND FUTURE WORK

This paper proposed a new taxonomy of video-game players: the *ACE2 model*. Building upon related work in player taxonomies, the ACE2 model sets out to refine the established Bartle's taxonomy of player types by incorporating the aspect of *creation play* (e.g., as exhibited in modern games such as *Minecraft*), thereby rendering the refined model more generally applicable to present-day video games. The paper considers the model part of an ongoing investigation into the relationship between aesthetics and mechanics in games. As such, the contribution of this paper lies not in proposing a definitive answer to taxonomic demarcation, it foremost attempts to highlight a creative play dimension that could be considered under-explored in classic player taxonomies. A model-validation method to this end, is to allow human participants to identify the subjective demarcation of creation play in a user study, in which the aesthetic / mechanics expressiveness of games is assessed by participants. The paper reported on the results of a first user study, set to obtain an early indication of the model's validity, prior to extensive validation experiments. These first studies that compare Bartle's model with the ACE2 model indicated that (1) the ACE2 model allows for a more articulate labelling of single-player video games, and that (2) even though creation play does not feature often, when it does it is a defining feature in modern games. In conclusion, it is suggested that (a) the descriptive expressiveness of the ACE2 model provides a substantial and functional refinement of Bartle's

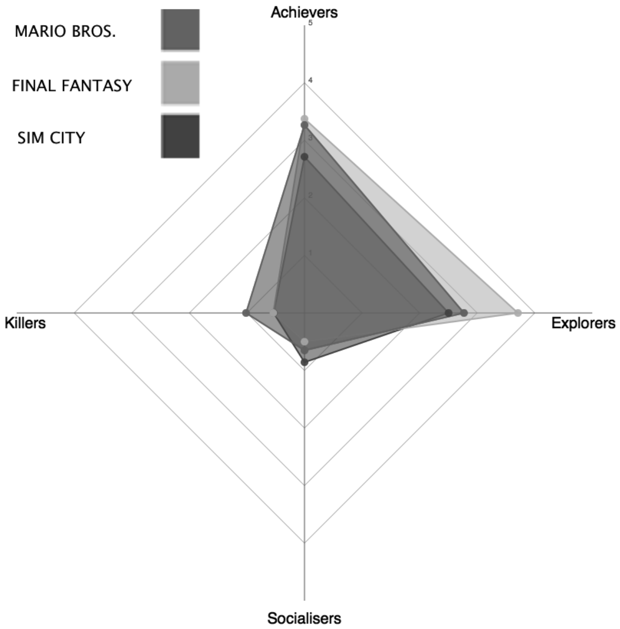

Figure 10: Super Mario Bros., Sim City, and Final Fantasy in Bartle's model. We observe that all three games score similar on the four axes despite belonging to three different genres.

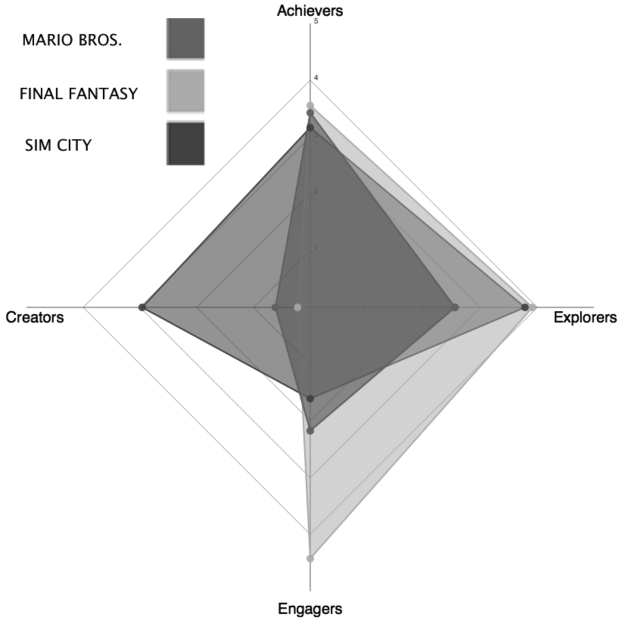

Figure 11: Super Mario Bros., Sim City, and Final Fantasy in the ACE2 model. We observe that the three games yield unique profiles on the four axes.

taxonomy of player types, and (b) further investigation of the interplay of aesthetics and mechanics – as experienced by game players – may yield important insight in (the taxonomic understanding of) creation play in games.

To do so, for future work, we will build upon the insights of the present paper, and will perform extensive validation experiments and data analysis that will draw correlates of creation play aspects of gaming, to how distinct players perceive the aesthetics / mechanic expressiveness of games – therein investigating the effect of player characteristics and personality, and game literacy.

REFERENCES

[1] Thomas Apperley. 2006. Genre and game studies: Toward a critical approach to video game genres. *Simulation & Gaming* 37, 1 (2006), 6–23.
[2] Elliott Avedon and Brian Sutton-Smith. 1971. *The study of games.*
[3] Sander Bakkes. 2010. Rapid adaptation of video game AI. (2010). TiCC Ph. D. Series 11.
[4] Sander Bakkes, Pieter Spronck, and Giel van Lankveld. 2012. Player behavioural modelling for video games. *Ent. Comp.* 3, 3 (2012), 71–79.
[5] Richard Bartle. 1996. Hearts, clubs, diamonds, spades: Players who suit MUDs. *Journal of MUD research* 1, 1 (1996), 19.
[6] Richard Bartle. 2003. A self of sense. *Multi User Entertainment [online]* (2003).
[7] Richard Bartle. 2012. Player Type Theory: Uses and Abuses. 2012. (2012).
[8] Richard Bartle. 2013. The decline of MMOs. (2013).
[9] Chris Bateman and Richard Boon. 2005. *21st Century Game Design.*
[10] Jonathan Blow. 2004. Game development: Harder than you think. *Queue* 1, 10 (2004), 28.
[11] Ian Bogost. 2007. *Persuasive games: The expressive power of videogames.*
[12] Roger Caillois and Meyer Barash. 1961. *Man, play, and games.*
[13] Alessandro Canossa. 2012. Give me a reason to dig: qualitative associations between player behavior in minecraft and life motives. In *FDG*. 282–283.
[14] Darryl Charles and Michaela Black. 2004. Dynamic player modeling: A framework for player-centered digital games. In *CGAIDE*. 29–35.
[15] Chris Fairclough, Michael Fagan, Brian Mac Namee, and Pádraig Cunningham. 2001. Research directions for AI in computer games. In *AICS*. 333–344.
[16] Johannes Fürnkranz. 2007. Recent advances in machine learning and game playing. *ÖGAI Journal* 26, 2 (2007), 19–28.
[17] Juho Hamari and Janne Tuunanen. 2014. Player types: A meta-synthesis. *Transactions of the Digital Games Research Association* 1, 2 (2014).
[18] Ryan Houlette. 2004. Player modeling for adaptive games. *AI game programming wisdom* 2, 2004 (2004), 557–566.
[19] Robin Hunicke, Marc LeBlanc, and Robert Zubek. 2004. MDA: A formal approach to game design and game research. In *AAAI Workshop on Challenges in Game AI.*
[20] Jesper Juul. 2011. *Half-real: Video games between real rules and fictional worlds.*
[21] David Keirsey. 1998. *Please understand me 2.*
[22] Raph Koster. 2013. *Theory of fun for game design.* O'Reilly Media, Inc.
[23] Ipsy Kyatric. 2013. *Bartle's Taxonomy of Player Types (And Why It Doesn't Apply to Everything).*
[24] Nicole Lazzaro. 2004. Why we play games: Four keys to more emotion without story. (2004).
[25] Bjarke Liboriussen. 2013. Craft, creativity, computer games: the fusion of play and material consciousness. *Philosophy & Technology* 26, 3 (2013), 273–282.
[26] Frans Mäyrä. 2008. *An introduction to game studies.* Sage.
[27] Matthijs Rohs. 2007. *Preference-based Player Modelling for Civilization IV.* Ph.D. Dissertation. B.Sc. thesis, Maastricht University.
[28] Mark Rosewater. 2002. Timmy, Johnny, and Spike. *Wizards of the Coast* 11 (2002).
[29] Katie Salen and Eric Zimmerman. 2004. *Rules of play: Game design fundamentals.*
[30] Jonathan Schaeffer. 2001. A gamut of games. *AI Magazine* 22, 3 (2001), 29.
[31] Chris Solarski. 2013. *Bartle's Taxonomy of Player Types (And Why It Doesn't Apply to Everything).*
[32] Pieter Spronck, Ida Sprinkhuizen-Kuyper, and Eric Postma. 2003. Online adaptation of game opponent AI in simulation and in practice. In *GAME-ON*. 93–100.
[33] Bart Stewart. 2011. Personality and play styles: A unified model. *Gamasutra, September* 1 (2011).
[34] David Thue, Vadim Bulitko, Marcia Spetch, and Eric Wasylishen. 2007. Interactive Storytelling: A Player Modelling Approach.. In *AIIDE*. 43–48.
[35] Jaap van den Herik, Jeroen Donkers, and Pieter Spronck. 2005. Opponent modelling and commercial games. In *CIG'05*. 15–25.
[36] Nick Webber and Matthew Wysocki. 2013. Controlling a Sandbox. *Ctrl-Alt-Play: Essays on Control in Video Gaming* (2013), 59–71.
[37] Nick Yee. 2006. The demographics, motivations, and derived experiences of users of massively multi-user online graphical environments. *Presence: Teleoperators and virtual environments* 15, 3 (2006), 309–329.

Implementing Drama Management for Improved Player Agency in Interactive Storytelling

Christopher McEvoy and David King
School of Design and Informatics
University of Abertay, Dundee
Bell St, Dundee
United Kingdom
E-mail: cmcevoy623@outlook.com

KEYWORDS
Drama Management, Intelligent Agent, Player Modelling, Heuristic Search Planning, Graph Theory.

ABSTRACT

With increases in complexity of graphics in video games, there exists a need to increase the complexity of game world narratives so that players feel they are an active part of an unfolding story, influenced by their actions and behaviours. Drama Management (DM) systems offer an attempt to facilitate this but are an area in need of further exploration for application in real-time narrative games. The aim of the project is to develop a prototype DM system for a real-time game that improves player agency and to analyse the effectiveness of the chosen techniques. An application was developed consisting of a 3D interactive environment, a possibility space of narrative plot points, and an Intelligent Agent that branches the story based on a Player Model, using Heuristic Search Planning. It was determined that the possibility space design has a major role in the application's effectiveness to invoke agency within players. The sense of agency can also be improved by combining the developed framework with additional extensions. This project determined that Drama Management systems are a viable method of improving the complexity of a narrative's discourse to promote player agency, but also require careful design alongside suitable algorithmic techniques to be fully effective.

INTRODUCTION

Game AI has been continually improving to match the complexity of computer graphics, to present players with believable and adaptive agents. In the context of narrative games, a similar advancement in the complexity of an interactive discourse, can improve the perception of player agency.

Agency is the degree of influence a player-character's actions have on the state of the game world. This can add to the enjoyment of games as players feel their choices matter.

Drama Management Systems are a computational approach to increasing narrative complexity to improve agency. DM systems use an Intelligent Agent (IA) that decides how to navigate a narrative possibility space defined by the designer.

Storytelling systems fall within two main classes: Emergent Narrative systems, and Drama Management systems. Where Emergent Narrative systems are simulations constructed from Intelligent Agents, Drama Management systems use a single Intelligent Agent, the Drama Manager, to monitor the game world and drive the authored story forward based on the player's actions (Reidl et al. 2012).

Individual approaches may borrow or combine traits from either class. A specific implementation may be considered to fall on a spectrum between prioritising player spontaneity and enforcing an authored narrative arc (Martens et al. 2017). This manifests as a trade-off between designer intent and player agency, which can become a problem. Compelling narratives can effectively balance these attributes.

Emergent Narrative systems are rich simulations underlying the game world that seek to generate unique stories through Intelligent Agents. Emergent Narrative systems have previously incorporated elements of DM systems in their design, creating a hybrid between the two system types. However, conventional DM systems are typically concerned with the structured ordering of authored story content rather than that of generated content. In such systems, a Drama Manager is used to perform the story planning. A Drama Manager is typically a single IA that makes decisions based on logical criteria and one or more additional metrics.

CURRENT TECHNIQUES

While DM systems vary in methodology, they all operate on similar principles to achieve different, but not dissimilar, goals. Events in a linear narrative can be called plot points, which are arranged sequentially. In non-linear games there can exist a state-space of plot points. DM systems use various techniques for constructing or ordering narrative arcs between plot points.

Many systems define plot points with two attributes: preconditions and effects. Preconditions describe what is required for the plot point to be accessible, and effects describe what changes are made to the story world by the plot point. These preconditions and effects connect plot points into an acyclic directed graph of nodes, and this format is used when planning the plot discourse.

DM systems navigate this directed graph but need at least one other module to inform the decision-making, where a module is a separate component that contributes to the overall system. Using information from its other modules, the DM can plan so that future problems can be avoided or to further enhance the story experience by selecting the most preferable path.

To select a plot point from a series of possibilities, many DM systems use a form of Heuristic Search Planning (HSP) to evaluate the best plot point, or trajectory, for a given situation. The heuristic can be calculated based on several attributes usually returned by another module to the DM.

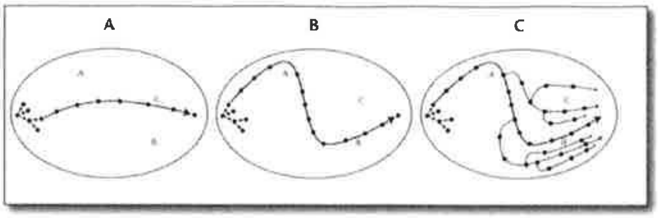

Figure 1: Visualisation of Trajectory Space as a Directed Graph

Figure 1 above (Roberts et al. 2011) is a visualization of trajectory space, as a directed graph, where each dot represents a plot point. There can exist a very large number of possible trajectories that, based on player interaction, are guided towards desirable plot points. Example A shows one possible path through the possibility space. Example B shows how a system can direct the path towards more favourable plot points and away from unfavourable points. Finally, example C shows how an interactive system with user input can lead to branches that navigate the space differently. HSP can be used to decide how favourable points are, and the plot can be branched using the heuristic and actions taken by the player.

The heuristic used to select plot points can be based on output from another module that communicates with the DM. Previous approaches have calculated the heuristic using a plot progression model, while other approaches use various forms of player modelling (PM). Player modelling has been used to model several different metrics through which a DM system can decide the possible future trajectories. DM systems incorporating player modelling attempt to model the player's state explicitly and shape the narrative specifically to influence it (Hernandez et al. 2014).

Player models have been based on a player's predicted emotional state. Others have been based on their predicted playstyle. Some approaches use the playstyle to infer a predicted player goal, using said predicted goal to predict the player's next set of actions, which is then used to change the state of the world in preparation for those actions.

For a playstyle prediction model, one method is to annotate player actions with a set of weights to several playstyle classes created by the designer (Weyhrauch et al. 1997). This annotation is to inform the system of what playstyle the player is likely to be using, to then infer predicted goals. The ordering of actions can then change the weights of the annotations, making a series of actions contribute more to one or more playstyles.

In essence, the way a player interacts with the world is used to determine what they are trying to accomplish, allowing the DM to consult the plot progression model and/or other metrics and use the result to decide what to do. The difficulty of this method is the balancing of the weights and player model algorithm to improve the accuracy of prediction. Additionally, the earlier predictions of player goals are more likely to be incorrect but improve in accuracy with a greater number of actions taken. This approach shares similar principles as reinforcement learning (RL), without the training results being retained.

Previous research and applications in Drama Management have involved the use of Adversarial Search and Partial-Order Causal Link (POCL) planning to structure a dramatic arc while avoiding problems (Roberts et al. 2011), Case Based Reasoning (CBR) to learn how interested participants are in different sections of the plot (Sharma et al. 2010), and player modelling for suspense (Reidl et al. 2011) to name a few. Additionally, natural language processing has been used in several applications with text-based interfaces. Most notably, *Façade* (Procedural Arts, 2005) allows players to communicate with the game's Intelligent Agent Non-Playable Characters (NPCs) using natural language through text.

Other approaches have used DM systems in conjunction with Intelligent Agent NPCs to add depth to an authored narrative, or to facilitate improvisation within a partially generated narrative space. One research example of such an approach is a Distributed Drama Management (DDM) system (Weallans, 2012) which seeks to retain the Emergent Narrative system benefits of believability and improvisation while still providing emotional and structural consistency.

METHODOLOGY

The created solution consists of a real-time 3D environment with limited methods of player interaction, a narrative possibility space represented by a directed acyclic graph and the DM system itself. The possibility space consists of plot points that have preconditions and effects.

Logical Structure

The Boolean condition of the precondition logic can be more complex than a simple 'AND' or 'OR' statement. There may be many conditions required and many logical operators. This project solves this problem by creating a method of defining condition logic, such that statements of any length and combination of ANDs and ORs can be specified. This project refers to the created solution as a 'dynamic Boolean matrix'.

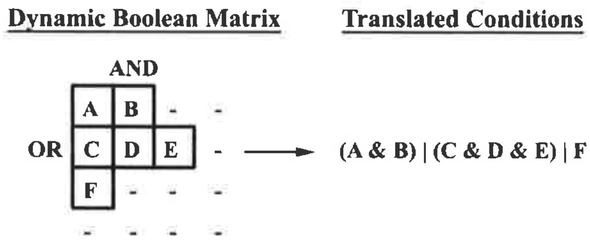

Figure 2: Visualisation of the Representation of Plot Point Precondition Logic

Figure 2 above shows that plot points in a row are ANDed together, and the result of each row is ORed with all other rows. The figure shows how this translates to a non-trivial Boolean statement. The matrix does not incorporate NOT operators as these are represented by the effects.

Player Modelling

The DM uses this logical structure to navigate the possibility space, and decides, out of those with met conditions, which plot points should be accessible in the world. This is determined by Heuristic Search Planning. The heuristic is a similarity measure between plot points and a Player Model, both of which have a set of weights. Plot points have a set of designer-annotated weights, used for HSP, and connected narrative text that is displayed when the point is fired. Once fired, plot point weights are used to update the Player Model and the weights of future points in the same branch as the fired point are also updated to improve their similarity to the new PM.

$$W_n = W_o + ((W_a - W_o) \cdot C_w + (W_a - m) \cdot C_m) \quad (1)$$

The above equation (1) adds a calculated value to the old PM weight to output an updated weight; where W_n is the new weight, W_o is the old weight, W_a is the adjusting weight, m is the constant mid value of 0.5, C_w is a weight constant and C_m is the mid line constant.

The increase value is generated using the difference between the old PM weight and the weight of the fired plot point, defined here as the 'adjusting weight', the difference between the adjusting weight and the 'mid value'. There is also a 'weight constant' and a 'mid line constant' which are used to determine the influence that these differences have on the incremental value.

The DM uses the possibility space structure and Player Model to decide which points to remove from the world, or prevent, and which points to place in the world. The weights of the PM, and of future plot points in the same branch, are updated after each point is fired.

Information Replacement

Information Replacement is the main feature of the Drama Management System. During this process, the IA evaluates what plot points are logically viable, with satisfied precondition logic, and then must decide which are preferred to put in the world. The IA puts the most preferable plot points in the world while removing the least preferable points. The Player Modeller's weight similarity calculation, the equation (2) below, is used as a heuristic for determining the preference of points.

$$S_b = W_p \cdot (1 - abs(W_p - W_c)) \quad (2)$$

This 'biased' similarity is calculated from the above equation (2); where S_b is the biased similarity, W_p is the player model weight, and W_c is the compared weight. The result is then subtracted from the maximum similarity to give a final distance heuristic. The similarity value is weighted, or biased, so that playstyles with a greater value will have more influence over the result. Weights have values between zero and one, but this biased similarity calculation will have a maximum result between these values, where the compared weights are equal to the player weights. The final similarity value is calculated by subtracting the biased similarity of a point's weights to the PM weights, from the maximum similarity value.

Using this as a heuristic, the DM decides which points not currently present in the environment, called the 'void-set', should replace points in the environment, called the 'world-set'. Points have potential 'spaces' they could be accessed from. During replacement, each space keeps a list of points trying to replace its current contents, called a competition list. When a point is replaced it is removed from the world set and added to the void-set where it can replace a less preferable world-set point.

Points are given a viability index based on their logical and PM preference. Points with a greater index can replace those with a lesser index.

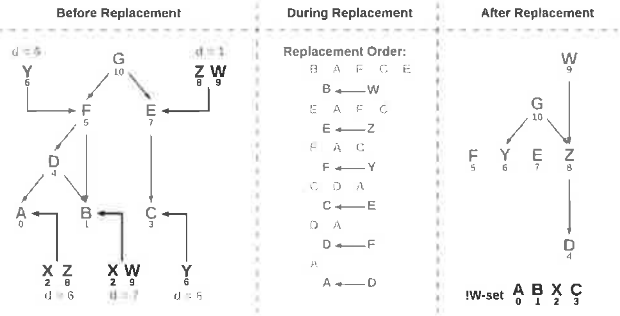

Figure 3: Example of a Round of Replacement

Figure 3 above shows an example of a round of replacement. The left tree represents the points in the world-set and how they could replace other points if they are themselves replaced. The middle column shows the order of replacement and steps through it. This order is determined by the difference in viability index between a space's top competitors. Once replacement is finished the tree becomes that shown on the right.

The values under the letters represent the viability index of competing points. The green values shown in the left tree represent the difference in viability index between the top two competing points. If only one competitor exists, the difference value is equal to the index of the lone competitor.

After each step of replacement, shown in the middle column, the competition lists are updated and new differences in viability indices are calculated. The replacement order is updated and the element space with the greatest difference value is replaced by its best competing point. This process is repeated until no more replacement can be done.

RESULTS

A possibility space containing 150 plot points was constructed, each with their own preconditions and effects, annotated weights, narrative text, and the actions/locations where the point can potentially be found. The dataset allowed for testing a variety of plot structures, with different constraints, so that the most effective design techniques can be found. Created branches led to 6 defined endpoints. Some branches were very specific in their player direction and possible locations, whereas others require more exploration, creating longer paths to the endpoint. 'Difference values' between the PM weights and the weights of the fired plot point was recorded throughout different playthroughs with different endpoints.

It was found that paths with more direction would have smaller PM difference values than those encouraging exploration. The explorative paths were typically longer and made progress towards more different endpoints than the directed paths, and consecutively fired points in the same branch less often than directed paths. Figure 4 below shows how the average PM difference value changes increases with path length. Points with more available locations-to-be-found-in were harder to find compared to those with fewer, but were more likely to be eventually found during longer paths, as the more location-specific points were more likely to be discarded by the DM.

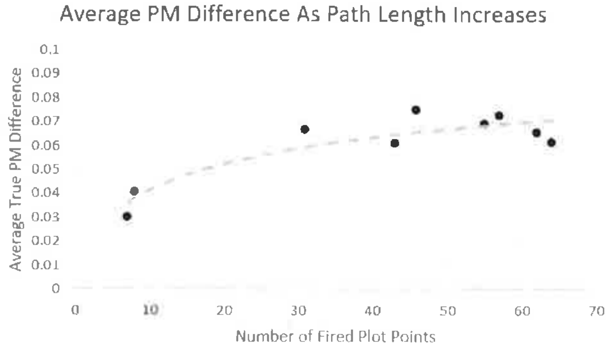

Figure 4: The Average PM Difference Value Increases with Path Length

DISCUSSION

It was determined that possibility space design has a large influence over the effectiveness of the system. Poorly-designed data will lead to poor performance. Exploring multiple separate branches granted improved player agency but made the overall story structure feel less focussed. It is suggested that games adopting this framework implement additional design constraints, simplify the location-specificity of plot points, and provide appropriate direction.

From the researcher's extensive playtesting of the system and its possibility space, this implementation was found to be effective at invoking player agency within well-constructed branches in real-time, yet improvements are needed to ready it for a commercial environment. Much time is needed to implement, debug and tweak the IA, but once the framework is created, it could speed up development of future similar projects. DM systems can also be combined with other AI techniques for specific requirements.

CONCLUSION

This style of Drama Management System offers an excellent supplement to conventional narrative structures, when well-designed and well-constrained, allowing for a great improvement to the variety of discourses available, and to player agency. However, limitations should be placed on the influence of this system over the overall narrative but can still be used to greatly improve the agency of a narrative's discourse. These solutions are also greatly scalable when the possibility space and narrative design are appropriately constrained, and small concessions are made in favour of speed over optimality.

REFERENCES

Hernandez, S.P.; V. Bulitko and E.S. Hilaire. 2014. "Emotion-Based Interactive Storytelling with Artificial Intelligence". *AIIDE.*

O'Neill, B. and M. Riedl. 2011. "Toward a Computational Framework of Suspense and Dramatic Arc" *Affective Computing and Intelligent Interaction* Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 246-255.

Procedural Arts (2005) *Façade* [Video Game]. Procedural Arts.

Riedl, M.O. and V. Bulitko. 2012a. "Interactive Narrative: A Novel Application of Artificial Intelligence for Computer Games", *AAAI.*

Riedl, M.O. and V. Bulitko. 2012b. "Interactive Narrative: An Intelligent Systems Approach" *AI Magazine* pp. 2013.

Roberts, D.; M.O. Riedl and C.L. Isbell. 2011. "Beyond Adversarial: The Case for Game AI as Storytelling".

Sharma, M.; S. Ontanon, M. Mehta and A. Ram. 2010. "Drama Management and Player Modeling for Interactive Fiction Games." *Computational Intelligence,* 26, pp. 183-211.

Weallans, A.; S. Louchart and R. Aylett. 2012. "Distributed Drama Management: Beyond Double Appraisal in Emergent Narrative". Springer, Berlin, Heidelberg, pp. 132.

Weyhrauch, P. and J. Bates. 1997. "Guiding Interactive Drama". Carnegie Mellon University Pittsburgh, PA.

WEB REFERENCES

Martens, C. and R. Cardona-Rivera. 2017. *Procedural Narrative Generation.* Available at: https://www.gdcvault.com/play/1024143/Procedural-Narrative (Accessed: Nov 11, 2018).

Serious and Entertaining

Edmund Carlo Louis van den Akker
Breda University of Applied Sciences
4817 JT Breda
The Netherlands
140520@ade-buas.nl

KEYWORDS

Gamification and Social Game Mechanics, Game Design, Serious Games and Gamification.

ABSTRACT

Entertainment in educational serious games is not the main purpose of these games. Research often suggests that entertainment plays an important role within educational serious games, while these educational serious games are often stigmatized for being boring. This research is done to better understand the effects of entertainment within these games. This effect was observed with participants playing different games and gauging their experiences. The results from this experiment yielded a result that entertainment does not influence the perception of learning in short term use cases of educational serious games. For long term uses future research would be required.

INTRODUCTION

Serious games are defined as games where the main purpose is beyond that of entertainment according to the Financial Times [1]. From this definition entertainment games are then defined as those games where the primary purpose is entertainment. This definition is derived from the article An Overview of Serious Games by Lamaarti, Eid, & El Saddik [2]. With educational serious games not primarily focusing on entertainment values, due to these games being serious games the following question arises:

- Does prioritizing entertainment over educational goals in educational serious games increase perceived learning?

To answer this question the currently existing research has been researched, which includes similar studies, game design methods, and cases of entertainment games which are educational. To fully answer this question research is done through a survey and interviews. Finally, the results are combined to give an answer to the research question.

EXISTING LITERATURE

To answer the research question, the background of educational serious games must be understood. This together in the context of entertainment games with educational content can show what is currently known in terms of the entertainment values in educational serious games. Therefore, literature regarding game design methods for entertainment games and educational serious games has been researched. Other literature regarding the effects of entertainment in educational serious games and entertainment in educational serious games has been researched as well.

Game Design Methods

Literature covering game design methods has been studied to better understand how both entertainment games and educational serious games work. Therefore, the most popular and applicable game design methods have been studied. The methods for educational serious games include methods such as LeGaDeE by Marfisi-Schottman [3], ABA as a design method explained by Kolić-Vehovec [4], and DPE by Winn [5]. The methods for entertainment games are MDA by Hunicke et al. [6], MTDA+N by Ralph & Monu [7], and Ubisoft Rational Design explained by McEntee [8].

The notable differences were that of educational serious games focusing heavily on the pedagogical nature of the games but with far less focus on the games themselves. An exception to this was the DPE model [5] which is based on the MDA [6] model as an extension. Another thing that was notable was the Ubisoft Rational Design model which covers the very fine details of game balancing and considers the flow of the game. Then this is further explained by the fact that this flow needs to be maintained by gradually increasing the in-game difficulty to accommodate for the player learning.

Entertainment in Educational Serious Games

Current existing research towards the importance of entertainment in educational serious games gives conflicting findings. Such is the case with the article Learning with serious games: Is fun playing the game a predictor of learning a success? by Iten & Petko [9]. In this study, tests were done that resulted in the answer that fun was not a predictor of learning with elementary school children. However, in the Triseum Game-Based Learning Validation Study Evaluation Report – 2018 by Tiede & Grafe [10], the results of several case studies resulted in successful learning results. Whilst this report does not go into the detailed effects of entertainment, it does hint towards entertainment playing a crucial role as motivator for the students involved in this project.

Educational Values in Entertainment Games

Whilst entertainment games focus primarily on entertainment first, there are cases of these games having educational elements. Cases like these are found in the article LEARNING ENGLISH THROUGH VIDEO GAMES by Vaisänen [11]. This article describes how boys who play more video games would get higher grades in their English class than people who did not. Another case of entertainment game being used in an educational setting is that of the article Age of Empires: “We used to get letters from schools” by Benson. This article covers how the developers of the mentioned game, Age of Empires would get letters from schools and that the students of these schools would learn from this game.

METHODOLOGY

With the background information known, the following methods have been executed to answer the research question of this research. The methods cover the analysis of the existing literature, the execution of mixed method interviews, and the release of a survey. The results of each method have been analyzed, and then combined with the results of the other methods. These results of all the methods combined have then formed the answer for the research question. However, due to the scope of the project the answer would only be valid for short term uses.

Mixed Method Interviews

The mixed method interviews cover the process of gathering participants and interviewing these participants. There is a survey section present in this interview which returns quantitative answers with room for qualitative quotes and feedback. The target audience for these mixed method interviews cover digital media students from Breda University of Applied Sciences within the ages of 18 to 25 years old. These students may have a bias for entertainment but do offer more critical insight towards their experiences with the offered test. These tests the participants have done consist of playing three different educational serious games. These games have been pre-tested with a group of people outside of the targeted audience in order not to contaminate this targeted group. Therefore, the presented games have predetermined entertainment values and very similar educational values. The selected games consist of Answer The Question [13], Everyday Genius: Squarelogic [14], and Math Problem Challenge [15]. The participant has played these three games and rated each game after they have finished it. They have rated it on its entertainment and educational values. These values have then been analyzed to determine if a correlation exists between the entertainment values and the perceived learning values.

Quantitative Surveys

Quantitative surveys have been released in the shape of personally asking other people within the target audience to participate in this survey. The participants were also selected from digital media students from Breda University of Applied Sciences within the ages of 18 and 25 years old. This did mean that the target audience likely has a bias towards entertainment values. This survey asked the participants about what they deemed important in an educational serious game, and if they had played one in the past, what these experiences were. Questions relating to these past experiences asked the participants if they had enjoyed these games, and if they felt that they have learned something from these games. Whilst these questions were from past events, and therefore not reliable they did form a good baseline to compare results. Further, something worth noting was that personally asking people to participate in these surveys was not the originally intended method, but a backup method. This is due to the original method of digital distribution falling through.

DATA RESULTS

The resulting data from the methods yielded a similar answer across the board. All methods returned the answer that entertainment values do not correlate with the user's perception of learning in educational serious games in short term uses. The biggest point of contention was the existing literature hinting at the existence of this correlation existing in the first place. In the existing literature there was only one case of this correlation not existing which was in the article by Iten & Petko [9]. However, the methodology in this article raised questions to the validity of the claim. This was due to the test being done with a game that could not be considered a game in the first place. This game only tutorialized web browsing, with somewhat pleasant aesthetics. There is however a similarity between the study mentioned in the article by Iten & Petko [9] and this research. Both projects focused on short term effects. However, the existing literature that does suggest that this correlation exists seems to suggest it exist in long term uses. These results did reflect a similar answer to both the conducted surveys, and mixed method interviews

Data Results Mixed Method Interviews

The data that has been gathered from the mixed method interviews yields a result that indicates there is no correlation between entertainment and perceived learning.

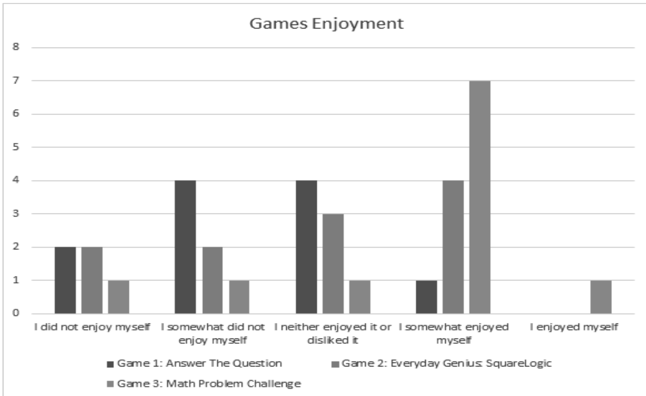

Figure 1, Enjoyment ratings of the tested games

In figure 1 can be observed that the third game, Math Problem Challenge [15] was rated the most enjoyable together with the second game, Everyday Genius SquareLogic [14].

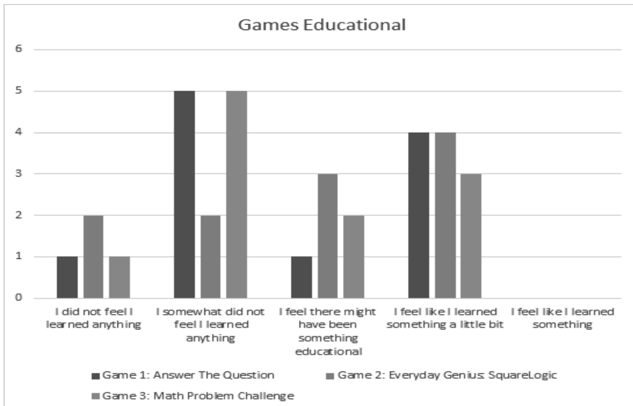

Figure 2, Perceived learning ratings of the tested games

Within figure two it can be observed that the amount people perceived to have learned varies greatly. There is no real indication between which game did better or worse.

Correlations

			Game 1 Enjoyement	Game 1 Education
Kendall's tau_b	Game 1 Enjoyement	Correlation Coefficient	1.000	.272
		Sig. (2-tailed)	.	.326
		N	11	11
	Game 1 Education	Correlation Coefficient	.272	1.000
		Sig. (2-tailed)	.326	.
		N	11	11

Figure 3, Correlation of the first game

The correlation test in figure 3 shows if there is an existence of a correlation between the enjoyment and perception of learning in the First game, namely Answer The Question [13]. As this figure shows with a correlation coefficient of 0.272 and a significance factor of 0.326 there is no existence of a significant correlation nor a strong correlation.

Correlations

			Game 2 Enjoyment	Game 2 Education
Kendall's tau_b	Game 2 Enjoyment	Correlation Coefficient	1.000	.465
		Sig. (2-tailed)	.	.088
		N	11	11
	Game 2 Education	Correlation Coefficient	.465	1.000
		Sig. (2-tailed)	.088	.
		N	11	11

Figure 4, Correlation of the second game

In figure 4 there is a correlation test of the second game, Everyday Genius: SquareLogic [14]. As shown in figure 4 with a correlation coefficient of 0.465 and significance of 0.088 the correlation was not strong nor significant.

Correlations

			Game 3 Enjoyment	Game 3 Education
Kendall's tau_b	Game 3 Enjoyment	Correlation Coefficient	1.000	.361
		Sig. (2-tailed)	.	.187
		N	11	11
	Game 3 Education	Correlation Coefficient	.361	1.000
		Sig. (2-tailed)	.187	.
		N	11	11

Figure 5, Correlation of the third game

In figure 5 a correlation test has been done for the third game, Math Problem Challenge [15]. As shown in figure 5 with a correlation coefficient of 0.361 and a significance of 0.187 there is no sign of a strong nor significant correlation.

The results from all games returned without any proof of a strong or significant correlation between entertainment values, and perceived learning. These values as shown in figures 3, 4, and 5 all show that in short term use entertainment does not influence the perception of learning in any significant way.

Data Results Quantitative Surveys

The results from the quantitative surveys yielded a very similar result to those results found in the mixed method interviews. In this survey the participant was asked to rate their past experiences with educational serious games. After this, they were also asked to state their opinion if they believed entertainment was positive or negative to the learning experience. After this data was collected it was checked for correlations. The question involving the participant's opinion then formed a basis to establish their perceptions of these games.

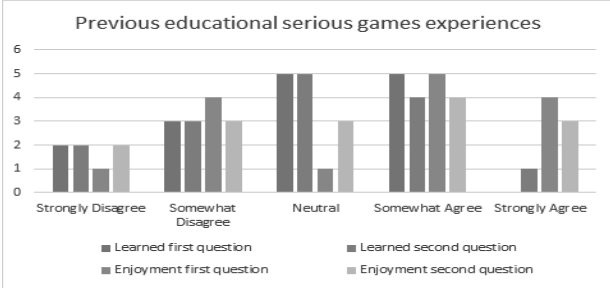

Figure 6, Learned and Enjoyment in past educational serious game experiences

In figure 6 the participant was asked at the start of the survey and as a check at the end of the survey if they had learned something and if they had enjoyed this game. This question was then later checked for accuracy through a correlation test, which resulted in it being accurate.

			I have enjoyed myself when I played this game	I have learned something from this game
Kendall's tau_b	I have enjoyed myself when I played this game	Correlation Coefficient	1.000	.354
		Sig. (2-tailed)	.	.117
		N	15	15
	I have learned something from this game	Correlation Coefficient	.354	1.000
		Sig. (2-tailed)	.117	.
		N	15	15

Figure 7, Correlation between learning and enjoyment

Figure 7 shows if there is a correlation between every individual's responses if they had learned something from the game, and if they had enjoyed themselves. As it turned out with a correlation coefficient of 0.354 and a significance value of 0.117 there was no evidence to suggest the existence of a correlation.

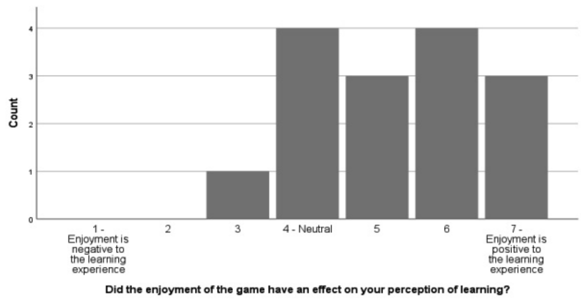

Figure 8, Target audience perception to the research question

Figure 8 shows the opinions of the participants on the importance of enjoyment in educational serious games.

Like the results of the mixed method interviews, the quantitative surveys returned the same answer. There is no evidence that a correlation between entertainment and perceived learning exists in short term uses as is shown in figure 7. However, people do believe that enjoyment in an educational serious game would be positive towards the learning experience as is shown in figure 8. This may imply that enjoyment might be important in long term uses, but not in short term uses.

CONCLUSION

As shown through the research methods there is no evidence that there is a strong or significant correlation between entertainment and perceived learning in short term uses of educational serious games. This means that entertainment would not play a significant role within short term use educational serious games. This would mean that a developer of such an educational serious game only has to design for the teaching aspects. However, since these results have only been found in short term uses, this does not mean that this result applies to long term uses. Since existing literature and the people's perceptions suggest, see figure 8, enjoyment may still play an important role for long term use. This means that current educational serious game design methods are fine as they are. The current methods cover teaching elements in detail and are seemingly fine for short term educational serious games. For long term educational serious games, it is currently speculated that entertainment might be more important and therefore the design methods could benefit from entertainment models.

To answer the research question, does prioritizing entertainment over educational goals in educational serious games increase perceived learning? The answer to this, at least in short use cases is no, it does not increase perceived learning.

For future research the effects of entertainment on the perception of learning in long term use cases is a good follow up to give a definitive answer. Further research could also involve the effects of player or student motivation of a subject and their motivations within an educational serious game, and what kind of effect this has on the learning experience.

ACKNOWLEDGEMENTS

I thank everyone who helped me during this project. I thank the staff of Breda University of Applied Sciences, and especially my project supervisor Thomas Buijtenweg for his guidance and insights. Furthermore, I would like to thank also the people who participated in the research as participants.

BIOGRAPHY

I, **Thomas (Edmund Carlo Louis) van den Akker**, have studied and graduated from NHTV with a bachelor's degree as a technical designer, and more recently graduated from Breda University of Applied Sciences with a master's degree in game technologies. Currently I am part time employed at ING as a junior technical game designer. I have previously whilst living in the United States of America gotten awards on a literature project, and I have been an honor student whilst attending college.
Current interests currently lie with game technology, such as game design, game scripting, and have incredible respect for game production and game artists. I would further like to improve my understanding in these fields of game technology as I continue to work in this industry.

REFERENCES

1. Financial Times. (n.d.). Serious Games Definition from Financial Times Lexicon. Retrieved October 3, 2018, from http://lexicon.ft.com/Term?term=serious-games

2. Lamaarti, F., Eid, M., & El Saddik, A. (2014, October 15). An Overview of Serious Games. Retrieved October 3, 2018, from https://www.hindawi.com/journals/ijcgt/2014/358152/

3. Marfisi-Schottman, I. (2012, December 08). Legadee. In LEGADEE Presentation. Retrieved September 27, 2018, from http://www-lium.univ-lemans.fr/legadee/

4. Kolić-Vehovec, S. (2018, September 12). Lecture presented at Evidence-based development of serious games for the educational sector: Challenges and perspectives. in Belgium, Brussels.

5. Winn, B. M. (2008). The Design, Play, and Experience Framework [PDF]. East Lansing, MI. Retrieved September 29, 2018, from http://ksuweb.kennesaw.edu/~jprest20/cgdd4303/winn-dpe-chapter.pdf

6. Hunicke, R., Zubek, R., & Leblanc, M. (2004, January). MDA: A Formal Approach to Game Design and Game Research. Retrieved September 23, 2018, from https://www.researchgate.net/publication/228884866_MDA_A_Formal_Approach_to_Game_Design_and_Game_Research

7. Ralph, P., & Monu, K. (2014, June 11). A Working Theory of Game Design. Retrieved October 31, 2018, from http://www.firstpersonscholar.com/a-working-theory-of-game-design/

8. McEntee, C. (2012, March 27). Rational Design: The Core of Rayman Origins. Retrieved October 24, 2018, from https://www.gamasutra.com/view/feature/167214/rational_design_the_core_of_.php?page=1

9. Iten, N., & Petko, D. (2014, November 11). Learning with serious games: Is fun playing the game a predictor of learning success? [PDF]. Retrieved October 9, 2018, from https://doi.org/10.1111/bjet.12226 Copyright was changed 25 February 2016F

10. Tiede, J. and Grafe, S. (2018). Triseum Game-Based Learning Validation Study Evaluation Report - 2018, European Schoolnet, Brussels

11. Väisänen, A. (2018). LEARNING ENGLISH THROUGH VIDEO GAMES: Finnish Learners' Experiences (Unpublished master's thesis). University of Jyväskylä. Retrieved January 17, 2019, from https://jyx.jyu.fi/bitstream/handle/123456789/57392/URN:NBN:fi:jyu-201803221813.pdf?sequence=1&isAllowed=yF

12. Benson, J. (2014, July 1). Age of Empires designers: "We used to get letters from schools". Retrieved April 10, 2019, from https://www.pcgamesn.com/age-empires-designers-we-used-get-letters-schools

13. Penny McDonald. (2018). Answer The Question [Computer software]. Say Again Studio.

14. Everyday Genius: SquareLogic [Computer software]. (2009). TrueThought.

15. Red Dolphin Games (2018). Math Problem Challenge [Computer software]. VT Publishing

A Taxonomy for Achievements in Digital Games

Sofiia Yermolaieva and Joseph Alexander Brown
Artificial Intelligence in Games Development Lab, Innopolis University, Russia, 420500
email: s.yermolaieva@innopolis.ru and j.brown@innopolis.ru

KEYWORDS
Digital Games, Game Design, Achievements, Taxonomy, Clustering, Decision Tree

ABSTRACT

Achievements are used to encourage desired behaviour and actions of a person. In the development of a game, it is important to determine the achievement set, as achievements reward desirable behaviour of players. This study presents a twenty-one characteristic taxonomy developed to investigate the logic behind achievement unlocks. The goal of the achievement unlocks created to expand and motivate players to discover new aspects of the game. This taxonomy was applied to the one hundred top-selling games parsed from the Steam store, a major online retailer, yielding 6775 achievements from the games. Data analysis is done using decision trees, decision tables, and clustering algorithms to determine relations in the outcomes of players receiving achievements.

INTRODUCTION

An achievement in a video game is a reward for performing actions predefined by the game developers, for example rewarding the player with an improved weapon for killing enemies, making user's character stronger or increasing health level for collecting inventory items, etc. Bernard Suits states that "playing a game is the voluntary attempt to overcome unnecessary obstacles" Suits (2005). Achievements add additional unnecessary obstacles which can be voluntarily overcome, thus acting as minigames to play in and of themselves in the larger framework of a game. Bertrand Suits also states that "[t]o play a game is to attempt to achieve a specific state of affairs [prelusory goal], using only means permitted by rules [lusory means], where the rules prohibit the use of more efficient means in favour of less efficient means [constitutive rules], and where the rules are accepted just because they make possible such activity [lusory attitude]." Suits (2005). Thus, the rules which open a new horizon in the gameplay could be implemented as achievements unlock conditions. Achievements unlock specifies accomplishing all the rules and actions required to get the achievement/reward.

The achievement system is a technique to retain players in the game and motivate them to continue playing. Algarabel and Dasi (2001) define achievement as an acquisition, learning, and form of knowledge representation and competence of a person in a specific area designed for the promotion of an expert level. The achievement is not only a result of satisfying the game's defined conditions by the player, but it is also an assessment tool of educational abilities and personal skills. Jane McGonigal points out that "Real-time data and quantitative benchmarks are the reason why gamers get consistently better at virtually any game they play: their performance is consistently measured and reflected back to them, with advancing progress bars, points, levels, and achievements. It is easy for players to see exactly how and when they are making progress. This kind of instantaneous, positive feedback drives players to try harder and to succeed at more difficult challenges" McGonigal (2011). Achievement systems are used in games to stimulate players not only complete the game but also discover all its embedded mysteries and players are faced with even more challenges Hamari and Eranti (2011).

Hamari and Eranti Hamari and Eranti (2011) discover and define designing components of the achievement based on the analysis of nine games: they are Signifier, Completion logic, and Reward. Authors state that a complete logic of achievement consists of a trigger (an action or an event), game setting's requirements, game state's conditional requirements, and a multiplier (amount of previous three components repetition). Blair's study on achievements influence on game-based learning presents a taxonomy of achievement design features Blair (2011). Some of the taxonomy items are: Expected or Unexpected, Negative or Positive, Easy or Difficult, etc. The taxonomy outlines a description of general achievements characteristics and their usage in a game.

Achievements drive players to be motivated to play the game, influence overall popularity, and have an influence on the development of player's vital behavioural habits and skills by rewarding desirable behaviour.

ACHIEVEMENT TAXONOMY

Since achievements have become an indispensable part of gaming, the following study gives a taxonomy, classification, and prediction model for achievement completion. We provide:1) A taxonomy involving 21 characteristics of achievements, based upon a common pattern among 279 achievements extracted from the top three games on April 17th, 2016 from the Steam website Valve Corporation (2016) and then applied to 6775 achievements of the 100 top sellers games in order to verify it. 2) A model developed for the prediction of the percentage of completion of achievements through the decision tree algorithm. The description of the algorithm given in Section V.3) Classification of achievements using clustering techniques: K-means and expectation maximization. K-means based classification can help to understand players' preferences of games based on their achievements. This can help in the construction of accurate user demographics. The expectation-maximization algorithm created a clustering of achievements which shows achievements on different stages of the game. This classification could contribute to the prediction of the various stages of the game at which players loses their interest. In this way, achievement clustering can help to develop attractive games for the users in future.

Results of the analysis is a guide for game developers in general and especially achievement designers to construct achievements by selecting the needed criteria to reach the desired percentage of completion. The results of the study benefit designers to improve the quality of achievements, make games more balanced, by having both easy and hard rewards and help motivate players to keep playing. These outcomes are the starting point for the development of a semi-automatic system for generation of achievements for games. The system will require from the user (in the case game designer) to select what criteria out of 21 presented in the paper are needed and will predict what percentage of completion by players, the achievement is going to have. In future researches, the system could be taught by neural network algorithms to pre-generate achievements templates by given criteria.

Through the analysis of different achievements description, we subjectively identify a common logic behind achievements and picked out 21 characteristics. By the common logic behind achievement unlocks, we mean a set of objective rules and actions required to unlock the achievement. All of these characteristics are discernible without speculation on the part of the designer at the design phase.

The Taxonomy

The games achievement taxonomy consists of 21 binary criteria, defined as follows:

• Duration — requires a certain period for it to be achieved (seconds, minutes, days, seasons).

• Gameplay experience earning basis — requires level-ups or gathering of experience points, etc.

• Invoked Action — requires well defined player-invoked action which leads to the game event for it to be achieved. This would for example not include such achievements as "play the game for some time".

• Compulsory compliance — required to be achieved for further game progress.

• Reattempt — requires to have a condition failure, where there exists the possibility to start again (mission failure, etc.).

• Repetition — requires the conditions for obtainment appears several times during the whole gaming period.

• Multiplier — requires a predefined number of triggers (actions, events).

• Character dependency — requires a predefined character is picked.

• Game setting dependency — requires some specific settings of the game (mode, difficulty).

• Functional value unlock — requires usable in-game and influences game process (more powerful weapon, more sturdy armour, new skills available, etc.).

• Non-functional value unlock — requires the influence of a social rank of the player in-game and outside-game (new appearance of the character, badges, game cards, pictures for the avatar, etc.).

• Handicaps — requires a player to meet with criteria which would negatively affect a player's skill.

• Pity — requires a linked to poor performance by a player.

• Multiplayer — requires interaction with other players.

• Survival — requires the player to survive in certain conditions.

• Game completion dependency — requires the game completion to a certain percentage or level.

• Inventory dependency — requires the usage of certain inventory (weapon, transport, etc.) .

• Enemy dependency —requires some actions from the player concerning a specific enemy.

• Achievements dependency —- requires earning specific other achievements to be unlocked.

• Story-related unlocks — requires some actions from the player to be done based on the game's plot.

• Discovering (exploration) — requires the player to explore gameplay map or some of part of it.

DATA COLLECTION AND TAXONOMY STATISTICS

The input data for the study consists of one hundred of the highest selling games with in-game achievements from the Steam by the 15th of November 2016.

Table 1 shows three examples of the data used for the study. Table 2 provides an example of Counter-Strike: Global Offensive showing the structure of the spreadsheet with the achievements of each game. It consists of such components as achievement name, description, and percentage of completion by players.

A 21 characteristics taxonomy mapped to 6775 achievements, and each achievement was evaluated on the objective 21 criteria by 22 people.

Table 1: Example Games Data Organization filled in from Steam Valve Corporation (2016).

Name of the game	Date of publication	Genre	Publisher	Developer
Counter-Strike: Global Offensive	21 Aug, 2012	Action	Valve	Valve
Project Genom	12 Oct, 2016	Action, Indie, Massively Multiplayer, RPG, Early Access	NeuronHaze	NeuronHaze
Grand Theft Auto V	14 Apr, 2015	Action, Adventure	Rockstar Games	Rockstar North

Table 2: Example Achievements from Counter-Strike: Global Offensive Data Organization.

Achievement name	Description	Percentage of completion by players
Points in Your Favor	Inflict 2,500 total points of damage to enemies	75.30%
Shot With Their Pants Down	Kill an enemy while they are reloading	75.30%
Body Bagger	Kill 25 enemies	74.40%

To identify achievements complexity, we calculated several criteria an average, satisfied by the achievement. We assume that achievements with more elements of the taxonomy are more challenging to accomplish. The average achievement satisfies four out of 21 criteria.

We examined which criteria are standard and those that are important based on the rate of their appearance in 100 games, see Table 3.

Invoked Action appeared to be the most common criteria for the data set; it is a component of 74.19% achievements. Other criteria which appeared rather often are Reattempt with 47.1%, Multiplier with 38.94% and Enemy dependency with the efficiency of occurrence 26.96%.

Two rare criteria for achievements are Pity with 1.24% and Handicaps with the efficiency of occurrence 1.61%.

Table 3: Statistics on how often each criteria appears in achievements.

Criteria name	Occurrence(%)
Duration	6.19
Gameplay experience earning basis	15.13
Invoked Action	74.19
Compulsory compliance	10.84
Reattempt	47.1
Repetition	23.78
Multiplier	38.94
Character dependency	18.62
Game setting dependency	17.5
Functional value unlock	5.91
Non-functional value unlock	6.16
Handicaps	1.61
Pity	1.24
Multiplayer	15.36
Survival	10.41
Inventory dependency	20.03
Enemy dependency	26.96
Achievements dependency	4.52
Story-related unlock	12.73
Discovering(exploration)	11.89
Game completion dependency	15.55

Decision Trees

Decision tree algorithms use a divide-and-conquer approach to build a tree from a function of one or more attributes with values in the leaf nodes of a tree Witten and Frank. (2005). It is a graph representation of decisions and their consequences Rokach and Maimon (2014).

The classification is performed via a sequence of questions Duda et al. (2000). Starting from the top to the bottom, each node with the question concerns specific property or characteristic of the pattern, leaf-nodes shows possible values (classes).

The Waikato Environment for Knowledge Analysis 3 (Weka 3) is a software tool, which contains a set of machine learning techniques for solving data mining tasks Frank et al. (2016). This tool was used to build decision tree models using the J48 Machine Learning Group at the University of Waikato (2001c) algorithm, which generates pruned and unpruned trees.

In the research, it was used to build a model for a percentage of completion prediction based on criteria values. Each criterion treated as a binary question; Yes and No to the question "Do achievement have this criterion?". By training the tree on the data set and using percentages of completion as outcomes, we trained the tree to create a prediction for a percentage of completion for different sets of criteria.

Clustering

Clustering is an approach to group sets of data by some similar characteristics Anderberg (2014).

For clustering, Weka Frank et al. (2016) algorithms are used. SimpleKMeans Machine Learning Group at the University of Waikato (2001b) is an algorithm based on

Table 4: Example of achievements belonging to the clusters created by K-means algorithm.

Cluster	Achievement name	Description
Cluster 1	Kill One, Get One Spree.	Kill an enemy player who has just killed four of your teammates within 15 seconds.
	Brief Acquaintance.	Play a Co-op multiplayer game through to the end.
Cluster 2	Not Just a Pretty Face.	Make every ape in the world intelligent.
	Visiting Diggers.	Make your way to the digger's cave.

K-means clustering.
K-means algorithm works as follows : 1) Initialization of K cluster centers. Usually chosen K objects, farthest from each other, or K random objects. Each of the chosen objects assigned to the 1...K clusters, as initial centers. By default, K equals two in SimpleKMeans. 2) Assignment of each analyzed object to its closest cluster center, using the Hamming distance, the Euclidean distance (default for SimpleKMeans) or the Manhattan distance (centroids computed as the component-wise median Jain (2010)). 3) Recalculate each clusters center after the addition of every new object. Averaging the cluster objects and assign this result as the clusters center. 4) Repetition of steps 2 and 3 until there were no changes in clusters centers.
Expectation maximization Machine Learning Group at the University of Waikato (2001a) is another algorithm for data classification from Weka 3 tool Frank et al. (2016), was used for data analysis. It is expectation-maximization on the base of multi-label classifiers.
The algorithm considers a set of data to be observed (X), a set of missing values or unobserved (latent) data (Z) and unknown parameters vector (Θ). It uses likelihood function Larraaga et al. (2018):

$$L(\Theta; X; Z) = \rho(X, Z|\Theta)$$

.
Maximum likelihood estimate of the unknown parameters is calculated as shown in the function Larraaga et al. (2018):

$$L(\Theta; X) = \rho(X|\Theta) = \sum_{Z} \rho(X, Z|\Theta)$$

.
Expectation-maximization algorithm is the iterative execution of two steps:

1. Expectation step (E step): calculates values of the log likelihood function, with respect to the current parameter estimate Little and Rubin (1987).

$$Q(\Theta|\Theta^{(t)}) = E_{(Z|\Theta^{(t)})}[\ln L(\Theta; X, Z)]$$

2. Maximization step (M step): search for parameter, which maximizes the likelihood of the quantity below , taking into account expected estimates of the unknown variables Little and Rubin (1987).

$$\Theta^{(t+1)} = argmaxQ(\Theta|\Theta^{(t)})$$

Number of clusters in EM Little and Rubin (1987) determined by the cross-validation or manually by user input.
Cross-validation works as follows Machine Learning Group at the University of Waikato (2001b):
1) Several clusters initially set to 1.
2) Split training set to ten subsets.
3) Perform EM ten times on the subsets from Step 2.
4) Calculate the average log-likelihood for all ten results.
5) In the case of likelihood increasing, increase the number of clusters by 1. Return to Step 2.

Task
In the research, it was used to create a classification of achievement by grouped criteria and to find descriptive criteria for each group. Descriptive criteria are those that have the most influence on the group.

RESULTS OF ANALYSIS

Decision Tree

Input
For the experiment following training data sets were considered:
First, the Top three selling games on April 17, 2016, from the Steam website: Counter-Strike: Global Offensive, Dark Souls III and Grand Theft Auto V, with a total of 279 achievements.
Secondly, The top 100 sellers grouped by year of release. According to the table, thirty-four percent of all top 100 sellers released in 2016. Games were analyzed separately, by the season.
• 737 achievements in the Spring 2016 season.
• 206 achievements in the Summer 2016 season.
• 451 achievements in the Autumn 2016 season.
• 565 achievements in the Winter 2016 season.
Each achievement was assigned to one of five classes by the value of percentage of completion in increments of twenty percent.

Output
Top three selling games gave the highest number of correctly classified instances — 74.898%, so the model is the most precise out of all created for five datasets. Precision value of the model was obtained based on the training (top three selling games) and test sets (four other sets considered as input) split. The tree built on the

Table 5: Example of achievements belonging to the clusters created by EM algorithm.

Cluster	Achievement name	Description
Cluster 1	Clean Sweep	Kill the entire opposing team without any members of your team taking damage.
	Wanderer	Play the game for 5 hours.
Cluster 2	Participation Award	Kill an enemy within three seconds after they recover a dropped bomb.
	Genetic Challenger	Play a VS multiplayer game through to the end.
Cluster 3	Non Starter	Stop DarkWater from analysing the Necroa Virus.
	Hide and Seek	Prevent link with Chernobyl exclusion zone.
Cluster 4	Frontier Justice	Have your sentry kill the enemy that just killed you within 10 seconds.
	HE Grenade Expert.	Kill 100 enemies with the HE grenade.
Cluster 5	Developer	Spend 1 hour on the DEVELOPER level.
	Sightseeing	Play 60 minutes each on Stoneshill Hillside, Dark Forest and Battlegrounds.

data consists of thirty-nine nodes and 20 leaves. The nodes represent characteristics from the taxonomy and leaves represent the percentage of completion for the achievement with all the prior set of nodes in the path. For the data from the top three most selling games, it was decided to build a decision tree, but with a broader number of classes. New decision tree considers ten classes. Outputted model of the J48 algorithm is less precise, than a model with five classes, only 56.63% of correctly classified instances.

As the model with five classes appeared to be more precise than the model with 10 classes, leading to the conclusion that five classes are the optimal number for the division of the model.

The model with the top three selling games appeared to be the most precise and working on test data, as well as outputted the most descriptive decision tree. Given that these top 100 sellers are popular, it is most likely that they are played more often and are therefore more likely to yield clear statistics as to the difficulty of achievements.

Clustering

With the EM algorithm, Weka 3 clustered input data set into five groups, with 3069 (45%) achievements in the first group, 2092 (31%) in the second group, 862 (13%) in the third cluster, 155 (2%) in the fourth and 597 (9%) in the fifth cluster.

SimpleKMeans resulted in two clusters, with 5421 (80%) achievements belonging to the first cluster and 1354 (20%) achievements from the second cluster.

Descriptive criteria for each cluster were chosen from the clustering outputs of both of these methods. Descriptive criteria are those which occur more often in a cluster. By the K-means clustering results, the efficiency of *Multiplayer* occurrence is 45% in the first cluster and only 14% in the other. Hence, the first cluster better is likely to contain *Multiplayer* achievements, which are intrinsic for Multiplayer games.

In the K-means clustering results, Cluster 1 contains multiplayer achievements with a time limit, inventory dependency and few actions required for the achievement, which are repetitive during the game. The achievements from the second cluster depend on game completion and discovering its plot; they require from player to survive and earn experience. Example of achievements from these two clusters, shown in Table 4.

According to the descriptive criteria and examples of randomly chosen achievements from each cluster, we created the names for the cluster. Achievements from the first cluster could be described as Time-limited players collaboration, while achievements from the second cluster, described as Single-player obstacles.

K-means classification is shown on the Figure 1. The figure represents descriptive criteria for each cluster and common criteria which appear in both clusters. Common criteria are shown in the middle of the figure in the intersection of clusters, for example, Invoked Action, Reattempt, etc. The classification separates the achievements by the types of games from the player's collaboration point of view.

For the clusters created by the EM algorithm, examples of achievements, shown in Table 5. Taking into consideration the descriptive criteria of each cluster, and examples of achievements, we developed descriptive names for clusters.

The first cluster's criteria are related with starting and experiencing the game, so its name is "Experience the game". The second cluster includes multiplayer achievements, where team fights with enemies, so it is called "Face the game with allies". The third cluster includes achievements which are directly related with the plot of the game, location exploration and other achievements, so it is named as "Fathom the game". The fourth cluster unites achievements, in which player should use different inventory and play for different characters, means to explore insides of the game, that is why it is called "Explore the game." The last, fifth cluster is called "Beat the game", because it requires a player to be experienced in the game to change game's settings and explore it from the other perspectives, such as changing the character or difficulty setting.

EM classification is shown on the Figure 2. The figure illustrates descriptive criteria for each cluster, criteria

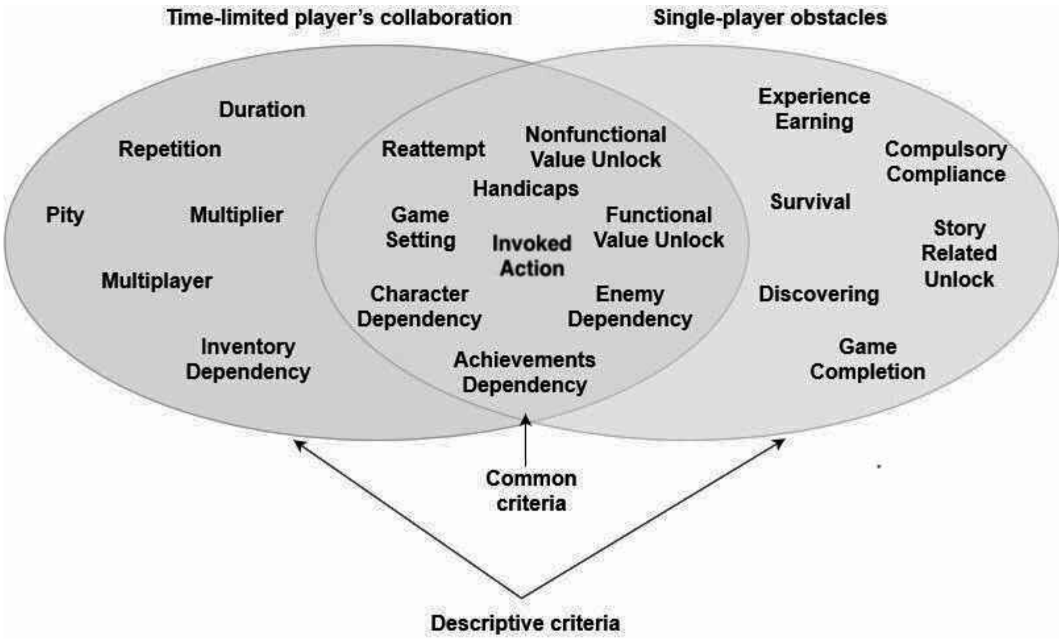

Figure 1: Classification of achievements using the K-means clustering method.

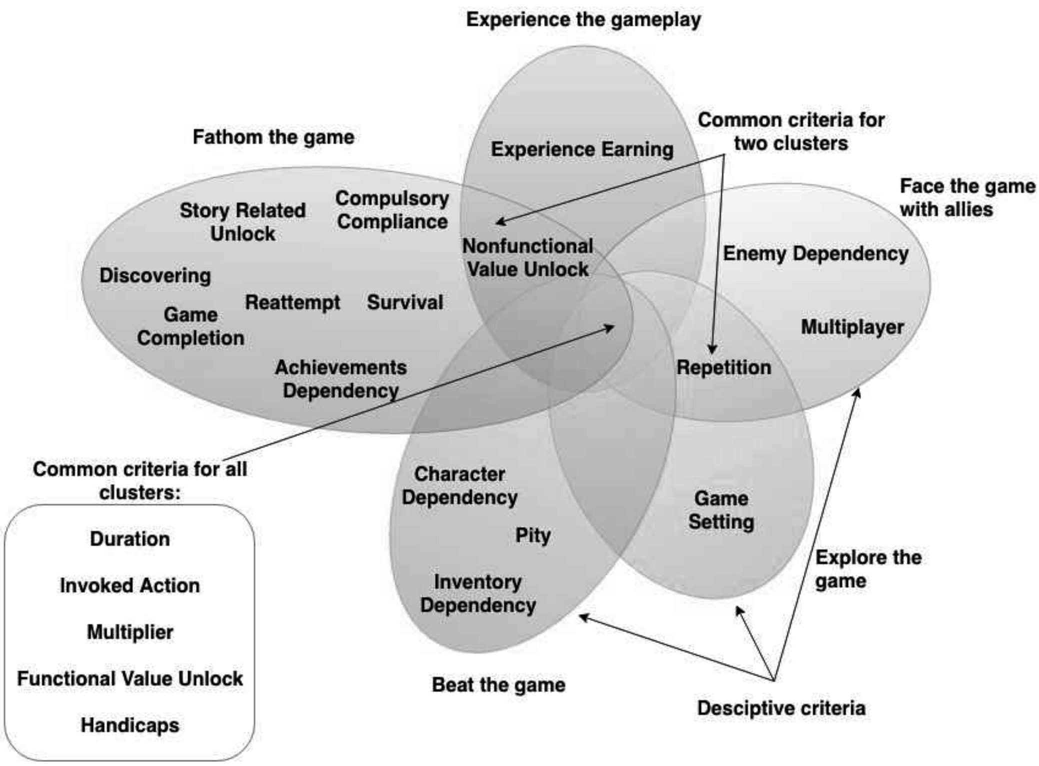

Figure 2: Classification of achievements using the EM clustering method.

which are common for two clusters (shown in the intersection of two clusters, for example Repetition) and those which are common for all clusters (shown in the intersection of all clusters, in the center of the figure, for example Duration, Invoked Action, Multiplier, etc.). This classification divides the unlocked achievements by the stage of the game.

CONCLUSIONS AND FUTURE WORK

This paper presents an analysis of achievements, conducted from one hundred top sellers' games from the Steam website. It is determined whether there is any dependency between achievement's characteristics and its percentage of completion. Twenty-one characterizing criteria were developed to perform the analysis.

Game designers can use the relation between sets of criteria and percentage of completion to create achievements which will lead to low completion rate as a result that will add an incentive for players to spend more time trying to achieve it. The time spent in a game is an essential factor for online games where players need to buy a subscription for some period.

In future work, user demographics according to their game preferences based on the classifications could be constructed. Further studies should include different analysis techniques, classification methods, and expanded data set for the possibility of the inclusion of other characteristics in the taxonomy.

The taxonomy can also be investigated from the game flow point of view, to determine which criteria are peculiar for specific actions. Further analysis can be done to determine which emotions an achievement forces player to feel and how these feelings are correlated with the flow, and those with the taxonomy. Determining the duration between the occurrence of different achievements can help to analyze taxonomy from another perspective. The time required for achievement completion should be considered as another critical completion criteria to find its dependency on the taxonomy. The use of more specific models of the clustering could also be examined, see Ashlock et al. (2010).

The results of the research are valuable for researches who study the psychological influence of video game achievements on players. The study will help to delve into the understanding of achievements logical content (with certain criteria) influence on the emotional state of the player who achieved it.

ACKNOWLEDGEMENTS

Thank you to the Valve Corporation, especially Gabe Newell and Dr. Mike Ambinder, for their answers to our questions about the Steam system. Thank you to Hamna Aslam and Elizabeth Reading for their comments. Thank you to the twenty taxonomy evaluators.

REFERENCES

Algarabel S. and Dasi C., 2001. *The definition of achievement and the construction of tests for its measurement: A review of the main trends. Psicológica*, 22, 43–66.

Anderberg M.R., 2014. *Cluster Analysis for Applications Probability and Mathematical Statistics: A Series of Monographs and Textbooks*. Academic Press.

Ashlock D.; Brown J.A.; and Corns S.M., 2010. *K-models Clustering, a Generalization of K-means Clustering. Intelligent Engineering Systems Through Artificial Neural Networks*, 20, 485–492.

Blair L., 2011. *The use of video game achievements to enhance player performance, self-efficacy, and motivation.*

Duda R.; Hart P.; and Stork D., 2000. *Pattern Classification*. Wiley-Interscience, second edition ed.

Frank E.; Hall M.A.; and Witten I.H., 2016. *The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques"*. Morgan Kaufmann, fourth edition ed.

Hamari J. and Eranti V., 2011. *Framework for Designing and Evaluating Game Achievements*. In *Proceedings of DiGRA 2011 Conference: Think Design Play.*

Jain A.K., 2010. *Data clustering: 50 years beyond K -means. Pattern recognition letters*, 31, 651–666.

Larraaga P.; Atienza D.; Diaz-Rozo J.; Ogbechie A.; Puerto-Santana C.E.; and Bielza C., 2018. *Industrial Applications of Machine Learning*. CRC Press.

Little R.J. and Rubin D.B., 1987. *Statistical Analysis with Missing Data*. In *Wiley Series in Probability and Mathematical Statistics*. 134–136.

Machine Learning Group at the University of Waikato, 2001a. *Class EM.*

Machine Learning Group at the University of Waikato, 2001b. *Class SimpleKMeans.*

Machine Learning Group at the University of Waikato, 2001c. *WEKA-J48.*

McGonigal J., 2011. *Reality is broken: why games make us better and how they can change the world*. Penguin Books, London, England.

Rokach L. and Maimon O.Z., 2014. *Data Mining With Decision Trees : Theory and Applications*. World Scientific Publishing Company.

Suits B., 2005. *The Grasshopper: Games, Life and Utopia*. Broadview Press, Ontario, Canada.

Valve Corporation, 2016. *The Steam. / Top sellers.* Online; accessed 17 October 2016.

Witten I.H. and Frank. E., 2005. *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, San Francisco, second edition ed.

UNDERSTANDING PLAYER ENGAGEMENT AND IN-GAME PURCHASING BEHAVIOR WITH ENSEMBLE LEARNING

Anna Guitart, Ana Fernández del Río and África Periáñez
Yokozuna Data, a Keywords Studio
Chiyoda 5F, Aoba No. 1 Bldg., Tokyo,
Japan
E-mail: {aguitart, afdelrio, aperianez}@yokozunadata.com

KEYWORDS
Churn Prediction, Ensemble Methods, Survival Analysis, Online Games, User Behavior

ABSTRACT

As video games attract more and more players, the major challenge for game studios is to retain them. We present a deep behavioral analysis of churn (game abandonment) and what we called "purchase churn" (the transition from paying to non-paying user). A series of churning behavior profiles are identified, which allows a classification of churners in terms of whether they eventually return to the game (false churners)—or start purchasing again (false purchase churners)—and their subsequent behavior. The impact of excluding some or all of these churners from the training sample is then explored in several churn and purchase churn prediction models. Our results suggest that discarding certain combinations of "zombies" (players whose activity is extremely sporadic) and false churners has a significant positive impact in all models considered.

INTRODUCTION

The concept of *churn* is as old as the customer–service relationships themselves. Churn occurs when a certain user stops using a service, i.e. when the relationship between the customer and the service provider ends (Mozer et al. 2000). This term is widely used in a variety of industries including retail banking (Mutanen et al. 2006), telecommunications (Hwang et al. 2004) and gaming (Runge et al. 2014, Periáñez et al. 2016).

Churn remains one of the most important metrics to evaluate a business, as it is directly linked to user loyalty (Hwang et al. 2004). High retention (i.e. low churn) points to a healthy business, and increases in user retention usually translate into higher revenues. In free-to-play games retention is crucial, since many of them have in-app purchases as their main source of revenue and, moreover, gaining new users through marketing and promotion campaigns is typically much costlier than retaining existing players (Fields 2014).

If there is a contractual relationship with the customer (as is normally the case in sectors such as telecommunications, see Mozer et al. 2000), the definition of churn is unambiguous: it happens when the customer cancels the contract or unsubscribes from the service. On the other hand, when there is no contract (or equivalent relationship) it is more difficult to assess whether a user has really churned or not. The appropriate way of defining churn in this kind of commercial activities must be carefully studied in light of their particularities and needs, and also of the purpose of the definition itself. This is the situation that applies to online games (Periáñez et al. 2016, Bertens et al. 2017, Kim et al. 2018, Chen et al. 2019), where most users stop playing without deleting their account. Additionally, free-to-play gamers who are active but make no purchases are of little or no economic value, and this allows us to introduce another type of churn definition within the video game context: *purchase churn*, which refers to paying users who cease to spend money on the title and is as tricky to define as conventional churn. (We will occasionally refer to the latter as *login churn*, for clarity.)

The usual strategy is to consider that a player has churned after a certain number of days of inactivity (Runge et al. 2014, Periáñez et al. 2016). Here we begin by examining how to choose a suitable (login/purchase) churn definition (in terms of days without activity/purchases). The goal of classifying players into active or churned is twofold: On the one hand, to have an accurate measure of the current health of the game. On the other, to label players in an appropriate way to successfully train churn prediction models.

Accurately predicting churn is of paramount importance for any business. In video games, the early detection of potential (login or purchase) churners may give studios the chance to target players individually—with personalized discounts, presents or contents—in an attempt to re-engage them. Previous works addressing churn prediction in video games have treated churn either as a classification (Sifa et al. 2015, Chen et al. 2019) or survival problem (Periáñez et al. 2016, Chen et al. 2019), with the latter approach being especially well-suited due to the censored nature of churn. Other related works used churn predictions to compute the lifetime value of individual players (Chen et al. 2018).

Player profiling (i.e. grouping users based on their behavior) is another noteworthy problem (Bauckhage et al. 2014, Drachen et al. 2012; 2014, Saas et al. 2016), which we also address here from a churn perspective. Our main goal is to characterize players who are identified as churners but eventually start playing again, namely *false churners*. Some of them are *genuine false churners*: in spite of meeting the corresponding churn definition, they never left the game, but just remained inactive for a relatively long time. Others (those who had a lengthier period of inactivity before returning to the game and thus can be considered to actually have churned) are more rightly regarded as *resurrected* players. In contrast, we will regard *all* players who start purchasing again after a prolonged lapse without spending any money as *purchase resurrected*. There is yet another group of interest in connection with churn: players whose activity is so sporadic that—regardless of whether or not they have been tagged as churners in the past based on the particular churn definition used—they can hardly be deemed as active users; we will refer to them as *zombies*. Such a classification of players according to their churn behavior is interesting on many levels, but in this work we focus on assessing its impact on the accuracy of churn prediction models.

The remainder of the paper is organized as follows: First we introduce the two main standard approaches used to define churn in video games, as well as the specific dataset and definitions adopted in our experiments. Then, we describe the churn prediction models analyzed in this study. Finally, after presenting and discussing the prediction results obtained by discarding different types of churners, we provide a brief summary of our findings and deliver our conclusions.

Our Contribution

To the best of our knowledge, this is the first work to simultaneously address login and purchase churn prediction, compare the classification and survival approaches and study the effect of excluding different kinds of churners from the training on the accuracy of the results.

DEFINITIONS AND DATASET

Defining Churn

Two main approaches to define churn in terms of player inactivity can be found in the literature:

1) Using a *fixed time window* for all players (Kim et al. 2018). For example, we could consider players who logged in during the previous month but not within the current one to be churners. This kind of strategy can be useful for some purposes—such as tracking game retention over long time scales—but it is not without shortcomings. In particular, it is fairly insensitive to specific player connection patterns, something especially problematic for churn prediction.

2) Using a *moving time window* for each player. To overcome the limitations of the above approach, most works measure the churn-defining inactivity period through a moving window, i.e., referred to individual player time instead of calendar time (Runge et al. 2014, Periáñez et al. 2016). While this method is computationally more demanding, it is also much better suited to model churn risk, and is thus the one followed in this paper.

The length of the optimal moving time window is highly game dependent. While in very casual titles a few days of inactivity typically signal a real user disengagement, in massively multiplayer online role-playing games time between sessions is usually much longer, and so longer time windows are required to correctly identify churners. The situation is analogous for purchase churn, as the typical purchase frequency may also vary greatly from game to game.

In this work, window lengths are selected so as to minimize two quantities: the *percentage of false churners* (number of churners who eventually return to the game over total number of churners) and the *percentage of missed sales* (sales from false churners after they return to the game over total sales). Considering long enough time windows can make both of these quantities vanish. However, our aim is to detect churn as soon as possible, both to have an accurate picture of player engagement at any given time and to have sufficient room for manoeuvre to try and re-engage potential churners. In particular, for our churn definitions we consider the shortest period of inactivity that keeps false churners under 5% and missed sales under 1% (although these figures can be fine-tuned according to the specific requirements of the analysis). Further details are given below.

Dataset

We used game data from the Japanese title *Age of Ishtaria* (a free-to-play, role-playing mobile card game developed by Silicon Studio), collected between 2014-10-02 and 2017-05-01. The data contains detailed daily information about each player, including level-ups, playtime, purchases and sessions. Only top spenders (*VIP players* or *whales*) were considered, as they are the most valuable users. We define VIP players as those with total outlay above a certain threshold (computed from the first two months of data so that whales provide at least 50% of the total revenue) and there were around 6000 of them in the studied dataset.

Data from other mobile games were also evaluated following the same methodology, and we obtained equivalent results, which shows the applicability of the proposed concepts to online games. These results are not included in the paper due to space limitations. In the case of non-online games, similar principles could be applied. However, as the purpose of this work is to give a solution that can be used in an operational environment, we focused on studying online games, where actions can be actively performed on the players and player information is continuously updated.

Age of Ishtaria's Churn Definition

Figure 1 shows graphically how the *login churn* definition was inferred from the first two months of data. The percentage of missed sales (left) and percentage of false churners (right) were evaluated for different churn definitions—letting the inactivity period after which a player is considered to have churned vary between 3 and 90 days—when considering all paying users (PUs, red curves) or just VIP players (blue curves). As already discussed, we require these percentages

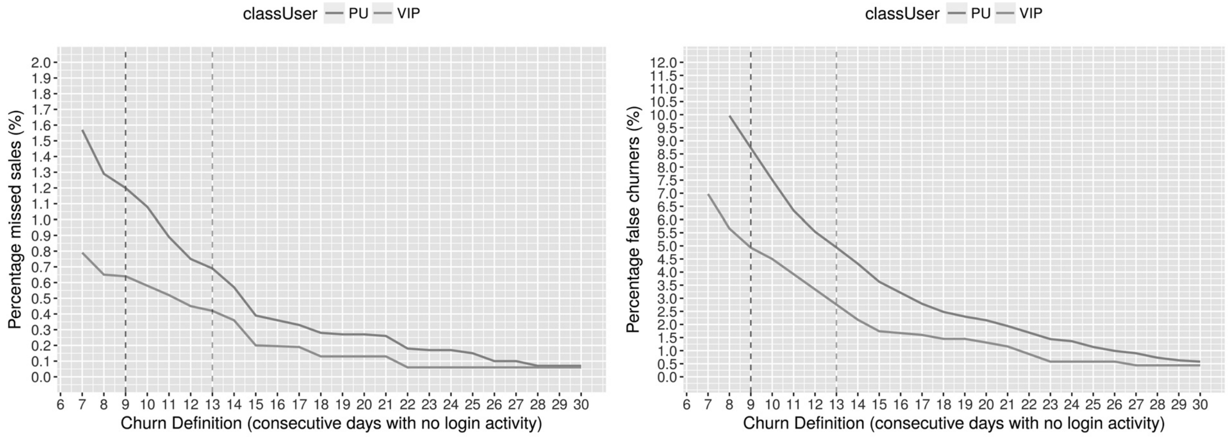

Figure 1: Determination of the login churn definition for VIP players (blue) and all paying users (PUs, red) based on two indicators: the percentages of missed sales (left) and false churners (right) during the first two months of data. By imposing these percentages to remain below 1% and 5%, respectively, we obtain 9 days (for VIP players) and 13 days (for PUs) as the inactivity period after which a player is considered to have churned.

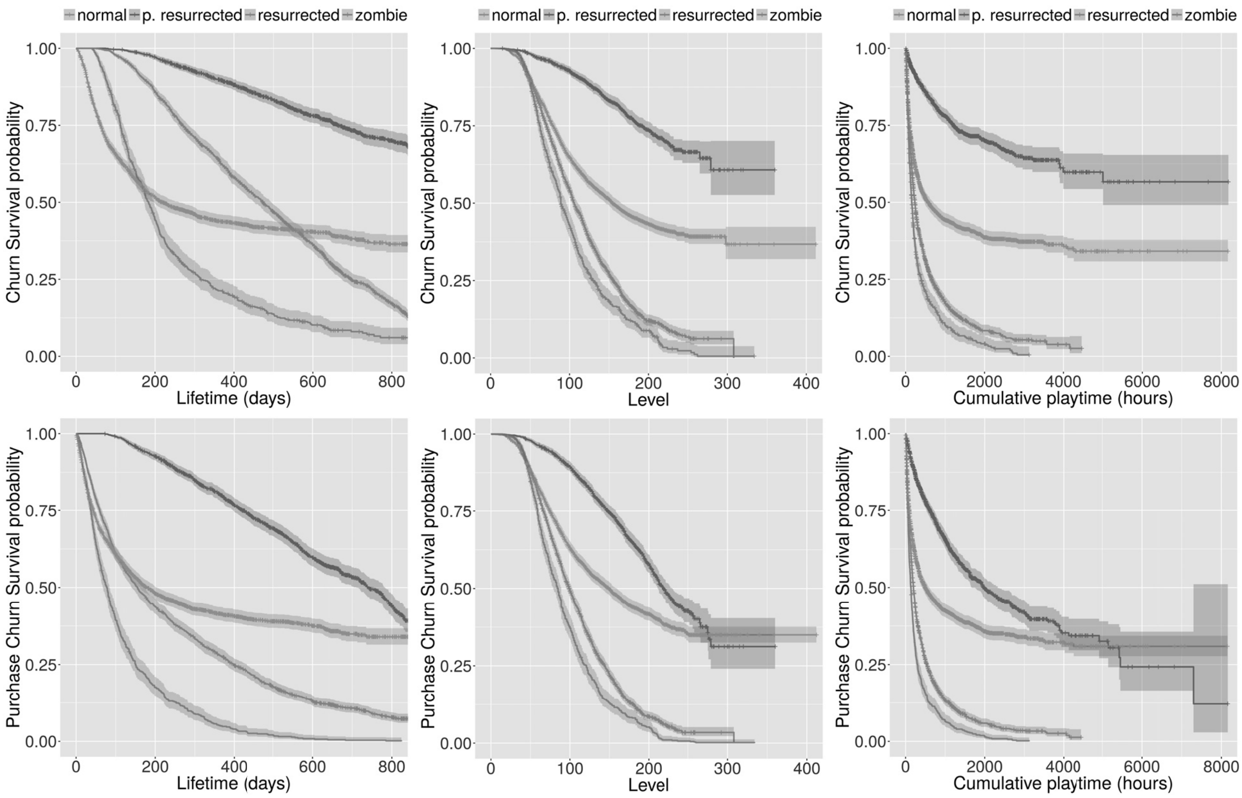

Figure 2: Cumulative survival probability (Kaplan–Meier estimates) as a function of time since first login (left), game level (center) and cumulative playtime (right) for VIP players. Top/bottom panels refer to login/purchase churn. Curves are stratified by churner type: *normal*, *zombie*, *resurrected* and *purchase resurrected* players. Shaded areas represent 95% confidence intervals.

to be less than 1% and 5%, respectively, which yields an inactivity period of 13 days for all PUs and 9 days for VIP players only. Since our analysis is restricted to top spenders, we will use the latter time window as our churn definition. Note that the percentage of false churners will tend to increase when considering extended data periods (longer than two months) but, for practical reasons, it is desirable to set the churn definition as soon as possible. In any case, we checked that such increase was not very significant—the percentage remained well below 10% even for longer windows of 6 months, taken at different dates across the full dataset—which means that the two-month data are representative of the overall churning behavior and supports our strategy. (Note that our real aim here is to restrict the number of *genuine* false churners. Thus the percentage of false churners can be higher when considering the whole dataset, due to the increase in the number of resurrected players.)

Following a similar approach we found that *purchase churn* should be defined as 50 days without any spending for VIP users. This inactivity period is much longer than in the previous (login) case, and thus a much larger (roughly by a factor 5) sample is needed to properly determine it. In practice, for a new title, it is possible to obtain a first working definition by using other games as reference, and then revisit it when a large enough data sample is available.

Churner Profiling

Three different groups of players with a particularly interesting churn-related behavior will be considered, and the impact of excluding them from the model training examined. These are

1) *Resurrected* players: Those who return to the game after churning and remaining inactive for a prolonged period of time. When churn is defined as less than 10 days of inactivity (as in our case), we require that period to be of at least 30 days. Users who return to the game before 30 days of inactivity are considered to be genuine false churners (i.e., to have been mistakenly marked as churners) rather than resurrected players.

2) *Purchase resurrected* players: In this study we identify all false purchase churners as purchase resurrected once they start spending again. (We thus disregard genuine false purchase churners.)

3) *Zombies*: Players who exhibit a too disengaged behavior to be considered active users (but who are not churners at that moment). In this study, players with less than 3 hours of playtime, no level-ups and no purchases in the previous 30 days were labeled as zombies.

Players who do not fall into any of the previous three groups will be referred to as *normal*.

In the sample considered, 21% of the players had churned and 5% had purchase churned by the end of the data period. Around 10% of all players were labeled as zombies, nearly 30% as resurrected and 23% as purchase resurrected at some point throughout their lifetime. Although the high percentage of resurrected players could suggest that our churn definition was not restrictive enough, we should recall that its aim is to limit the presence of *genuine* false churners rather than resurrected players (who typically churn for good shortly after returning to the game and thus do not increase the percentage of false churners in the long run).

Figure 2 shows Kaplan–Meier survival curves for VIP players—as a function of playtime, lifetime (time since first login) and game level—stratified by user type. Purchase resurrected players have the highest survival probabilities against both churn and purchase churn. (The only exception could be purchase survival for very high game levels or playtime, where normal players seemingly have higher probabilities, although it is not possible to ascertain that due to the large uncertainties.) On the other hand, zombies have the lowest survival and purchase survival probabilities across all variables. Interestingly, for small lifetime (though not level or playtime) values, resurrected players present higher survival rates (against login churn) than normal players. After more than a year the trend is inverted, as the survival probability for normal players stabilizes whereas that of resurrected players continues decreasing at the same pace. Note also that, in general, purchase survival curves are steeper than the corresponding (login) survival curves. This highlights the fact that all churners are also purchase churners (while the opposite is not true).

MODELING

We analyzed both binary and survival churn prediction models, exploring the effect of removing zombies, resurrected players, purchase resurrected players and combinations of them from the model training. The aim is to elucidate whether the presence of these players might be introducing noise that prevents the models from learning the *typical* VIP churn behavior more efficiently.

To get the results shown in this paper, data until 2018-03-01 was used for training and the remaining data until 2018-05-01, for validation. Nonetheless, we also evaluated the impact of varying the training and validation data ranges, obtaining similar results in all cases.

Model Specification

Specifically, we investigated the performance of a *conditional inference survival ensemble* model (Hothorn et al. 2006), described in detail in previous churn prediction studies (Periáñez et al. 2016, Bertens et al. 2017) of which the present work constitutes an extension. Player survival was described in terms of three different variables: playtime, lifetime (time since first login) and game level reached. On the other hand, binary classification was explored through *conditional inference trees* (Hothorn et al. 2006). Ensembles of size 1000 were used in all cases.

Feature Selection

Feature selection was also based on previous studies (Periáñez et al. 2016, Bertens et al. 2017) that constructed game-independent features measurable in most titles, such as playtime, purchases or number of actions of each player. We evaluated the best feature combination as a function of the model (binary or survival) and survival variable (lifetime, level, playtime). The possibility of adding a flag to identify the type of user (e.g. 1 for zombies and 0 for normal players) was also investigated. However, these variables proved

Table 1: Login and purchase churn prediction results for the binary and survival models, measured through the area under the curve (AUC) and the integrated Brier score (IBS), respectively. Survival results are given in terms of different predictor variables: lifetime, level and cumulative playtime. We consider different situations with regard to the training sample: including all users (*none*) vs. excluding zombie, resurrected or purchase resurrected players (or combinations of them). The best results for each model and variable are highlighted in bold.

CHURN	Binary models (AUC)		Survival models (IBS)					
excluding from training	by login	by purchase	by login			by purchase		
			lifetime	level	playtime	lifetime	level	playtime
none	0.95	0.69	0.072	0.069	0.060	0.070	0.080	0.077
zombie	0.93	0.69	0.034	0.047	**0.035**	0.055	0.067	0.086
resurrected	0.90	0.68	0.043	0.048	0.041	0.070	0.080	0.080
p. resurrected	0.95	0.72	0.104	0.084	0.060	0.065	0.076	0.062
zombie, resurrected	0.94	0.69	**0.029**	**0.041**	**0.035**	0.055	**0.057**	0.086
zombie, p. resurrected	0.93	0.72	0.057	0.068	0.049	**0.053**	0.067	**0.050**
resurrected, p. resurrected	0.92	0.73	0.071	0.068	0.057	0.065	0.068	0.057
zombie, resurrected, p.resurrected	0.94	0.73	0.053	0.059	0.050	**0.053**	**0.056**	**0.051**

to bias the models towards the behavior of the special users (affecting the accuracy of the predictions for normal players) and were discarded in the end.

Model Validation

For conditional inference ensembles, model validation was performed through specific survival analysis error curves and the integrated Brier score (IBS) (Graf et al. 1999), in the way described by Periáñez et al. (2016). The binary models performance was assessed using the area under the receiver operating characteristic curve (AUC); see e.g. Bradley (1997).

The set of players used for validation was the same in all cases (excluding zombies, resurrected and purchase resurrected players) so that we can fully assess the impact that training on different groups of users has on the predictions for the same group of players. This strategy was adopted to avoid massaging the data, which may lead to biased results.

RESULTS

The login and purchase churn prediction results for the different models (binary and survival) and survival variables (lifetime, level and playtime) are summarized in Table 1. Prediction error curves from the survival analysis of churn and purchase churn are shown in Figures 3 and 4, respectively. Both the table and the figures explore how the prediction accuracy of the models varies when we exclude one or several of the previously described player groups (zombies, resurrected, purchase resurrected) from the training sample.

The impact of including or excluding these groups is large in the survival analysis, but small to non-existent in the binary classification (where the only action that seems to have a relatively noticeable effect is removing purchase resurrected players when predicting purchase churn). This seems reasonable, as the former method relies on learning probabilities throughout the whole lifetime of each player and is thus much more sensitive to the noise introduced by erratic churn behaviors. Remarkably, the choices that optimize the survival results (discussed in detail in what follows) have a negligible to slightly positive impact on the binary models, and thus the same approach could be safely taken for both the classification and survival problems.

Focusing on the left column of Figure 3 (where only individual groups of players have been excluded from the training) we see that, for small lifetime, level and playtime values, the most significant error reduction in login churn predictions is achieved by removing zombies (although there is no such reduction for very short lifetimes), which is also reflected in the IBS scores in Table 1. The improvement is further enhanced as lifetime increases; for high playtime and level, however, the trend is reversed and errors are lower (albeit not significantly) when considering all players. Removal of (only) resurrected players exhibits similar patterns, but with a generally lower impact. Curiously, discarding purchase resurrected players has almost the opposite effect: it affects very negatively the performance for small values of all three survival variables, but improves it at large scales—to the point of yielding the best results for high level and playtime. However, the IBS values in Table 1 clearly indicate that the overall performance is degraded when removing these users.

As suggested by the previous discussion, the best overall results for login churn prediction are obtained by excluding both zombies and resurrected (but not purchase resurrected) players from the training sample; see Table 1. On the other hand, the overall negative impact of removing purchase churners can be deemed reasonable, which may be explained by the fact that these players—despite going for long periods without any spending—can maintain typical activity levels in terms of session frequency and duration and in-game progression, thus providing the models with additional valuable information to learn from.

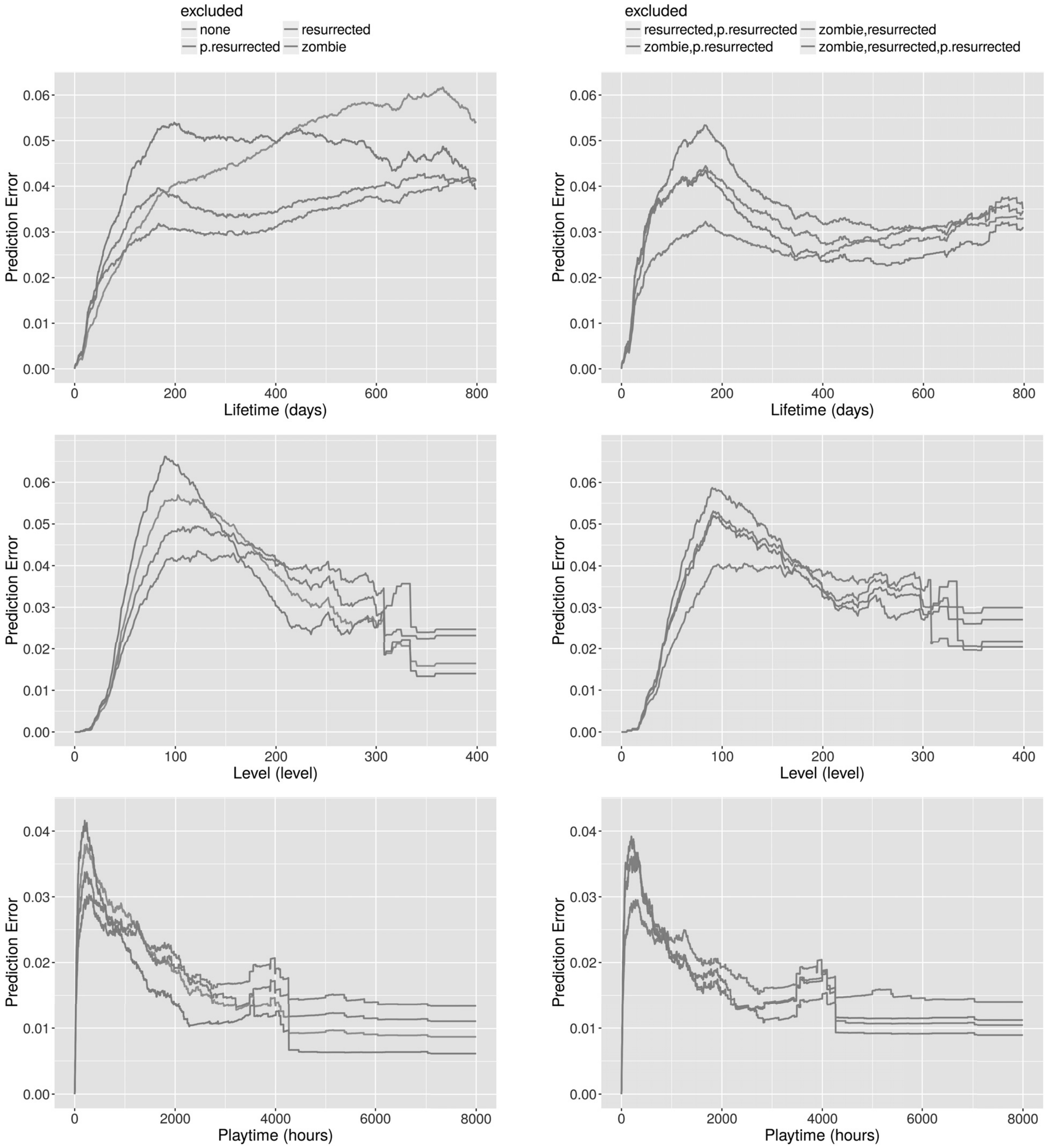

Figure 3: Prediction error curves for *login* churn as a function of lifetime (top), game level (center) and playtime (bottom). They have been computed using a conditional inference survival ensemble model, upon excluding zombie, resurrected or purchase resurrected players (left) and combinations thereof (right) from the training sample.

Turning now to purchase churn, the effects of excluding only zombies or only resurrected players (see Figure 4, left column) are qualitatively similar to the ones discussed for login churn. Purchase resurrected players could have been anticipated to play a major role in understanding purchase churn, and indeed their exclusion does provide an overall improvement in all variables (lifetime, level and playtime) as shown by the IBS values in Table 1. Interestingly, discarding these players has a negative impact for short lifetimes—an effect compensated for by the great gains at large values. This could be suggesting that a more restrictive definition of purchase resurrected players (by requiring them to start purchasing again after periods not just longer but *much longer* than the purchase churn definition, namely the approach followed for

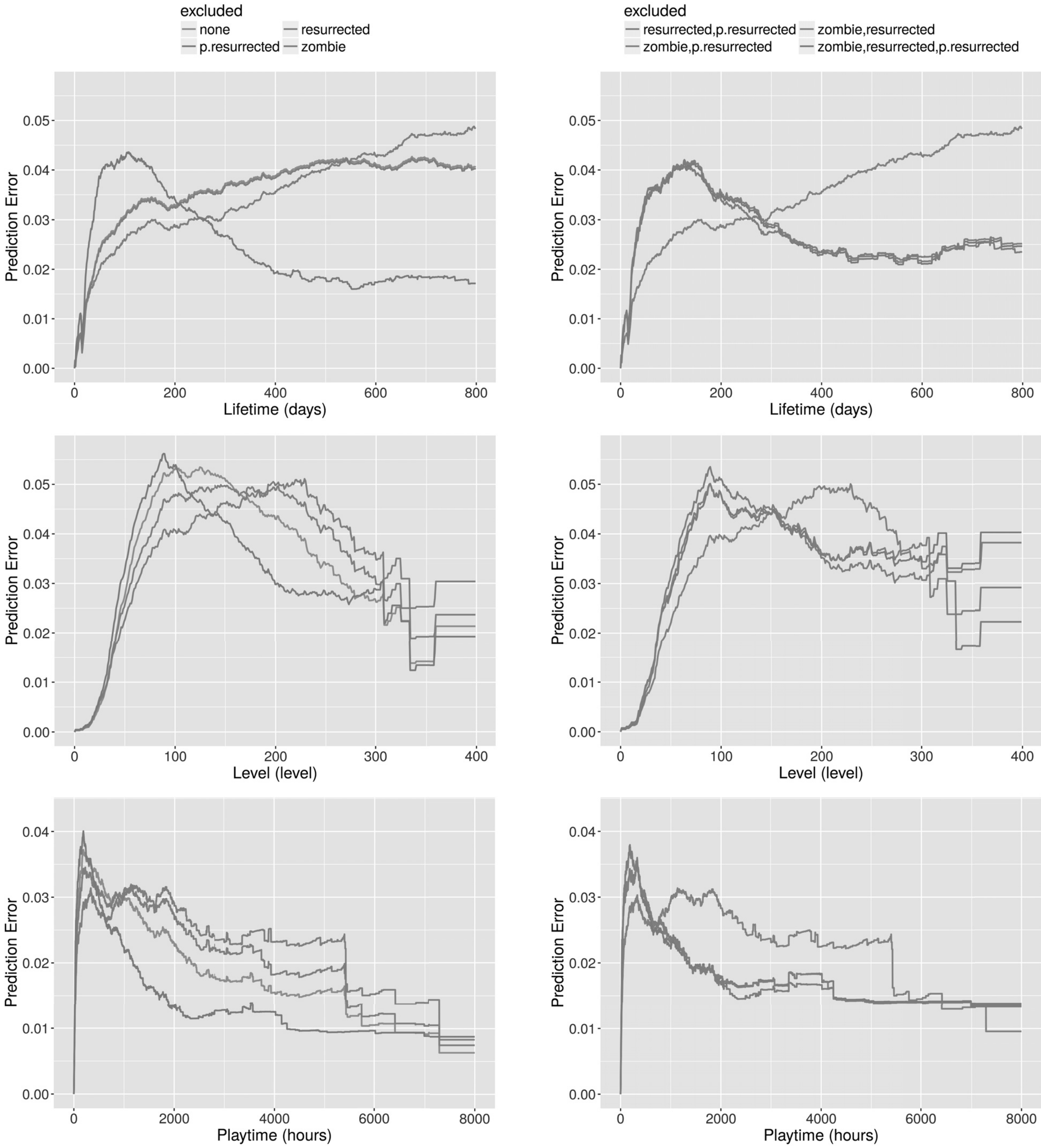

Figure 4: Prediction error curves for *purchase* churn as a function of lifetime (top), game level (center) and playtime (bottom). They have been computed using a conditional inference survival ensemble model, upon excluding zombies, resurrected or purchase resurrected players (left) and combinations thereof (right) from the training sample.

login churn) might be needed.

As in the case of login churn, excluding both resurrected and zombie players yields good results in terms of lifetime and level; however, for playtime it is better to consider all players. The highest overall accuracy is achieved by discarding zombies and purchase resurrected players (being almost irrelevant whether or not resurrected players are also discarded).

SUMMARY AND CONCLUSION

This study shows that excluding certain types of players (with a particular behavior regarding churn) from the training sample can lead to better churn predictions in the context of video games. Both binary classification and survival models were evaluated. Even though both approaches yield accurate results, the latter seems better suited for churn prediction, since (as discussed in Periáñez et al. 2016) it takes into account the censored nature of the problem and provides a much richer output. Our results show that, in general, removing active players with very limited activity (zombies) and those who return to the game or after a long period of inactivity (resurrected players) leads to more accurate churn and purchase churn forecasts. (In the latter case, optimal results are obtained by removing also players who start purchasing again after a long period without spending.) Moreover, excluding certain players from the modeling might be helpful from an operational perspective, as it would reduce the size of game datasets.

This work proposes three new types of players based on their churn behavior and aims to establish a basic framework for further related studies (in a similar vein e.g. to the already extended use of the "VIP player" concept). It also opens new questions in game data science research, such as whether it could be possible to foresee if a certain player will resurrect and how many times she will do so, or to get an accurate time-to-resurrection prediction. Finally, it represents a first step towards finding better and increasingly complex ways to characterize churn behavior that will improve our understanding of the phenomenon and the performance of churn prediction models.

ACKNOWLEDGEMENT

We thank Javier Grande for reviewing the manuscript.

REFERENCES

Bauckhage C.; Drachen A.; and Sifa R., 2014. *Clustering game behavior data*. *IEEE Transactions on Computational Intelligence and AI in Games*, 7, no. 3, 266–278.

Bertens P.; Guitart A.; and Periáñez Á., 2017. *Games and Big Data: A Scalable Multi-Dimensional Churn Prediction Model*. In *2017 IEEE Conference on Computational Intelligence and Games (CIG)*. IEEE, 33–36.

Bradley A.P., 1997. *The Use of the Area Under the ROC Curve in the Evaluation of Machine Learning Algorithms*. *Pattern Recogn*, 30, no. 7, 1145–1159.

Chen P.P.; Guitart A.; and Periáñez Á., 2019. *The Winning Solution to the IEEE CIG 2017 Game Data Mining Competition*. *Machine Learning Knowledge Extraction*, 1(1), 252–264.

Chen P.P.; Guitart A.; Periáñez Á.; and Fernández del Río A., 2018. *Customer Lifetime Value in Video Games Using Deep Learning and Parametric Models*. *IEEE International Conference on Big Data*, 2134–2140.

Drachen A.; Sifa R.; Bauckhage C.; and Thurau C., 2012. *Guns, swords and data: Clustering of player behavior in computer games in the wild*. In *2012 IEEE conference on Computational Intelligence and Games (CIG)*. IEEE, 163–170.

Drachen A.; Thurau C.; Sifa R.; and Bauckhage C., 2014. *A comparison of methods for player clustering via behavioral telemetry*. *arXiv preprint arXiv:14073950*.

Fields T., 2014. *Mobile and Social Game Design: Monetization Methods and Mechanics*. CRC Press, 2 ed.

Graf E.; Schmoor C.; Sauerbrei W.; and Schumacher M., 1999. *Assessment and comparison of prognostic classification schemes for survival data*. *Statistics in medicine*, 18, no. 17-18, 2529–2545.

Hothorn T.; Hornik K.; and Zeileis A., 2006. *Unbiased recursive partitioning: A conditional inference framework*. *Journal of Computational and Graphical Statistics*, 15, no. 3, 651–674.

Hwang H.; Jung T.; and Suh E., 2004. *An LTV model and customer segmentation based on customer value: a case study on the wireless telecommunication industry*. *Expert systems with applications*, 26, no. 2, 181–188.

Kim K.J.; Yoon D.; Jeon J.; Yang S.i.; Lee S.K.; Lee E.; Jang Y.; Kim D.W.; Chen P.P.; Guitart A.; Bertens P.; Periáñez Á.; Hadiji F.; MÃijller M.; Joo Y.; Lee J.; and Hwang I., 2018. *Game Data Mining Competition on Churn Prediction and Survival Analysis using Commercial Game Log Data*. *IEEE Transactions on Games*, 1–1.

Mozer M.C.; Wolniewicz R.H.; Grimes D.B.; Johnson E.; and Kaushansky H., 2000. *Churn reduction in the wireless industry*. In *Advances in Neural Information Processing Systems*. 935–941.

Mutanen T.; Ahola J.; and Nousiainen S., 2006. *Customer churn prediction-a case study in retail banking*. In *Proc. of ECML/PKDD Workshop on Practical Data Mining*. 13–19.

Periáñez Á.; Saas A.; Guitart A.; and Magne C., 2016. *Churn Prediction in Mobile Social Games: Towards a Complete Assessment Using Survival Ensembles*. In *2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*. IEEE, 564–573.

Runge J.; Gao P.; Garcin F.; and Faltings B., 2014. *Churn Prediction for High-value Players in Casual Social Games*. In *Computational Intelligence and Games (CIG), 2014 IEEE Conference on*. IEEE, 1–8.

Saas A.; Guitart A.; and Periáñez Á., 2016. *Discovering Playing Patterns: Time series clustering of free-to-play game data*. In *2016 IEEE Conference on Computational Intelligence and Games (CIG)*. IEEE, 1–8.

Sifa R.; Hadiji F.; Runge J.; Drachen A.; Kersting K.; and Bauckhage C., 2015. *Predicting purchase decisions in mobile free-to-play games*. In *Eleventh Artificial Intelligence and Interactive Digital Entertainment Conference*.

REAL
WORLD
GAMIFICATION

A SYSTEMATIC LITERATURE REVIEW OF GAMIFICATION DESIGN

Rokia Bouzidi[*], Antonio De Nicola[**], Fahima Nader[*] and Rachid Chalal[*]

[*]Laboratoire des Méthodes de Conception des Systèmes, Ecole nationale Supérieure d'Informatique, BP 68M, 16309, Oued-Smar, Alger, Algérie. http://www.esi.dz

[**]Italian National Agency for New Technologies, Energy and Sustainable Economic Development (ENEA), Casaccia Research Centre, Via Anguillarese 301, 00123, Rome, Italy

KEYWORDS

Gamification design, systematic literature review, framework

ABSTRACT

Gamification is the use of game design elements in non-game contexts with the aim of motivating users and improving their productivity. It gave promising results when applied to many domains. However, analysts predict that most gamification systems are doomed to failure due to the lack of understanding of the gamification design process. Many attempts tried to propose design frameworks for gamification. Nevertheless, only few works were conducted to review the literature in this area of research. The present work is an attempt to review the literature for gamification design. It analyses and compares 28 candidate papers dealing with gamification design for generic application from different digital libraries. It results in a synthesis of the existing body of knowledge and identifies gaps that need to be completed.

INTRODUCTION

The concept of gamification defined as the use of game design elements in non-game contexts (Deterding et al. 2011) gained a great interest from both academics and industrials during the last years. It gave promising results when applied to many domains like education and health among others. However, analysts predicted that most of the gamification solutions are doomed to failure due to poor understanding of the gamification design process (Pettey and van der Meulen 2012). Different approaches have been proposed to support the gamification design process. However, only few works have been interested in reviewing the existing body of knowledge around gamification design. Accordingly, we present this work as an attempt to characterize the state of the art around gamification design. The rest of this article is organized as follow: Section 2 presents related work to reviewing literature for gamification design. Section 3 describes the research methodology we followed to conduct this review. Section 4 presents the results of this literature review. Section 5 is a discussion of these results. Finally, we provide conclusions and perspectives of this work.

RELATED WORK

We present in this section existing literature reviews for gamification design:

(Mora et al. 2015; 2017) carried out reviews of gamification design frameworks. They analyzed gamified design frameworks according to related game design items clustered in five categories (i.e. economic, logic, measurement, psychology and interaction). This analysis aims to determine if gamification frameworks inherit game design principles for their development. Authors concluded that the analyzed approaches are on the right way. However, they don't consider some important keys for effective gamification design like involving stakeholders, preventing risks and using metrics for measurement.

Recently, (Azouz and Lafdaoui 2018) conducted a systematic mapping study for gamification design frameworks using 58 documents for comparison and analysis. The main interest of this study was to determine the current state of the art for gamification design frameworks. However, as stated in (Petersen et al. 2008) systematic mappings do not study articles in enough detail. They emphasize on providing classifications and thematic investigations to identify publication trends.

(Bouzidi et al. 2019) propose an integrated ontology for the domain of gamification. The ontology counts modular sub-ontologies for gamification concepts namely: gamification core concepts, user concepts, psychological concepts, organizational concepts, ethical concepts, risk concepts and evaluation concepts. Authors carried out a review of the literature to build their ontology. The review includes concepts dedicated to gamification design. For instance, gamification approaches, gamification elements to be used (i.e. gamification mechanics and dynamics) and user types. Yet, it does not analyze in enough detail design aspects since it mainly covers other sub-domains of gamification.

To sum up, there exist in the literature only two systematic reviews for gamification design where the considered works date back to 2015 with only one systematic mapping study. Based on the aforementioned findings, the need for a systematic review for gamification design is clearly required.

RESEARCH METHODOLOGY

We present in this section the research methodology we adopted to conduct this review.

Research questions

We enumerated 3 research questions for this review as follows: 1) What are the frameworks used for the design of gamification for generic application in the literature? 2) What are the common design aspects that gamification design processes consider? 3.1) Does it need technology to deploy gamification? 3.2) Does is it recommend implementing the designed gamified system or it uses existing platforms?

Data collection process

The string used for the search consists of two parts. The first part includes gamification and gameful synonyms. The second part includes terms that can be used to describe ways to build gamified systems. We list 9 terms namely; framework, model, design, approach, method, process, strategy, implementation and technique. Then, we chose sources to make our search. We adopted ACM, IEE, AIS Electronic Library, Science direct, Springer, Wiley InterScience, Emerald Insight, Taylor and Francis Online, JSTOR as target sources.

Data extraction process

We designed two data extraction forms. The first form is dedicated to papers proposing a way to design gamification. It consists of a set of 17 questions grouped in 3 categories according to the three research questions. The second form was used for review papers about gamification design. It focuses on the selected works in each review with the analysis criteria and results of the review.

RESULTS OF THE SEARCH

In this section, we report the results of this review. The first part gave general results providing statistics about the selected studies. The second part answered the research questions discussed in this paper.

General results

Through the research methodology, we identified 28 studies on gamification design. The number of studies increased significantly in 2013 with 8 research studies. This number dropped down in 2014 to mark 5 studies in the next year. Starting from 2016, the number of studies continues its gradual change to attend the peak during the year of 2018 with 9 research studies. This shows the great interest the field has gained in recent years. We also note that conference proceedings and journals are the predominant venues for gamification design with respectively 13 and 6 studies.

Answering research questions

RQ1: What are the frameworks used for the design of gamification for generic application?

We summarize in table 1 answers to the first research question. For each framework: it gives a brief description of the work, it identifies the adopted definition of gamification, it determines if the proposed work is a guideline only or a complete framework (✓: A guideline, x: A complete framework), it determines if the paper considers related work exclusively from the academic or professional world or both of them, (A: Academic, P: Professional, B: Both, NM: Not mentioned) and it describes the domain of application of the proposed framework.

RQ2: What are the common design aspects that gamification design processes consider?

We present in the following a list of common design aspects the selected papers consider when presenting gamification design approaches.

User-centric design

The user-centric design requires that the user's needs and goals are the primary concerns in each step of the design process. Results revealed that applying user-centered design in gamification is attracting more and more the attention of designers. Almost half of the frameworks support explicitly the adoption of the user-centric design.

User types

User types is considered as an important input for gamification design, since every type results in different preferences to certain gamification elements. In this context, researchers are working on player typologies and the corresponding gamification elements. However, we noticed that only 8 papers out of 25 explicitly cite the player types to be adopted in the gamification design. The most frequent player types are the Hexad player types, Bartle's player types and Octalysis frameworks of motivational drivers (Böckle et al. 2018).

Gamification elements

Based on the results we reported, only 8 out of 25 papers propose a list of gamification elements to be used during the design process. Other authors state that selection of gamification elements is a creative process. Thus, limiting the choice of gamification element will restrict designers' creativity (Morschheuser et al. 2018).

Involving stakeholders

The objective of involving stakeholders in the design process is to let them understand and participate in the gamification solution. Even though recent studies recommend involving stakeholders in the early stage of the gamification design, only 6 studies out of 25 recommend doing so.

Feasibility step in the gamification process

Feasibility is an important steps that more than half of the studies do not consider. We refer to this step by using different names like feasibility, requirement , and declaration step or by asking questions namely; determine if gameful design fit? Or check whether gameful design is an effective and efficient strategy?

Table 1: Summary of answers for RQ1

Paper	Description	Adopted definition of gamification	Guide line	Related work	Domain
Morschheuser et al. (2018)	A detailed method for engineering gamified software with key design principles.	Deterding et al. (2011), Huotari and Hamari (2017)	x	B	General
(Morschheuser et al. 2017)	This method is the first version of the one proposed in (Morschheuser et al. 2018).	(Huotari and Hamari 2017)	x	B	General
(Deterding 2015)	A gameful design method that combines design lenses with skill atoms and intrinsic integration.	(Deterding et al. 2011)	x	B	General
(Böckle et al. 2018)	A design framework for the emerging research stream of adaptive gamification where the gamified system adapts to user characteristics	(Deterding et al. 2011)	x	A	Adaptive
(Gunta et al. 2018)	A design framework for Web apps accommodating gamification strategies.	(Deterding et al. 2011)	✓	A	Web apps
(Herzig et al. 2015)	A gamification development process that describes roles and tasks involved in gamification projects with conceptual requirements for gamification.	(Deterding et al. 2011)	x	A	General
(Chen 2018)	Guidelines of how to conduct a gamification project with user-centered design.	(Deterding et al. 2011)	✓	B	General
(Kumar 2013)	A practical guide for user experience designers, product managers and developers to incorporate the principles of gamification into their software.	NM	x	NM	Business
(Ning 2018)	A design method of gamification systems which corresponds user experiences in three levels of nature, process, and interface.	(Werbach and Hunter 2012)	✓	A	General
(Rapp 2017)	A set of recommendations for the gamification design of interactive systems.	(Deterding et al. 2011)	✓	B	General
(Versteeg 2013)	An ethical framework to serve as a basis of moral gamification design.	(Deterding et al. 2011, Huotari and Hamari 2017)	✓	A	General
(Marache-Francisco and Brangier 2013)	A user-centered design approach for gamification design with main factor to be considered (intention, situation, task, users)	(Deterding et al. 2011)	x	A	General
(Nicholson 2012)	A framework based on user-centered design for meaningful gamification.	(Deterding et al. 2011)	x	A	General
(Chou 2015)	A gamification design framework called Octalysis, which derives its name from an octagonal shape with 8 Core Drives representing each side of the framework.	(Chou 2015)	x	NM	General
(Werbach and Hunter 2012)	A design process of six steps for gamification design, also known as 6D process.	(Werbach and Hunter 2012)	x	NM	General
(Klevers et al. 2016)	An implementation model of three phases for gamification implementation.	(Deterding et al. 2011)	x	A	Business
(Francisco-Aparicio et al. 2013)	A method for applying gamification as a tool to improve motivation of people.	(Deterding et al. 2011)	x	NM	General
(Gears and Braun 2013)	A development process for gamification systems.	(Deterding et al. 2011)	x	A	Business
(Julius and Salo 2013)	A framework for designing a gamification system suited for marketing.	(Deterding et al. 2011)	x	B	General
(Kappen and Nacke 2013)	A design-centric model and analysis tool for building gamification systems.	(Deterding et al. 2011)	✓	A	Business
(de Paz 2013)	Generic guidelines for the design of gamified systems.	(Deterding et al. 2011)	x	A	General
(Mora Carreño 2018)	A framework for the design of personalized gamification services. The framework is called FRAGGLE for: FRamework for AGile Gamifcation of personalized Learning Experiences.	(Deterding et al. 2011, Pelling 2011)	x	A	Education
(Fitz-Walter 2015)	A framework for designing gamification, derived from an iterative process of evaluation.	(Huotari and Hamari 2017, Guin et al. 2012, Deterding et al. 2011)	x	A	Mobile apps
(Knutas et al. 2018)	A design process based on Deterding's framework for gameful design using machine learning algorithm for gamification implementation.	(Deterding et al. 2011)	x	A	General
(Lopez and Tucker 2018)	A machine learning method that uses task information and an individual's facial expression data to predict the performance of the user on a gamified task.	(Deterding et al. 2011)	x	A	Adaptive

In addition to the aforementioned aspects that gamification design should consider, we present in the following additional criteria that can be used to assess the quality of a gamification design process.

Ethical consideration

Although gamification gave positive results to motivate users, it is considered as a source of tension and pressure that may affect social and mental well-being of the users. Despite the importance of ethical regulation in the gamification process, only 9 of the selected studies consider this aspect when designing gamification solutions.

Risk consideration

Designers need to take in consideration potential risks when making decisions about applying gamification like task quality, cheating the system, and privacy of the user. A good gamification process should prevent risks and deal with them during the design of the gamified system. Among the papers we analyzed, only 6 papers address the risk concept in gamification design.

Evaluation of the proposed process

To evaluate the validity of a gamification design process, authors in the literature used mainly case studies with 54% to verify the applicability of their propositions. Other works adopt expert evaluation using questionnaires or interviews to assess the quality of their work or combine both case studies and expert evaluation.

RQ3: How is the gamified system implemented?
RQ3.1 Does it need technology to deploy gamification?

Almost all papers where this aspect was discussed consider that gamified systems need technology to be implemented e.g (Morschheuser et al. 2018; 2017, Deterding 2015). While half of the discussed papers do not address this issue, (Werbach and Hunter 2012) is the only work that states that gamification does not necessary require technology but it fits perfectly with online systems.

Does is it recommend implementing the designed gamified system or it uses existing platforms?

An average of 80% of the studies do not mention how to implement gamified systems. Only 4% (Deterding 2015) recommends implementing the gamified system entirely while reusing existing gamification platforms was never recommended alone. 16% of the studies recommend both implementing and reusing existing platforms depending on the context and IT expert choices (Morschheuser et al. 2018; 2017, Herzig et al. 2015).

DISCUSSION

In this section we provide some discussions about the results we reported during this review.

Gamification theoretical concepts

During this review, we could identify that gamification design is a trending area of research. Although early works about gamification design date back to 2012, there is still a lack of agreement on the main concepts involved in the design process. For instance, gamification definition has always been considered as the starting point of each work proposing a way to design gamification. However, there is not an agreed upon definition for gamification among the selected papers. Each paper adopts subjectively a definition based on the authors' choice. Other concepts that need to be discussed are gamification elements to be used and player types to be considered. These findings highlight the need for more philosophical research in order to establish the foundation for a mature area of research.

Gamification design tendencies

In addition to usual gamification design approaches, we identified the emergence of a new stream in gamification design that adapts the gamified system to the user by considering the needs and goals of the end-user. We refer to this concept by adaptive or personalized gamification design (Böckle et al. 2018, Lopez and Tucker 2018). Another tendency is the use of machine learning methods in the design process by using task information and individual's facial expression data to predict the performance of a user on a gamified task as reported in (Lopez and Tucker 2018).

Gamification design gaps

Few interest was given to the way gamification systems are implemented. Existing frameworks propose either reusing existing platforms or implementing the new gamified system from scratch but no evaluation nor validation research with rigorous experimental methods was conducted in this context. Moreover, the technical architecture of the gamified system was not included in the design process even though separate works proposed ways to deploy the gamified system (Sripada et al. 2016).

Ethical considerations are an important gap that design work should consider. Only few recent studies include ethical concerns when designing a gamification solution. To this we can add preventing risks that may occur due to gamification like cheating, undesired competition and resistance to change. Future work should be able to prevent these risks and deal with them if they take place.

CONCLUSION

In this study, the objective is to identify the existing body of knowledge around designing gamification solutions in generic contexts via 28 papers from the literature. Results of the review state that there is a rapid and increasing interest on the design of gamification solutions. Based on the results of this review, we identify a lack of an agreed upon body of knowledge around gamification design. The result of this study may be used to build a holistic approach based on the existing isolated methods. Furthermore, this review highlights new streams in the field of gamification design namely, the adaptive gamification. This stream aims to provide personalized gamification solutions tailored to different users and contexts to optimize gamification results.

REFERENCES

Azouz O. and Lafdaoui Y., 2018. *Gamification design frameworks: a systematic mapping study.* In *2018 6th International Conference on Multimedia Computing and Systems (ICMCS)*. IEEE, 1–9.

Böckle M.; Micheel I.; Bick M.; and Novak J., 2018. *A design framework for adaptive gamification applications.* In *Proceedings of the 51st Hawaii International Conference on System Sciences.*

Bouzidi R.; De Nicola A.; Nader F.; and Chalal R., 2019. *OntoGamif: A modular ontology for integrated gamification. Applied Ontology*, , no. Preprint, 1–35.

Chen Y., 2018. *Exploring Design Guidelines of Using User-Centered Design in Gamification Development: A Delphi Study. International Journal of Human–Computer Interaction*, 1–12.

Chou Y.k., 2015. *Actionable gamification: Beyond points, badges, and leaderboards.* Octalysis Group.

de Paz B.M., 2013. *Gamification: A tool to improve sustainability efforts.* Ph.D. thesis, Ph. D. dissertation.

Deterding S., 2015. *The lens of intrinsic skill atoms: A method for gameful design. Human–Computer Interaction*, 30, no. 3-4, 294–335.

Deterding S.; Dixon D.; Khaled R.; and Nacke L., 2011. *From game design elements to gamefulness: defining gamification. New York, NY: ACM.*

Fitz-Walter Z.J., 2015. *Achievement unlocked: Investigating the design of effective gamification experiences for mobile applications and devices.* Ph.D. thesis, Queensland University of Technology.

Francisco-Aparicio A.; Gutiérrez-Vela F.L.; Isla-Montes J.L.; and Sanchez J.L.G., 2013. *Gamification: analysis and application.* In *New trends in interaction, virtual reality and modeling*, Springer. 113–126.

Gears D. and Braun K., 2013. *Gamification in business: Designing motivating solutions to problem situations.* In *Proceedings of the CHI 2013 Gamification Workshop.*

Guin T.D.L.; Baker R.; Mechling J.; and Ruyle E., 2012. *Myths and realities of respondent engagement in online surveys. International Journal of Market Research*, 54, no. 5, 613–633.

Gunta K.; Sivaselvan B.; and Oswald C., 2018. *Gamification Paradigm for WebApps Design Framework.* In *2018 International Conference on Computer, Communication, and Signal Processing (ICCCSP)*. IEEE, 1–5.

Herzig P.; Ameling M.; Wolf B.; and Schill A., 2015. *Implementing gamification: requirements and gamification platforms.* In *Gamification in education and business*, Springer. 431–450.

Huotari K. and Hamari J., 2017. *A definition for gamification: anchoring gamification in the service marketing literature. Electronic Markets*, 27, no. 1, 21–31.

Julius K. and Salo J., 2013. *Designing gamification. Marketing.*

Kappen D.L. and Nacke L.E., 2013. *The kaleidoscope of effective gamification: deconstructing gamification in business applications.* In *Proceedings of the First International Conference on Gameful Design, Research, and Applications.* ACM, 119–122.

Klevers M.; Sailer M.; and Günthner W.A., 2016. *Implementation model for the gamification of business processes: a study from the field of material handling.* In *Simulation and Gaming in the Network Society*, Springer. 173–184.

Knutas A.; van Roy R.; Hynninen T.; Granato M.; Kasurinen J.; and Ikonen J., 2018. *A process for designing algorithm-based personalized gamification. Multimedia Tools and Applications*, 1–20.

Kumar J., 2013. *Gamification at work: Designing engaging business software.* In *International conference of design, user experience, and usability.* Springer, 528–537.

Lopez C. and Tucker C., 2018. *Towards Personalized Adaptive Gamification: A Machine Learning Model for Predicting Performance. IEEE Transactions on Games.*

Marache-Francisco C. and Brangier E., 2013. *Process of gamification. Proceedings of the 6th Centric*, 126–131.

Mora A.; Riera D.; Gonzalez C.; and Arnedo-Moreno J., 2015. *A literature review of gamification design frameworks.* In *2015 7th International Conference on Games and Virtual Worlds for Serious Applications (VS-Games).* IEEE, 1–8.

Mora A.; Riera D.; González C.; and Arnedo-Moreno J., 2017. *Gamification: a systematic review of design frameworks. Journal of Computing in Higher Education*, 29, no. 3, 516–548.

Mora Carreño A., 2018. *A framework for agile design of personalized gamification services.*

Morschheuser B.; Hamari J.; Werder K.; and Abe J., 2017. *How to gamify? A method for designing gamification.*

Morschheuser B.; Hassan L.; Werder K.; and Hamari J., 2018. *How to design gamification? A method for engineering gamified software. Information and Software Technology*, 95, 219–237.

Nicholson S., 2012. *A user-Centered theoretical framework for meaningful gamification, paper presented at the Games+ Learning+ Society 8.0. 80, Madison, USA.*

Ning B., 2018. *A UX-Driven Design Method for Building Gamification System.* In *International Conference of Design, User Experience, and Usability.* Springer, 112–124.

Pelling N., 2011. *The (short) prehistory of gamification. Funding Startups (& other impossibilities)*, 9.

Petersen K.; Feldt R.; Mujtaba S.; and Mattsson M., 2008. *Systematic mapping studies in software engineering.* In *Ease.* vol. 8, 68–77.

Pettey C. and van der Meulen R., 2012. *Gartner says by 2014, 80 percent of current gamified applications will fail to meet business objectives primarily due to poor design. Gartner: USA.*

Rapp A., 2017. *Designing interactive systems through a game lens: An ethnographic approach. Computers in human behavior*, 71, 455–468.

Sripada S.K.; Reddy Y.R.; and Khandelwal S., 2016. *Architecting an extensible framework for gamifying software engineering concepts.* In *Proceedings of the 9th India Software Engineering Conference.* ACM, 119–130.

Versteeg M., 2013. *Ethics & Gamification design: a moral framework for taking responsibility.* Master's thesis.

Werbach K. and Hunter D., 2012. *For the win: How game thinking can revolutionize your business.* Wharton Digital Press.

GAME AND HUMANITARIAN: FROM AWARENESS TO FIELD INTERVENTION

Emmanuel Guardiola
Cologne Game Lab
TH Köln
51063, Cologne,
Germany
E-mail: eg@colognegamelab.de

KEYWORDS
Video game, humanitarian, refugee, Do No Harm principle, game design.

ABSTRACT

Since the early 2000s, game developers are more and more interested or involved in topics related to Humanitarian aid. From the first initiatives that targeted people totally external to the humanitarian situation to the most recent initiatives that tried to impact the situation. In recent years, video games for humanitarian causes reached a pivotal point: the apparition of games that try to have immediate measurable impact in particular regarding the population affected. For these kind of games new design paradigms seem necessary, as for instance the Do No Harm humanitarian principle. A continuum analysis offers a new perspective on this progressive involvement. After an overview of the evolution of connections between game and humanitarian action, we will examine one representative case of these pivotal games for direct field intervention, to illustrate the way designers should integrate humanitarian intervention principles.

INTRODUCTION

Board games are used to understand humanitarian crises and to train to prevent them, as for instance PAX Sims board games *Aftershock: A Humanitarian Crisis Game* (PaxSim2015) or *ISIS Crisis* (PaxSim 2016). The use of non digital games as field intervention tools has also been explored by many humanitarian organizations for decades. They could aim, for instance, to help children suffering from trauma and need psychosocial support. Several manuals are available from field organization like Unicef (Macy 2002) or Terre des Hommes (Meuwly 2007).

On the video game side, it took 15 years from the first initiatives to sprout real field impact. Most of this paper will document this progression and then provide a framework to position the current and future humanitarian games initiatives, from addressing audiences out of the conflict to affecting the victims themselves. Initially the developers used games to promote awareness about a specific crisis. Over time, the objective shifted progressively to other kinds of impacts such as teaching about the role of some nonprofit or Non-Governmental Organizations. Also, the topic of humanitarian crisis became a context in entertainment games, and even began to feed art oriented interactive experiences. Later, games aimed to have more impact, even being used to raise fund for NGOs. Finally, there appeared several initiatives made to be deployed in the field: games or gamified applications developed for the prevention of risks during crisis; or to help on the educational or psychosocial levels.

Developers, designers and researchers have reached a pivotal point these last years with the apparition of this last category of games. The design, production and distribution of these humanitarian field games raised new questions and now need specific paradigms to achieve their goals. The discussion part will evoke the integration of some of the core humanitarian principles into the process of making game suitable with crisis environment.

Note that in this paper, when we refer to *humanitarian games*, it covers exclusively Video Games, whatever is the form of connection to the topic, from entertainment games, that use humanitarian crisis as contexts, to games that aims to impact a current crisis.

THE EVOLUTION OF HUMANITARIAN VIDEO GAMES

In our research we try to reflect a particular aspect of the evolution of humanitarian games. In the ensemble of titles fitting with this category, the origins of the projects (funders, clients, etc) vary a lot. Their aims also cover many types of impacts (awareness, educational, entertainment, etc.). Regarding past attempts to distinguish categories of games with an intended impact, mainly from serious game field, we try to establish our own lenses.

Sorting humanitarian games using past taxonomies

There are already several taxonomies of video games with measurable impact, but none really help to visualize this progression. For instance, the Sawyer et al (2008) taxonomy proposes a grid with two dimensions: one (Rows) for the type of market segment that the game tries to impact (Government and NGO, Defense, Healthcare, etc.), and the second (Column) related to the type of purpose (health, training, advertising, etc.). This central grid is completed of sub-grids dedicated to different purposes. In the main grid, the market row “government and NGO” might help to classify several of the games we want to investigate: for instance, the games produced or made for NGOs. But a part of the games related to humanitarian games does not find a proper slot, even in the sub segmentations proposed in the overall Sawyer proposition. Games from the entertainment realm engaging in awareness or empathic intentions like *Bury me, My Love* (The Pixel Hunt et al. 2017) couldn't fit in this strict serious game perspective. Also, the fact that, in Sawyer's taxonomy, we have to zoom in to view several sub categories of intended impacts does not help to clearly illustrate the key dimension of our approach, the link to the field of humanitarian work.

Several following taxonomies investigated by Djaouti et al (2011) are related to specific purpose (Bergeron 2006; Alvarez et al., 2007) or to the market (Zyda 2005; Chen et al. 2005; Alvarez et al. 2008). However, they all suffer from the same issues: they are strictly categorizing serious games, making it difficult to use them to illustrate the link to the field.

Djaouti et al. (2011) propose a simplified model: the G/P/S, for Gameplay / Purpose / Scope. This approach can potentially help us to analyze our set of games, but again the authors apply their model exclusively to serious games. If we accept that an entertainment game could also aim to have an impact (promote a culture, rise the attention on a societal issue or situation, etc.) then this model could eventually help to classify all humanitarian games. Modifying the "scope" dimension to include the audience and market of use could be a way to include our main dimension of analysis, the connection to the field, but by making all these transformations we completely change the current classification model. To achieve our main objective, by illustrating a progression of connection from the activities and themes of the game to the humanitarian crisis itself, we must set up a new approach.

Dimension to observe the evolution

Our goal is to observe the evolution of humanitarian video games from the last 15 years. The origins of projects (Funding, client) and types of audience are interesting parameters to classify the different types of initiatives. We also noted another interesting parameter, one that illustrates an evolution: How the game is connected to an actual situation and impacts it? We can set this dimension in a continuum from games that address an audience outside of the crisis, to games that directly address the victims with a clear attempt to impact the field situation.

Along the continuum

The following list of games is not exhaustive but an attempt to get representative ones, with examples of positioning all along this continuum. We start games most distant from the actual humanitarian situation and with no explicit link to people affected by a crisis. The more we progress in the list, however, the closer we become to the field.

Entertainment with fantasy context

The first examples come from entertainment. Numerous games use humanitarian crisis as elements of their narrative contexts or gameplay. In our continuum, these types of games, using fantasy contexts, are the most distant from the situation. The users and makers are distant from the field and there is no impact on an actual, identifiable crisis.

The refugee figure is particularly present. We can cite *Neverwinter Online* (Cryptic Studios 2013) and *Final Fantasy XV* (Square Enix 2016) as representative examples. In Final Fantasy XV the player meets the Comrade Refugees during a specific quest chain and has to rescue them. In *Neverwinter Online* an entire environment is dedicated to a crisis: the Lonelywood. In this scenario, the player must complete various quests in order to protect refugees. These tasks include but are not limited to: finding wood to make fire, fetching medications, escorting groups of refugees, and defending the camp.

Art and experimentation with no specific situation

In another distant position from the field, we can find some art installations or experimental games. The following examples use humanitarian crisis as a topic and, here again, the concept of refugee is a strong figure. If the approaches could be abstract or conceptual, they find their inspiration into current humanitarian crisis. Compared to the previous examples, this is a step in our continuum.

Several games and playable installations were proposed to the public during the Art Game Demos#4 artistic event in Lyon in December 2017, dedicated to the borders and refugees topics (http://www.kareron.com/art-games-demos-evenements/#agd4). For instance the game *North*, (Helfenstein et al. 2016) invites the player to experience applying for asylum status in a strange city (Fig. 1).

Figure 1: Screenshot of the art game *North*, where the player is confronted as a refugee in a strange city

It tries to transmit the feeling of being immerge in another culture environment, something impossible for most people to understand. In another installation, *Fuir la guerre* (Alineaire 2015), the player encounters metaphorical stages of a refugee family's trip to liberty (Fig. 2).

Figure 2: The installation *Fuir la guerre*, at the Art Game Demos#4 exposition

Empathy or awareness to specific situation
Initiated by artists, another category of games are a step more in the direction of directly representing an existing situation. In the following games, the authors refer to specific contexts such as the Syrian refugee crisis or the American-Mexican border. Here again the figure of the refugee is recurrent, but in different ways.
Borders (Alvarez et al. 2017) proposes a game installation where the player must pass the Americano-Mexican frontier (Fig. 3). The content is directly inspired from the story of the author's mother and father. Game mechanics such as avoiding patrols in the desert or preserving water are part of the experience.

Figure 3: *Borders* installation, highlighting the desert environment of the game

Passengers (Alliot et al. 2015) takes a completely different angle. The player experiences the point of view of smuggler, who must funnel the refugees across the Mediterranean Sea to Europe (Fig.4). In this minimalist tycoon game, the process of dehumanization of refugees is very well illustrated as the player tries to optimize his "business".

Figure 4: Screenshots of *Passengers*, illustrating the exploitation of refugees

With a more empathic approach, several games also propose to share the experience of being a refugee. *21 days* (Hardtalk Sudio 2017) is a simulation adventure game where the player follows a man, already arrived in Europe who seeks to help his family to rejoin him. The player faces many of the issues related to this situation, from language barriers (Fig.5) to hunger.

Figure 5: Screenshot of 21 Days, when the player faces the difficulty of understanding a foreign language as a refugee

With an alternate approach *Bury me my love* (The Pixel Hunt et al. 2017) offers to depict the refugee situation from the perspective of someone staying in the conflict area. The player is the husband of a woman trying to find her way to Europe. The only way to influence the relative's travel and decisions is through the text message exchange of the husband's phone (Fig.6). This game, inspired from actual refugees stories, received critical success, awards and was released on many platforms (iOS, Android, Windows PC, Nintendo Switch).

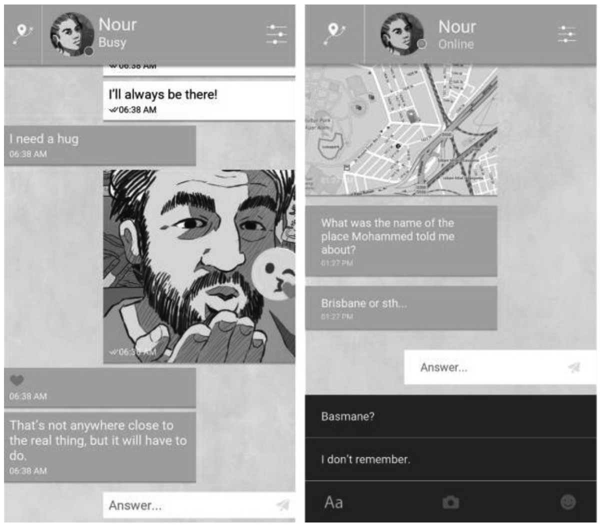

Figure 6: Two screenshots from the iOS version of *Bury Me, My Love* simulating interactions of a Syrian couple through their mobile phones

NGOs' Awareness of specific situations
Engaging with the next step in our continuum we now find games that are not only connected to actual field situations but additionally the funders or initiators are from humanitarian world. We can see several examples of organizations or foundations that fund projects to generate awareness on specific situations.
One well-known title from this category is *Darfur is dying* (Ruiz 2006). The funding and support came from humanitarian aid organizations such as the Rebook Human Right Foundation and the International Crisis Group and was developed by students of the University of Southern California. The goal of this flash-based game was to generate some awareness in the occidental audience about

the terrible situation of refugees in Darfur. In the game, the player must manage a family struggling for survival, but is quickly confronted to situation where you lose family members (Fig.7). The message was explicit and the game generated 700 000 views in the first month of its publication online (Sayre et al. 2010).

Battle Kid (Amnesty International 2012) is a smart example of an alternate use of game. Under the guise of a real shooter, Amnesty International invites the player to choose a conflict on a world map (actual conflicts involving child soldiers) and then to customize this child, by equipping him with weapons and armor. The disproportion of equipment and fragility of the child is obvious (Fig.8). When the player launches the game, a loading screen starts, and then a splash screen delivers a text from Amnesty International about the current situation of children in war. The core message is "you don't play with a kid's life". There is no real game outside the use of classical multiplayer shooter loadout customization features to convey this message.

Figure 7: Screenshots from *Darfur is* dying. On the left a phase of water gathering where you are trying to avoid hostile militias. On the right: the camp where you survive with your family

Figure 8: Screenshot of the fake game *Battle Kid*, where the player customizes her child soldier.

Promote the role of Nonprofit NGO

Taking a step further, we found games close to the previous (funded by NGOs, related to actual situation) but that tries to show how to solve the situation. Their main purpose is to promote the NGOs' impact on the field or inform the players about the way these organizations works.

Food force (Deepend et al. 2005) was certainly a first in this category. Initiated by the World Food Programme, this game takes place in an imaginary but credible food-scarcity crisis. The player has to evaluate the number of people to feed, to balance the different components of the menus regarding the budget, and manage to reach out to the starved population. Supported by major names in video games like Ubisoft and Konami, the game saw more than 4 million downloads in one year.

Fund raising

If all previous examples could initiate a donation to NGOs or tease a voluntary engagement in a cause, these games were not made for these objectives. As we progress toward the field, the next step is explicit fund raising for the activity of NGOs. Whatever these game contexts are, whether imaginary or actual, they impact the situation by increasing the potential goodwill given towards the victims.

In 2010, Zynga game company used *Farmville*, their successful social game, to fund the World Food Programme action in Haiti after a violent earthquake. They solicited their community of player by selling specific items through their in game purchase feature. They design items with explicit connection to the crisis as shown on figure 9.

Figure 9: Two items that *Farmville* players could buy to fund WFP operation in Haïti

More recently, the International Committee of the Red Cross collaborated with the *Arma 3 Laws of War DLC* from Bohemia Interactive (2017). This mode for *Arma 3* (Bohemia Interactive 2017) features fictional humanitarian organizations in the game world. The intention was double: to familiarize the players with the laws of war (Fig.10) and to encourage fund raising for the Red Cross through their virtual-turned actual awareness of NGO influence in large and small-scale geopolitical conflict.

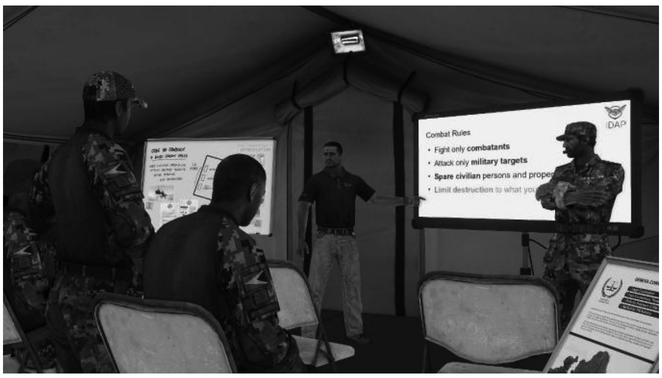

Figure 10: A brief sequence in *Arma 3 Laws of War DLC* with a focus on Laws of war.

Train or help humanitarian workers
The examples of this category reduce the distance to the victims: they are designed to train members from NGOs or Governmental agencies who intervene on the field.
For instance, *Gamoteca* (Technovatio 2016) is a gamified application intended to train humanitarian workers to work in various scenarios through a mix of face to face and use of the digital application.
Another interesting use of gamification can be observed in a running initiative of the organization Terre des Hommes. They aimed to increase the engagement of participants on a web page dedicated to child protection: https://childhub.org. To achieve this goal, they implemented very classic gamification features such as badges, quests, levels, leaderboard. According to an interview with the team, this approach did in fact increase users' participation and sharing of key game moments, which for them constituted an improvement of the community knowledge about the situation and the practices.

Prevent risk, behavior changes
From the humanitarian worker and general public, we shift to the victims or persons at risk. The following examples of games aims to improve their safety or comfort.
Games can provide a sense of solace after a crisis. The pilot research project *Mantra* (Soriano 2018) was created to improve the level of health knowledge in rural areas of Nepal, in particular after the consequences of the earthquake of 2015. The game teaches how to evaluate the severity of health symptoms encountered during pregnancy and in new born children. From this game the audience should get a strong sense of the appropriate reaction and understanding of potential urgency in the event of certain symptoms. These aspects are critical as initiate a movement in the mountain to reach the nearest medical center could be dangerous.
Games could also be designed with the intent of preventing potential crises. For instance, there are many initiatives in the cause to promote hygiene and educate players about germ theory. In 2016 the *Handwashing Innovation Sprint* received 120 propositions of applications and games to improve the handwashing and hygiene among children at risk. The winner of this particular contest *Play with Nazeef* was a game disseminated in close to 50 Palestinian schools and improved the handwashing in toilet from 65.2% to 98.2%. This is a clear measurable impact on an actual situation.

Direct impact on a crisis situation
Finally, we arrive at the extremity of the continuum. Now video games can even aim to be intervention tools in the crisis itself. This time, final users are victims and funding and development are connected to governmental or NGO concerned by the unfolding humanitarian situation.
The game *Antura & the Letters – Arabic* (Cologne Game Lab TH-Köln et al. 2018), is among the first widely distributed games in this category. Funded by an international call, *EduApp4Syria*, the aims are to provide literacy and psychosocial support to Syrian children in displaced populations. Deployed worldwide, this free game is mainly promoted in countries around Syria (Fig.11). Supported by All Children Reading and the UNICEF, a large evaluation was done in some refugee camps in Jordan in 2017 (Koval-Saifi et al. 2018), demonstrating that even in the worst conditions (no school, no internet, illiterate parents, etc.) the game is accessible and also able to teach and to improve psychosocial well-being of the children.

Figure 11: *Antura & the Letters* Photos from playtest (left side) and impact evaluation (Right side), both with children refugee in Jordan

Sum up of the continuum
We might miss certain categories of games related to humanitarian even if we tried to be exhaustive. The following table (Table1) is an illustration of this continuum principle, from out of the field to direct intervention. If other categories appear, they might find a position on this continuum.

Table 1: Continuum of categories from the most external to the closest to the crisis situation

Criteria Current Identified categories	Audience, main targeted users	Funding or client	Area of diffusion and impact	Examples of games
Entertainment with fantasy contest	Broad audience, gamers	Entertainment game companies	Broad diffusion, not pointing at real crisis situations, and without explicit way to have impact	*Neverwinter Online* *Final Fantasy XV*
Art and experimental with no specific situation	Broad audience, interested in art or the topic	Indie game companies, artist, art funding	Limited diffusion, not pointing at specific crisis situation, and without explicit way to have impact	*North* *Fuir la Guerre*
Empathy or awareness to specific situation	Broad audience, with interested in the topic	Indie game companies, sometimes with cultural/media states grants	Limited diffusion pointing at specific crisis situations but without explicit way to have impact	*Borders* *Passengers* *21 Days* *Bury me my love*
NGOs' Awareness of specific situations	Broad audience, with interested in the topic	NGO's, foundations, government	Broad diffusion with identified indirect impact (Donation)	*Darfur is dying* *Battlekid*
Promote the role of Nonprofit NGO	Broad audience, with interested in the topic	NGO's, foundations, government	Broad diffusion, with identified indirect impact (Donation, enrolment)	*Food Force 1* *Food Force 2*
Fund raising	Broad audience, with interested in the topic	NGO's, foundations, government , entertainment game companies	Broad diffusion, indirect impact on the situation by funding interventions	*Farmville* Haiti operation *Arma 3 Laws of war* DLC
Train or help humanitarian workers	Humanitarian workers	NGO's, foundations, government ,	In training session of NGOs, indirectly impact on the situation by improving workers' skills	*Gamoteca* *https://childhub.org*

Prevent risk, behavior changes	People affected by the crisis	NGO's, foundations, government ,	In the territories affected by a crisis, prior or after it, with direct impact on the victims situation	*Play with Nazeef Mantra*
Direct impact on a crisis situation	People affected by the crisis	NGO's, foundations, government ,	In the territories affected by a crisis, during the crisis, with direct impact on victims situation	*Antura & the Letters - Arabic*

DISCUSSION: HUMANITARIAN PRINCIPLE AND GAME DESIGN

This progression into all these categories of games, with this *proximity to the field* perspective, leads us to the destination. In the past years, we have reached a pivotal point: video games are now directly addressing the situation by being in contact with the populations affected by humanitarian crisis, at the moment and in the place where these crises occur.
These video games are intervention tools among the other, in the middle of the crisis. and as any form of humanitarian field intervention, video games teams should adopt ethic and principles developed by NGOs. As for instance, among UNICEF humanitarian principles: *Neutrality*, *Respect for culture and Custom* or the *Do No Harm*. This last principle means that the intervention should be designed in a way that it is not causing more damage than it solves problems, and propose tools to have an as good as possible perception of the ecosystem of the crisis.
During the development of *Antura & the Letters – Arabic*, the team adopted several of these principles. For instance, for the *Do No Harm Respect for culture and Custom*, they framed the production with a very strong user centered approach. Among the initiatives: Focus groups with families in the concept phase; Involvement of Syrian designers and developers from the very early stages; Curriculum designed with Syrian elementary school teachers; Numerous playtests of prototypes, alpha and beta versions, with child refugees, in Europe and Middle East. Media studies and surveys on the elements of content were also conducted to be sure the game was able to fit with the culture and customs of the affected populace. Investigation of the numerous social media groups and pages related to refugees were also made in order to identify neutral ones. The team also focused on worldwide distribution to offer the game to the refugees at the same time it was available to inhabitants of their host countries. The communication strategy focused on directly addressing the families through digital campaigns which utilized UNHCR data. Even on the technical side, a bundle of constraints came to ensure the attempt to reach in priority the most fragile, for instance: the game had to consume less than 100mb of space so it would be easy and cheap to download in the crisis related countries. The game also had to allow the freedom to users to transfer of the executable from a phone to another. The game also was designed to be completely playable offline, capable of running on middle range hardware as well as old version of operating systems and adhered to other various technical accessibility guidelines. All these cumulated strategies and constraints appeared to answer the humanitarian context and the adherence to its principles.

The game is currently considered as a success in term of impact with 200 000 downloads in the region hosting most of the Syrian refugees, 1 million views of the trailer on YouTube and a community of 8000 persons on Facebook. The Middle East press covered the launch with more than 40 articles. Antura is played each week several thousands of times.
The adoption of humanitarian principles might be seen as constraints but they also are some sort of guarantee of the quality of the impact. Quality presents itself both in terms of intended real life benefit, here literacy and psychosocial well-being, and also the accessibility of adoption by the population.

CONCLUSION

UNICEF, the Red Cross and Terre des Hommes, are among the organizations that understand the role that game could play in the humanitarian context. There is now time to time international calls for game content, collaboration with students in game oriented training, contests and awards, and integration of gamification appearing in some of their operations. The initiatives are multiple and testify of the increasing of the interaction between the two worlds of digital entertainment and humanitarian aid.
The journey through different game types that we brought to light in the paper is not totally chronologic. But regarding the dates of the first of their kind in connection to the humanitarian context we can observe a progression toward the field that takes around 15 years. We used three ensemble of criteria to be able to point this progression. 1: the type of audience or targeted users; 2: the source of funding or main client; 3: the area of diffusion and intended impact. We use as milestones nine categories of games, with different properties in the 3 ensembles of criteria. From one of this category to the next one, we could identify a step toward the field, considered as the ultimate level of impact.
With the unlocking of this level comes the responsibility of any humanitarian intervention tool and the respect of some core principles. As shown by the summarized post mortem of *Antura & the Letters – Arabic* this new kind of framework positively affects the design, production and distribution of the game.
In light of these multiple initiatives between the world of play and the humanitarian world, we can legitimately say that there is a growing common ground that is profitable for collaboration. By increasing contact with the field, games for humanitarian causes will need to establish a strong basis of best practice. The first steps could be to understand how we can adapt the core principle of humanitarian intervention to the case of video game tools (such as freedom in accessibility). Also, we might miss specific opportunity to share our experiences when establishing these guidelines. Workshops, conference, there is a fertile ground to formalize more efficiently the current state of art and prepare the next generations of video games. The quality of the experience and the impact felt by those involved will beneficiate from this growing collaboration. This improvement might one day be seen all along the continuum, from a better understanding and transmission of

the core concepts of the situation to the audience abroad at one extremity, to more efficient intervention tools on the other side.

ACKNOWLEDGEMENTS

We would like to thank Alex Boccia for his editorial assistance in the preparation of this article. We would like also to thank to the *Antura & The Letters* team for their dedication. The Antura project was funded and supported by the Norwegian Ministry of Foreign Affairs, as part of EduApp4Syria innovation competition, UNICEF and All Children Reading.

REFERENCES

Alineaire. 2015. "Fuir la guerre". Art interactive installation. Alineaire, Cergy, France.

Alliot, F. & De Bock, A. 2015. *Passengers*. PC game. http://ludumdare.com/compo/author/nerial/ (last visited 09/02/2019)

Alvarez, G., Digiacomo, J. & Genero V. R. 2017. "Borders". PC game. https://gonzzink.itch.io/borders (last visited 09/02/2019)

Alvarez, J., & Michaud, L. 2008. *Serious Games: Advergaming, edugaming, training and more*. IDATE, France.

Alvarez, J., & Rampnoux, O., & Jessel, J-P., & Methel, G. 2007. "Serious Game: just a question of posture?". In *Artificial and Ambient Intelligence convention* (Artificial Societies for Ambient Intelligence) - AISB (ASAMi) (pp 420-426), University of Newcastle, UK.

Amnesty International. 2012. *Battle Kid*. PC game. Londres, UK: Amnesty International. Game not available, video at https://www.youtube.com/watch?v=tL7IEfdvj0o (last visited 09/02/2019)

Bergeron, B. 2006. "Developing Serious Games". Charles River Media, USA.

Bohemian Interactive. 2017. *Arma 3*. Bohemian Interactive, Prague, Czech Republic.

Bohemian Interactive. 2017. *Arma 3 Laws of War*. DLC for Windows PC game. Bohemian Interactive, Prague, Czech Republic.

Chen, S., & Michael, D. 2005. *Serious Games: Games that Educate, Train and Inform*. Thomson Course Technology, USA.

Cologne Game Lab - TH Köln, Video Games Without Borders, & Wixel Studios. 2018. *Antura & the Letters – Arabic.* Android, iOS, Windows PC. Video Game Without Borders, Sant Cugat del Valles, Spain.

Cryptic Studios 2013 *Neverwinter online.* Online Windows PC. Redwood City, CA, USA: Perfect World Entertainment

Deepend & Playerthree. 2005. "Food force". Windows and Mac OS game.

Djaouti D., Alvarez J., Jessel JP, 2011. "Classifying Serious Games: The G/P/S Model" in *Handbook of Research on Improving Learning and Motivation through Educational Games: Multidisciplinary Approaches* edited by Felicia, P. IGI Global, USA.

Square Enix. 2016. *Final Fantasy XV*. Playstation 4 / Xbox One. Square Enix, Tokyo, Japon.

Hardtalk Sudio. 2017. *21 days.* PC game. Hardtalk studio.

Helfenstein, G. & Neu, T. 2016. *North.* PC game. Outlands Games. http://outlands-games.com/portfolio/o3-cabinets/ (last visited 09/02/2019)

Koval-Saifi, N., & Plass, J. 2018. *Antura and the Letters: Impact and technical evaluation*. World Vision and Foundation for Information Technology Education and Development, Washington, DC, USA.

Macy, R. D., Macy, D. J., Gross, S., Brighton, P. and Stuvland, R. 2002. *Psychosocial Play and Activity Book For Children and Youth Exposed to Difficult Circumstances, Version 1.02* UNICEF- MENA, Center for Crisis Psychology,Norway.http://nchd.org.pk/ws/downloads/5.2%20Psychosocial%20Play%20and%20Activity%20Book%20-%20UNICEF%20.pdf (last visited 22/01/2019).

Meuwly, M. and Heiniger, J-P. 2007. *Laugh, run and move to develop together, Games with a psychosocial aim*. Terre des Hommes International Federation, Swiss. https://www.sportanddev.org/sites/default/files/downloads/1__laugh__run_and_move_to_develop_together__games_with_a_psychosocial_aim.pdf (last visited 22/01/2019).

PAX sims. 2015. *Aftershock: A Humanitarian Crisis Game.* Board Game. PAX sims, Canada.

PAX sims. 2016. *ISIS Crisis* Board Game. PAX sims, Canada. https://paxsims.wordpress.com/2016/06/03/exploring-matrix-games-for-mass-atrocity-prevention-and-response/ (last visited 09/02/2019)

Ruiz, S. 2006. *Darfur is Dying.* Flash based game. University of Southern California, Los Angeles, USA.

Sayre, S., & King, C. 2010. *Entertainment and society: influence, impacts, and innovations.* New York, USA: Routledge.

Sawyer, B. and P. Smith. 2008. "Serious Games taxonomy" *Presentation at the Serious Games Summit at the Game Developers Conference*, San Francisco, CA.URL:https://thedigitalentertainmentalliance.files.wordpress.com/2011/08/serious-games-taxonomy.pdf (last visited 22/01/2019).

Soriano, D., Arjyal, A., Bara, S., Boscor, A., Mueller, S., Hearn, G.J., Saville, N.M., Fordham, M., Le Masson, V., and Kostkova, P. & al. "Engaging Pictograms! A Methodology for Graphic Design in Enhancing Player Engagement: As Applied to the Design of a Serious Game for Nepalese Women with Low Literacy". In *proceedings of the DH'18*, April 23–26, 2018, Lyon, France. ACM, New York, USA.

Technovatio. 2016. *Gamoteca*. Android/iOS game. Technovatio Limited, London, UK. http://www.gamoteca.com/about/ (last visited 09/02/2019)

Terre des hommes foundation. 2016. *Play with Nazeef*. Android game. Terre des hommes foundation, Lausanne, Swiss.

The Pixel Hunt & Figs. 2016. *Bury me My Love.* iOS game. Plug In Digital, Fresne, France.

Zyda, M. 2005. "From Visual Simulation to Virtual Reality to Games". Computer, 38(9), 25-32.

Zynga. 2009. *Farmville.* Social network game. Zynga, San Francisco, USA.

BIOGRAPHY

PROFESSOR EMMANUEL GUARDIOLA, with expertise in game design methodologies, is a veteran of the video game industry with over 30 major titles released for publishers such as Ubisoft and independent studios such as Dontnod Entertainment. He works on licenses in various genres, for example: *Far Cry, Prince of Persia, Rainbow6, Life is Strange, Ghost Recon, Frank Herbert's Dune*. At Ubisoft, after driving research on game design and emotion or meaning, he was one of the creators of the *Game For*

Everyone brand, focused on the way to bring real life benefits to the users.

PhD in computer science, Emmanuel Guardiola is now Professor at the Cologne Game Lab – TH-Köln (applied sciences university of Cologne). In the field of game design research he focuses on models, processes, characteristics of the gameplay experience, and how we can apply this findings to education, health, psychology or humanitarian cause.

Regularly funded by National, European and International grants, he earned several awards as the European Game For Change Europe award, the eVirtuos R&D award or the Interactive Author of the Year award by the SACD (French Dramatic Author and Composer Society).

The *Antura & the Letters*, laureate of the international call EduApp4Syria, is his last major project.The Antura project addresses refugee children and uses mobile gaming to provide Arabic literacy and psychosocial wellbeing support. The game is currently in exploitation on the field and has more than 200 000 users. Antura received 7 international awards in the field of education and humanitarian.

SHARING SUSTAINABILITY DATA THROUGH AN OPEN DATA GAME

Kaisa Könnölä, Marko Loponen, Markus Krusberg, Teijo Lehtonen
Department of Future Technologies
20014, University of Turku
Finland
E-mail: kaisa.konnola@utu.fi, majulop@utu.fi, markus.krusberg@utu.fi tetale@utu.fi

KEYWORDS
sustainability, game design, gamification, open data

ABSTRACT
Both legislation based and voluntary sustainability reporting of companies has increased during recent years. Although reporting schemes, such as GRI (General Reporting Initiative) exist, sustainability reporting takes various forms. While this kind of sustainability data exists, the presentation can be tedious. In order to make sustainability information more fun, gamification can be utilized. In this experience report, the design and experiences of an open data game for sharing the sustainability commitments of Finnish Maritime Industry are presented. As an open data game, the focus is on letting the player to explore the data freely. Information sharing is the key of this application, and while most information was text-based, also badges were used, the game was based on a story, and the player was presented his/her interests at the end of the game. The experiences of the players were measured with a questionnaire which included In-Game Game Experience Questionnaire. The experiences of the players were positive and interest in this kind of application was also seen as a learning medium.

INTRODUCTION

Sustainability research, sustainability reporting, as well as general knowledge on sustainability has seen an increase during recent years. The concept itself is often split into three categories: environmental sustainability (sometimes ecological sustainability), economic sustainability and social sustainability (Elkington, 1999). The focus is usually on environmental sustainability, but attention is also paid to the other two categories. United Nations has created the 2030 Agenda for Sustainable Development, including 17 Sustainable Development Goals (SDG), which cover all aspects of sustainability (United Nations, 2015). These SDGs are widely known and utilized in both education and reporting of sustainability. The focus on the company side has changed from distinct social and environmental reporting gradually to include both of these areas and even the third dimension, i.e. economic sustainability (Hahn and Kühnen 2013). According to Hahn and Kühnen (2013), the research gaps of sustainability reporting can be found in regulation, governance, quality and stakeholder perception.

In scientific research, sustainability can also be seen in other roles: sustainability applications are considered as a research media. These applications typically attempt to change the personal behaviour of the target groups (e.g. Jylhä et al. 2013; Nguyen 2014; Brewer et al. 2015) and they tend to focus on changing economically and environmentally sustainable behaviour, such as the household energy usage or the transport method (Könnölä et al. 2018).

Although reporting sustainability actions exists, these reports can be tedious can be left out of public attention. In this paper, existing data of sustainable actions taken in the Finnish Maritime Industry is presented in an open data game to encourage the users to explore the data. The scientific scope is widened from changing the personal behaviour to informing the user. In the design and development of the application, previous knowledge of gamification and open data games is utilized. The aim of the research is to share information and measure the perception of the users on the gamified reporting of sustainable actions.

BACKGROUND

In the design of sustainability applications, pro-environmental elements, as well as game design elements from gamification, are used (Könnölä et al. 2018). Pro-environmental behaviour is defined as conscious actions that aim at reducing negative human impact on the environment (Kollmuss and Agyeman 2002), whereas pro-environmental elements are used to motivate this kind of behaviour (Froehlich 2010). Several definitions of game design elements have been created by going through existing gamified applications (Weiser et al. 2015; Sailer et al. 2013; Hamari et al. 2014) and these definitions resemble each other quite well. For example, assignments, quests and goals; points, credits and levels; achievements and badges; and leaderboards and collections are found in all of them. Some game design and pro-environmental elements are more commonly used in both research applications as well as popular sustainability applications: Information sharing can be considered as a critical element, whereas feedback and rewards are used to encourage sustainable behaviour through points, levels and badges, and commitment and comparison are created through a story or a leaderboard (Könnölä et al. 2018).

Open data have become more available due to the inner desire and external pressure for different groups to be as transparent as possible. For example, many governments and cities have published a wide range of data in their application programming interfaces (APIs) (Friberger et al.

2013). In open data games, the content comes from an external source (such as APIs and websites) and the gameplay mechanics are developed on top of the existing data (Macklin et al. 2009). Open data games differ from serious games since they provide a possibility to explore the data without restrictions instead of creating a learning environment or trying to change the behaviour of the player. Data games are strongly related to the data used in it, making the quality of the data play a significant role in the game. (Friberger et al. 2013) Visualizing the data in new ways is practical and straightforward, but by giving the user an active role, the player can bring their interest and behaviour to the game itself (Macklin et al. 2009).

RESEARCH METHOD

The research method can be divided into two parts: game development and game testing. Research questions focus on an open data game based on open company-related sustainability data:

1. Which game design and pro-environmental elements are suited for the application?
2. What kind of challenges were faced during the development of the application?
3. How the players experience a gamified application and its informative value?

The game was developed based on the existing open sustainability data and knowledge on open data games as well as sustainability applications. In the game design, several design choices were made on both the game design elements as well as in the ways a user can explore the data. The results of this part answer to the first two research questions and are described in Game Description Section.

After the creation of the game, it was tested with real users in a one-day science event aimed at students from the first grade to the end of high school. In addition to the students, adult participants (e.g. teachers, organizers, parents) were able to test the game. Taking part in testing was voluntary. As there were many underaged participants, permission from their parents was asked either in advance or they got permission leaflet about taking part in the research to be brought home. The application was used with tablets and the questionnaires to be filled were printed ones. The testing included a pre-game and an after-game questionnaire. In the pre-game questionnaire, the players were questioned about their age, gender and previous knowledge of sustainability in the maritime industry. The after-game questionnaire included a partial In-Game Game Experience Questionnaire (GEQ) (Ijsselsteijn, Kort & Poels 2008) and questions about learning from different sustainability areas. These results are presented in Game Experiences Section.

Even though GEQ is recommended to be used in its longer form, only In-Game GEQ was used due to the short time period each player was present in the game testing. The In-Game GEQ was also modified by dropping out challenge and competence questions: the game was more about telling a story, than having competitive or challenging elements. In addition to the questionnaire, the game logged data about the choices the players made during gameplay as well as the time it took them to play the game.

The data was collected in only one event, where the participants were mostly underaged. This can be seen to have an effect on the generalization of the results. With a longer version of GEQ, the generalization of the results would be more reliable.

GAME DESIGN

The game design aims at creating an alternative to the web page as a fun experience on the exploration of the sustainable actions of the Finnish Maritime Industry. The design of the game follows Friberger et al. (2013), i.e. the player should be able to explore the data without restrictions. The idea behind ResponSea - and correspondingly the game - was to invoke interest in the maritime industry and to emphasize that sustainability includes all three categories, not only the environmental category.

The game was based on ResponSea, i.e. sustainable commitments of Finnish Maritime Industry openly available on a website (https://sitoumus2050.fi/en/). In addition to the general description of the commitment, each company has defined United Nations Sustainability Development Goals their commitment is related to, one to five sustainable actions they are going to take, indicators used for the follow-up of each action, as well as agreed to report on their progress of the action.

The open data fetch routine was designed to be as automatic as possible. Since Sitoumus2050 did not have an API, the data used in the game is collected by web scraping the website (https://sitoumus2050.fi/en/) to produce the file for the game. This is done frequently to enable the most recent information into the game. In each commitment, the SDGs are defined only in company level, and thus the data has to be manipulated to relate the SDGs to the action level. This is done only once for each action, and was done in co-operation with the coordination of ResponSea: each company was asked to choose a maximum of three SDGs they considered fitting to each of their actions. When this manual work was done once for each action, everything else could be automatic. While the fetch routine is automatic, the data quality can have various challenges: the companies inputted the data to the system manually, which causes syntactic or grammatical errors violating data quality as well as missing or incorrect items in the data.

The graphics design of the game had two main targets: it should not draw attention from the information content of the application and it should fit the quite simplistic style of the current ResponSea web page. For these two reasons, the style was selected to be pixel graphics.

Game Description and Game Design Elements

The *story* of the game is a cruise in a fictional M/S ResponSea. In the beginning, the player is in a lobby shown in Figure 1 and selects whether to start exploring the actions of the companies through an SDGs' corridor or a companies' corridor. By selecting the SDGs' corridor, the player moves to a corridor which has all seventeen SDGs as crossings to corridors. This view can be seen in Figure 2. By selecting an interesting crossing, the player moves to a new corridor where cabin doors represent all actions of different companies relating to that SDG. Similarly, if the player selects the companies' corridor in the beginning, he/she is presented crossings which have the names of all the participating companies. When the player enters the corridor of a specific company, he/she sees all the actions of that company as cabin doors. Then the player opens one of the cabin doors and enters the cabin which has more information about that specific action.

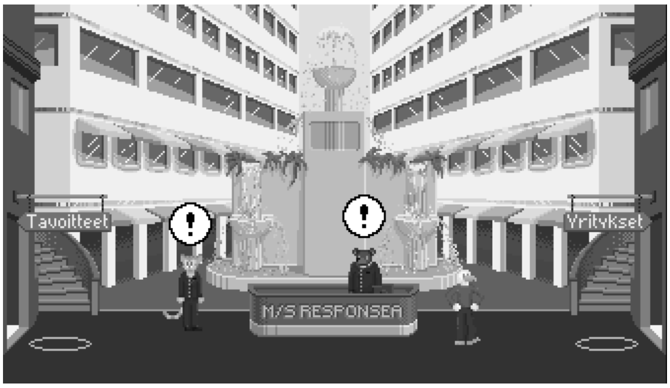

Figure 1. Starting lobby. SDGs' corridor on the left, companies' corridor on the right.

Figure 2. SDG's corridor, where the player can select which SDG related actions he/she wants to explore.

Figure 3 presents the cabin, where the player can explore the action. Here the actual open data is transferred to gamified and pro-environmental elements. *Information* is presented in textual form when the player opens television, including a title, a description of the action, as well as a description and start and target levels of the indicator related to the action. Although gamified elements were considered, the automatic fetch routine of the open data and the aim of not using manual work were the main reasons for selecting textual input. The cabin has up to three paintings each of which represent one SDG involved in this action. These can be considered as *badges* that visually share the information to the player. Viewing the information on the TV is obligatory and the player cannot continue the gameplay before reading it.

Figure 3. The cabin text includes the company name, and the action the company aims at. More information can be found by opening the television and the SDG paintings tell which SDGs the action is related to.

After exploring the action, the player continues the game by selecting one of the existing SDGs from the paintings or the company by exiting through the door and ends up back to the SDGs' corridor or the company corridor. The player can go back to previous corridor or even to the lobby. This way, the player continues to explore the actions of the companies.

When the player has explored a predetermined number of actions (five in the test setting), the game moves to the ending scene in Figure 4. The three company names are the top three companies the player was most interested in and the shipping containers represented the SDGs by the corresponding colour and as a portion to the overall SDGs in the chosen commitments. This way, the player is presented their own collection of their interests, which can be related to collecting *points* or *badges*.

Figure 4. End of the game scene. The companies the player is most interested in are presented as text, whereas the SDGs are presented as containers in the cargo ship.

GAME EXPERIENCES

In the one-day science event, 52 players tested the game, as presented in Table 1. According to the in-game data collected by the game, the median time for playing the game was 4 minutes 13 seconds, whereas the minimum and maximum time were 2 minutes 25 seconds and 8 minutes 36 seconds respectively.

Table 1. Information of the game testers.

	Female	Male	Unknown	Total
Under 11 years	0	6	0	6
11-15 years	1	18	2	21
16-20 years	13	5	0	18
Over 20 years	4	1	0	5
Unknown	1	1	0	2
Total	**19**	**31**	**2**	**52**

The In-Game Game Experience Questionnaire included a total of ten questions two from each of the selected components: positive affect, immersion, flow, negative affect and tension. Each question was a statement, where the player could answer using the five-step Likert scale (not at all, slightly, moderately, fairly, extremely). From these two questions, an average was used for each player in order to evaluate their experience in each component.

The averages and standard deviations of all players of the five components are presented in Table 2. The positive affect component consisted of questions about how good and content players felt while playing the game. The average for this component was between "moderately" and "fairly", meaning that most people enjoyed playing the game. Immersion and flow components are also such, that the better the score, the better the game has succeeded. The immersion component, which measures how impressive the game is and how interesting is its story, got its average a little over moderate. The flow of the game, i.e. how absorbed the player was in the game and did they forget everything around them, got in average a little bit smaller scores, i.e. under "moderately".

Table 2. In-Game Game Experience Questionnaire component averages and standard deviations of all players (1=not at all, 2=slightly, 3=moderately, 4=fairly, 5=extremely).

	Average	Standard deviation
Positive affect	3,39	0,91
Immersion	3,10	0,82
Flow	2,45	0,97
Negative affect	2,33	0,92
Tension	1,88	0,80

The negative affect and tension components are opposite to the other three components: The lower the scores, the better the results. The negative affect, i.e. how boring and tiresome the game was experienced, received on average score close to "little". The tension component, which measured frustration and irritation, was at an even smaller level.

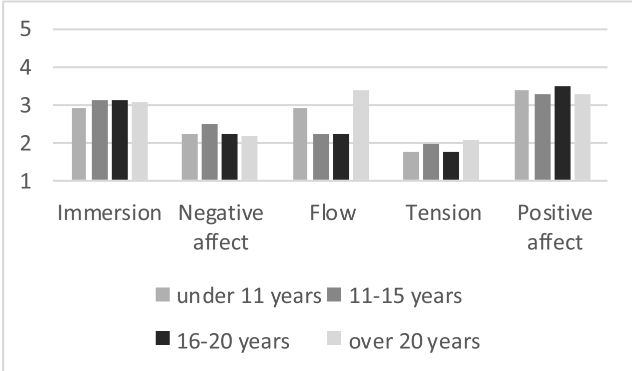

Figure 5. Averages of different age groups in each of the components.

Figure 5 presents the averages of each component in different age groups. Although most of the experiences in the age groups resemble each other, the flow component has clear differences, i.e. the youngest and the oldest seem to have better flow than the teenagers.

The aim of the game was to share information about sustainability actions including all three sustainability categories. According to Figure 6, this aim was quite well achieved, since on average the players thought they learned from all the sustainability areas moderately and only a few players thought that they did not learn at all.

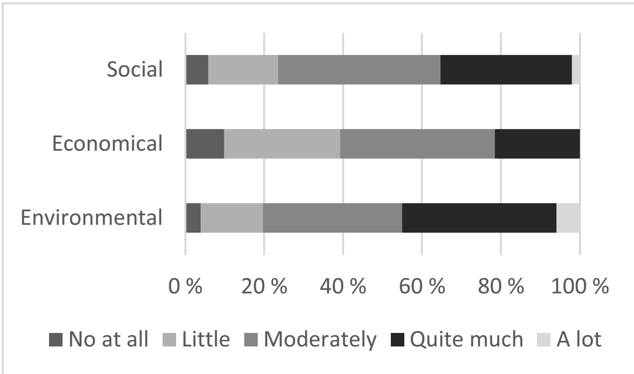

Figure 6. Which sustainability areas did you learn about in the game?

DISCUSSION

In the game design, similar components as were found in (Könnölä et al. 2018) were utilized. *Information* sharing was the main aim of the application, but also *commitment* was created through the usage of a *story* and *comparison* possibility was created by showing different coloured containers and names of the companies, which can be related to collecting *points*. Since the aim was not to change the sustainable behaviour, *feedback* and *rewards* were not present similarly to many other sustainability applications on scientific research.

Similar difficulties were encountered as other studies relating to data games (Friberg et al. 2013, Dunwell et al. 2016), which can be presumed to be a challenge for every game using open data. The data collection system was handled by two different parties, the website maintenance and companies adding commitments to it, causing it to have no validation regarding the quality of data. A decision was made not to alter the data in any way in between the game and the website. The data the companies varied substantially: for example, some companies had created long explanations of the carefully thought indicators, their start and target levels, whereas other companies had only entered the name of the indicator without start or target levels. This caused some data points to have insufficient information and could affect the learning and interpretation of the data by the player.

Although the game time on average was less than five minutes, the players considered the game positively and learned how three sustainability categories are present as actions in the Finnish Marine Industry. The goal of invoking interest in the sustainability of maritime industry was fulfilled. The environment, i.e. the noisy event, could have an effect on the results of immersion and especially flow, which had lower scores than the positive affect component in the Game Experience Questionnaire. A calmer environment could give other results. Mostly each age grouped seemed to have similar averages on the GEQ components, but a small difference was seen in the flow component. Since the youngest and oldest age groups were small in size, generalization cannot be made from this, but more data should be collected to understand the phenomena better.

CONCLUSIONS AND FUTURE WORK

In this research, an open data game for sustainability information sharing was developed and tested with mostly young users in a science event. The design decisions were well in-line in the current research on the design decisions of sustainability applications: information sharing is the key, that was coated with a story and finalized with the commitment of showing SDGs and companies the player was most interested in. Data quality played a great role in what kind of information the player received in the game. By improving the data quality, it could be possible to enhance the overall experience of the game and the impact of the information on the player. Despite the challenges with the data quality, the design choices were effective since the game was experienced positively according to the Game Experience Questionnaire. Also, the aim of the game, i.e. sharing sustainability information was accomplished as almost all the respondents thought that they learned about different areas of sustainability of maritime industry from the game.

In the future, the application should be further developed by also adding the reporting information. In addition, the player should be able to continue to explore the actions after ending the game. Since the test group was quite young, it would be interesting to extend the approach to older customer groups to cover a wider population and gain knowledge of experience differences between different age groups.

ACKNOWLEDGEMENT

The research reported in this article has been conducted as a part of SUSTIS (Sustainability and Transparency in Shipbuilding Networks) project. The project is carried out in collaboration with VTT Technical Research Centre of Finland, DNV GL Business Assurance, Evac, Lautex, Meriteollisuus, Meyer Turku, NIT Naval Interior Team, Paattimaakarit, Piikkio Works, Sininen Polku and SSAB Europe. The project is mainly funded by Tekes – the Finnish Funding Agency for Technology and Innovation.

REFERENCES

Brewer R. S.; N. Verdezoto; T. Holst; and M. K. Rasmussen. 2015. “Tough Shift: Exploring the Complexities of Shifting Residential Electricity Use Through a Casual Mobile Game”. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '15)* (London, UK, Oct. 5-7). ACM, New York, NY, USA, 307-317.

Froehlich J.; L. Findlater; and J. Landay. 2010. “The design of eco-feedback technology”. In Proceedings of the SIGCHI *Conference on Human Factors in Computing Systems (CHI '10)* (Atlanta, Georgia, USA, Apr. 10-15). ACM, New York, NY, USA, GRI, 2000.

Elkington, J. 1999. “Triple Bottom Line Revolution: Reporting for the Third Millenium”. *Australian CPA*. 69, 75.

Global Reporting Initiative, G4 Sustainability Reporting Guidelines.https://www2.globalreporting.org/standards/g4/Pages/default.aspx (accessed 3.6.2019).

Gustafsson Friberger, Marie; J. Togelius; A. Borg Cardona; M. Ermacora; A. Mousten; M. Møller Jensen; V.-A. Tanase; and U. Brøndsted. 2013. “Data games”. *In Proceedings of the Fourth workshop on Procedural Content Generation in Games*. ACM.

Hahn, R. and M. Kühnen. 2013. “Determinants of sustainability reporting: a review of results, trends, theory, and opportunities in an expanding field of research”. *Journal of Cleaner Production*,59, pp.5-21.

Hamari J.; J. Koivisto; and H. Sarsa. 2014. ”Does Gamification Work? A Literature Review of Empirical Studies on Gamification”. In Proceedings of 2014 47th Hawaii *International Conference on System Sciences* (Waikoloa, HI, USA, Jan 6-9). IEEE, 3025-3034.

IJsselsteijn, W. A.; Y.A.W. de Kort; and K. Poels. 2014. ”The Game Experience Questionnaire”. *Eindhoven: Technische Universiteit Eindhoven.*

Jylhä A.;P. Nurmi; M. Sirén; S. Hemminki; and G. Jacucci. 2013. “MatkaHupi: a persuasive mobile application for sustainable mobility”. In Proceedings of the 2013 *ACM conference on Pervasive and ubiquitous computing adjunct publication (UbiComp '13 Adjunct)*. ACM, New York, NY, USA, 227-230.

Kollmuss, A. and J. Agyeman. 2002. “Mind the Gap: why do people act environmentally and what are the barriers to pro-environmental behaviour?” *Environmental Education Research* Vol. 8, No. 3, 239–260.

Könnölä, K; T. Mäkilä.; K. Seppälä ; and T. Lehtonen. 2018. ”Motivational game-design and pro-environmental elements in sustainability applications”. GAME-ON’2018: 19th *International Conference on Intelligent Games and Simulations.* Dundee, UK.

Macklin, C.; J. Wargaski; M. Edwards; and K.Y. Li. 2009. “DATAPLAY: Mapping Game Mechanics to Traditional Data Visualization”. In *DiGRA Conference.*

Nguyen S. P. 2014. "Mobile application for household energy consumption feedback using smart meters: Increasing energy awareness, encouraging energy savings and avoiding energy peaks," *2014 International Conference on Collaboration Technologies and Systems (CTS)*. Minneapolis, MN, 2014, pp. 291-296.

Sailer, M.; J. Hanse; H. Mandl; and M. Klevers. 2013. “Psychological perspectives on motivation through gamification.” *Interact. Des. Archit.*, 19, 18–37.

United Nations. 2015. “Transforming our world: the 2030 Agenda for Sustainable Development”. https://sustainabledevelopment.un.org/post2015/transformingourworld (accessed 3.6.2019) .

Weiser P.; D. Bucher; F. Cellina; and V. De Luca. 2015. “A Taxonomy of Motivational Affordances for Meaningful Gamified and Persuasive Technologies”. *In Atlantis Press. Atlantis Press*, 271-280.

WEB REFERENCES

https://sitoumus2050.fi/en/

A SHIPPING SIMULATION THROUGH PATHFINDING: *SEL* WITHIN THE *MSP CHALLENGE* SIMULATION PLATFORM

Phil de Groot, Wilco Boode, Carlos Pereira Santos, Harald Warmelink, Igor Mayer
Academy for Digital Entertainment
Breda University of Applied Sciences
Monseigneur Hopmansstraat 2, 4817 JT BREDA, The Netherlands
E-mail: groot.p@buas.nl

KEYWORDS
Shipping Simulation, Maritime Spatial Planning, Simulation Gaming, Serious Gaming

ABSTRACT

The authors present the design of the shipping simulation SEL and its integration in the MSP Challenge Simulation Platform. This platform is designed to give policymakers and planners insight into the complexity of Maritime Spatial Planning (MSP) and can be used for interactive planning support. It uses advanced game technology to link real geo- and marine data with simulations for ecology, energy and shipping. The shipping sector is an important economic sector with influential stakeholders. SEL calculates the (future) impact of MSP decisions on shipping routes. This is dynamically shown in key performance indicators (e.g. route efficiencies) and visualised in heat maps of ship traffic. SEL uses a heuristic-based graph-searching algorithm to find paths from one port to another during each simulated month. The performance of SEL was tested for three sea basins: the firth of Clyde, Scotland (smallest), North Sea (with limited data) and Baltic Sea regions (largest, with most complete data). The behaviour of the model is stable and valid. SEL takes between 4 and 17 seconds to generate the desired monthly output. Experiences in 20 sessions with 302 planners, stakeholders and students indicate that SEL is a valuable addition to MSP Challenge, and thereby to MSP.

INTRODUCTION

The *MSP Challenge Simulation Platform* (henceforth MSP Challenge) is designed to give policymakers and planners insight into the complexity of Maritime Spatial Planning (MSP). It can be used for interactive planning support and general learning purposes. MSP is a process by which a country 'analyse[s] and organise[s] human activities in marine areas to achieve ecological, economic and social objectives' (European Union 2014), ending in a spatial plan. This spatial plan is essentially a highly annotated map of the sea area with spatial designations for specific human activities and marine protection measures for the medium-term future, often a period of 5 to 10 years. The MSP Challenge was first conceived and developed as a computer simulation game in 2011, and has been applied in sessions with MSP authorities, stakeholders and students many times since (Mayer et al. 2014, 2013; Stolte et al. 2013). Since early 2016, it has been further developed at Breda University of Applied Sciences within the context of the EU projects and consortia *NorthSEE, Baltic LINes* and *SIMCelt*. It has now become a platform allowing for all sorts of simulation game sessions: in different sea basins, with different data sources, and with different simulation models running in the background.

Shipping is an important sector to take into consideration for three reasons. First, it is one of the oldest and thus best-established sectors to use the seas and oceans. Second, the sector has been one of the key drivers of global economic prosperity by transporting goods and people all over the world (Ferreira et al. 2018). Third, it is legally a strong sector as well; freedom of navigation has for centuries been an important principle in international maritime law (Wolfrum 2008). For this reason, shipping has always been an important theme and consideration in MSP Challenge sessions, especially since recent technical and social developments are creating new offshore human activities (e.g. wind farms, aquaculture) or new environmental protection measures (e.g. Marine Protected Areas, MPA) which are directly impacting the shipping sector.

To better involve shipping in MSP, different governments and private companies developed spatial maps of specific sea regions (e.g. the North Sea region) showing ship movements over a certain period of time (Nilsson et al. 2018). These static 'heat maps' are developed using ship movement data captured through the ships' Automatic Identification System (AIS). The compendium of all ship movements generates heat maps which are very useful since they indicate the intensity of ships in specific areas over a specific time period. The maps can be used to recognise patterns, identify congestion areas, and evaluate risks, thus allowing planners to take important shipping information into account when they plan the use of sea space.

Although these maps offer a great utility for planners, they have no predictive power and do not allow MSP officials to forecast and develop different scenarios. While certain shipping patterns are generally constant (e.g. cargo or tanker vessels taking fixed routes and avoiding shallow waters), the influences of, for example, new wind farms or new traffic separation schemes introduced by the International Maritime Organization (IMO) are hard to grasp in a static map.

This is where the MSP Challenge could provide great value. MSP Challenge allows players to plan different scenarios for long periods of time (10 to 40 years), encouraging international discussion and cooperation to reach a coherent plan for an entire sea basin. However, the MSP Challenge cannot rely on static shipping information, since any new plan made and implemented will invalidate the information. An example would be planning a wind farm over an existing shipping lane. A dynamic and responsive shipping model would be much more useful and insightful.

Several shipping simulations already exist and could theoretically be reused in the MSP Challenge. Integrating (parts of) any of the following existing simulators is technically possible:

1. MARIN's Vessel Traffic Service simulates individual ships over the course of a couple of hours. It is mainly used for training harbour personnel.
2. Sea Traffic Management, rolled out at several locations affiliated to the European Maritime Simulator Network, simulates individual ships in often particularly busy areas to test and teach novel traffic management technologies and techniques and thus optimise routes and reduce risks.
3. SEATRAS simulates sea traffic in particularly congested areas to e.g. enable calculations of collision risks and tests of collision avoidance technologies and techniques (Itoh et al. 2003).

However, when evaluating these existing solutions, we were concerned with the following:

1. None of these simulation goals are well-aligned with ours. The simulations are created for other purposes than those of MSP Challenge. They offer some functionalities that we could use, but need some functionalities that we would still need to develop.
2. Assuming we could adapt the existing simulations to fit our needs, MSP Challenge would simultaneously also handle dozens of large-scale data layers, as well as a simulation of offshore energy production and distribution (Hutchinson et al. 2018), and of ecosystem dynamics (Steenbeek et al. 2019). We thus require an efficient, well-targeted shipping simulation to keep system requirements at levels acceptable for our sessions and target audience.

We therefore decided to explore how we could create our own shipping simulation. In this paper, we offer an answer to the question of *how a convincing shipping simulation can be designed and implemented within the MSP Challenge, allowing for players to make MSPs that could include shipping measures and showing players within a reasonable timeframe the effects of their plans on ship traffic.*

We answer this question by explaining the design, implementation and results of the shipping simulation *SEL* (Shipping Emulation Layers) within the MSP Challenge. The bulk of this work took place over almost one year, involving co-design with shipping experts, programming and extensive testing and application in three sea basins through 20 MSP Challenge sessions since 2018. In the remainder of this paper we first formalise the requirements that SEL needed to fulfil, before we explain how SEL solves this pathfinding problem efficiently yet realistically.

FORMALISING REQUIREMENTS

The MSP Challenge platform architecture and desired shipping functionality led us to define a number of requirements for input, output and throughput of the shipping simulation. We explain the platform architecture and our chosen requirements in this section.

MSP Challenge Architecture

MSP Challenge is a data-driven client-server platform, enabling sessions with different scenarios, regions and time frames (Figure 1). The platform uses advanced game technology to link real geo- and marine data with simulators for specific maritime sectors, mainly ecology, energy and shipping. These simulators are satellite applications interconnected with the game server. They add dynamic data to the game on the levels of ecology, energy and indeed shipping. They have a discrete-event architecture, with each discrete event representing one simulated month. A single game client can act as a player or game master and connects to the server, which is responsible for maintaining the current game state and interfacing with the simulations. The actual time between each discrete event is defined by the game master and depends on how long they want the entire session to last. In this manner, the MSP Challenge simulates MSP in up to four rounds of planning and simulation, each round representing as many years as the game master defines.

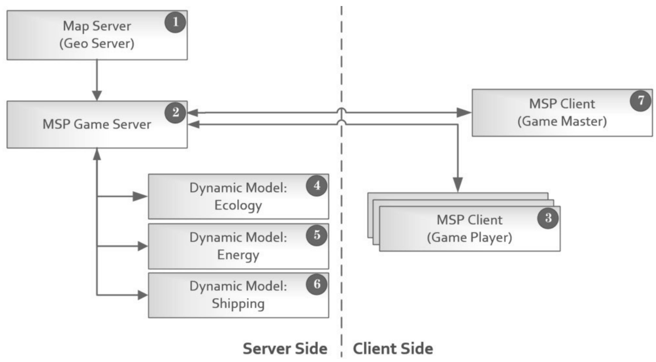

Figure 1: MSP Challenge high level architecture

A typical MSP Challenge session takes at least half a day, during which around 30 players are grouped into 5 to 9 country teams. They design and implement at least 20 independent maritime spatial plans that each alter any of the roughly 40 data layers, and analyse and evaluate resulting key performance indicators on the levels of ecology, energy, and indeed shipping. For players to analyse and evaluate results, the MSP Challenge needs to be able to obtain and pass through data of each month reasonably quickly.

The different background simulations must read the maritime spatial plans defined by all players, and calculate and feed back the combined results and consequences. Obviously, the quicker the simulation can do this, the better. Yet, if the shipping simulation would take between 5 to 10 seconds of computation per month, this would translate to around 15 minutes per 10 years. This is an acceptable performance level, as it still fits the typical dynamic of an MSP Challenge session.

Shipping Emulation Layer Objectives

We defined SEL's objective as to generate reasonably realistic ship intensity information per discrete event (each simulated month) for a gameplay period of several decades. The information needed to be split over several different types of ships, each with different behaviour: cargo, tanker, maintenance, passenger and ferry ships.

The primary output of SEL needed to consist of rasterised heat maps showing the intensity of ship traffic in the simulated area. Given multiple ship types, multiple heat maps would need to be generated. The outputted heat maps would need to be shown in the MSP Challenge client, but would also need to be integrated into other simulations. Particularly, we would require the ability to define from ship intensity information particular ecological pressures that we could then feed into the ecosystem simulation MEL (Steenbeek et al. 2019).

Furthermore, SEL also needed to output certain key performance indicators (KPIs). KPIs give insight into certain aspects of the sea basin state, and are typically highly contextual quantifications. Interpretations of whether or not changes in KPI are improvements or setbacks is part of the game and thus up to the players. Some KPIs are directly influenced by players and others are more for informational purposes. Three main KPI types were defined for SEL to generate:

- The number of ships per port in the simulated month;
- The routing efficiency between any two ports (actual route distance compared to the rectilinear distance);
- The amount of ships travelled over a shipping lane.

As a third and final output, SEL also needed to create shipping routing issues. When SEL was unable to find a route between two points, it should report an issue to the MSP Challenge platform. The issue is then shown to the players in the game client indicating that possibly one of their plans has created a problem for shipping and needs to be investigated. For example, players might create plans which define restriction zones that prevent specific ships to reach destination areas or ports.

SEL Input Data

As input SEL would firstly need all the data in the MSP Challenge sea basin in question to obtain a representation of the simulated world. The server divides data into certain data layers, where data layers can contain any number of planned geometry instances. The data layers of the MSP platform can be classified as:

- *Constant data*. This is data that cannot be edited by players or other dynamic models while a session is running. This data is only requested and fed into the simulation upon startup. An example of static data would be of the landmass or bathymetry (sea depth).
- *Dynamic data*. This is data that may change throughout the session by players' actions, or as the result of another simulation. When a layer changes, it is flagged on the server and will be re-acquired and fed into SEL at the next discrete event (i.e. the next simulated month). Anything planned by players themselves, such as shipping routes, is dynamic data.

The following data is subsequently interpreted by SEL in particular ways:

- *Shipping lanes*. These are route segments in the sea basin. Ships might prefer to take such routes because it is, for example a mandatory route, company policy, or safer. Designated shipping lanes are mostly ship type specific and only present in busy and/or otherwise risky areas. Thus they never comprise complete routes from port to port, but are segments scattered over a sea basin.
- *Ports*. Ports are considered producers and consumers of ship intensity, and are defined as point geometry. Each port has relevant metadata, such as the available fuelling types, port facilities, and the expected number of vessels (per type) arriving or departing per simulated month.
- *Restriction areas*. These either block pathing for all or some ship types. An example of restriction geometry would be the landmass layer which blocks pathing completely for all ship types. Ship traffic separation areas, aquaculture and wind farms are other examples of restriction areas taken into consideration when pathing.

As described above, there is specific metadata behind each port. It was a game design decision to keep the number of vessels "generated" by each port configurable per game session, and allowing the game master to tweak the number of vessels per port before the game session starts. This way different scenarios can easily be configured.

SEL Ship Navigation Considerations

In order for SEL to find realistic paths, we defined the following common ship navigation considerations:

1. *Freedom of navigation and basic economics*. Under international maritime law, ship captains can, in principle, choose their own paths. Ideally, they would choose the most direct and thus most efficient path.
2. *IMO route adherence.* Shipping companies have been following predetermined routes for safety reasons for over a hundred years. Nowadays, traffic separation schemes and shipping routes are the responsibility of the IMO, regulating congested sea areas in the world. An IMO designated route has a strong legal status and the benefit of increased safety in particularly busy areas. We take some exceptions to these rules into account (see final point) but other than those the simulated ships will follow IMO routes.
3. *Obstructions*. Certain obstructions do not allow specific vessels to enter specific areas, notably:
   a. Specific ships are not allowed to go through human-made structures (notably wind farms, oil and gas installations).

   b. Ship restrictions may be in place, for example for traffic separation, during wind farm construction, or no-shipping zones.
   c. A shallow area can represent an obstruction to bigger cargo and tanker vessels requiring deeper waters. This is implemented through a type-specific penalty system. Larger ships will have a larger penalty than smaller ships. We cannot treat the shallowest waters (0 - 20 metres) as no-go zones as ships will need to cross these depths to get to certain ports.
4. *Ship type differences*. Although generally each ship wants to take the most direct route possible, there are noticeable differences between each ship type:
   a. Tankers and cargo vessels will try to follow the defined (IMO) shipping lanes as long as they do not create too big a detour.
   b. Ferry ships will take the most direct route regardless of whether there are shipping lanes it can use.
   c. Maintenance ships construct and maintain offshore man-made structures. They are the only type allowed to go to and cross over any offshore energy areas (notably wind farms). Maintenance ships are normally smaller, will always take the most direct route, and always originate from the port closest to the port with maintenance facilities.

SEL'S ARCHITECTURE

With all requirements described, in this section we explain the approach we took for building SEL. We specify how each simulated month the simulation finds the paths for all ships to generate the desired outputs.

Pathfinding

The simulation uses a heuristic-based, graph-searching algorithm to find paths between two points on the internal graph (A*) (Hart et al. 1968). There are three main steps to take the data provided by MSP Challenge and transform it into a usable structure for pathfinding.

Step 1: Connection Graph Setup
SEL internally builds a complex graph from the geometry data received from the game server. Port information is added as graph vertices. Similarly, all the geometry points defined in the shipping lane layers are added as graph vertices. All shipping lane connections between points are added as edges on the graph. The rest of the geometry is converted into restriction areas, forming rules that are reviewed when generating the rest of the connectivity graph.

We created a separate layer invisible to players with a set of points in a grid pattern on navigable areas. These grid points create more graph vertices for populating the graph and supporting the pathfinding algorithm in finding alternative paths when required. Moreover, they define the alternative path's resolution, important for defining the degree of resolution for our heat maps.

To simulate the different ship navigation behaviours we implemented a system of rulesets working with restriction zones. Depending on the ruleset configuration we can force specific shipping routes to be very strict and have ships always take them if possible, or let them be more flexible to allow ships to take the shortest paths available. As a result of the A* heuristic function that is used the different routes that SEL calculates are usually sub-optimal in terms of distance travelled, but closer to the real-world results.

Step 2: Route Calculation
Once the input data is set up for the graph we start connecting the entire graph together by creating more edges. SEL loops over every vertex in the graph and connects it to the closest neighbours that are within a certain direction, as long as there is no blocking restriction geometry in between.

To give an example, for every vertex in our graph a connection is made to the closest navigable vertex north, east, south and west of it, as long as there is no blocking restrictions in between. These edges that are created are marked as being *implicit* edges, as opposed to the *explicit* edges which are shipping lanes defined by the data. We uphold this difference between *implicit* and *explicit* edges for the pathfinding algorithm. In the pathfinding implementation travelling over *implicit* edges incurs a configurable cost penalty. This cost penalty influences the pathing in such a way that we can control how likely the ships are to adhere to *explicit* edges (official shipping lanes, e.g. IMO routes) by increasing and decreasing the penalty.

Each edge also stores information about what ship types are allowed to use it and with which direction. The restriction zones the edge crosses influences the types of ships allowed to cross the edge. By default all ships are allowed to path over all edges, but this is changed when an edge is created over a restriction zone that only allows a subset of ship types. The edge copies the allowed ship types from the restriction zone it crosses. The directionality setting restricts in what direction the edge can be crossed and can be set to *unidirectional* (only from start to end) or *bidirectional* (either way) for every edge. This mimics IMO traffic separation schemes.

For finding paths, we query the constructed graph using an A* algorithm that takes into account the edges the ship type can cross and respects the directionality of the edge. Depending on the ship type configuration *implicit* edges are penalised by using a cost multiplier for crossing that edge. Additionally, restriction geometry can specify cost multipliers to make ships only cross the geometry when alternatives are either not found or significantly more costly. A usage example of this restriction geometry penalty is the bathymetry layer. The 0 - 20 metre depth bathymetry layer specifies a large cost penalty for crossing the layer by large ships. This causes large ships to avoid coastal areas that are shallow unless they need to cross it to get to a harbour.

During the pathfinding calculation stage we cache all created routes. Before we calculate a new path, we check if there is a path that already matches our requirements of source, destination, ship type and directionality. For instance routes from point A to B can be reused for paths from point B to A provided that they do not contain any *unidirectional* edges and allow for the required ship type.

Step 3: Storing the Connection Graph.
The connection graph generated in Step 1 and the paths generated in Step 2 are stored for use in future calculations. By keeping the data, we can significantly reduce the required calculation time for months that do not influence any layers that affect shipping. The graph and routes are completely discarded and recalculated when one of the layers taken into account for the shipping simulation changes.

Heat Map Generation

Generating heat maps is an important step for our implementation. This takes the abstract data of shipping intensities over a route to data that can be visualised in the form of a heat map. The process consists of three steps.

First, SEL generates an unbounded raster of intensity values. This starts with a two-dimensional array of a size equal to the final output image initialised with 0 values. For every route that was calculated the algorithm walks the edges that make up the route. We project the edge onto the raster using a line rasterisation method (Bresenham 1965). When we use this method, every cell the edge crosses has the cell's value increased by the intensity of that route. The rasterised data obtained (Figure 2) contains very sharp results of the actual intensity values for each pixel on the simulated raster.


Figure 2: Unbounded intensity map

Second, SEL creates a raster mask defining what areas are inaccessible to each ship type. These images are either fully black or fully white, where all pixels covered by an unpassable restriction zone are white (Figure 3). This mask serves a purpose in the next step, ensuring that we do not blur ship traffic over areas where ships are not allowed to go, for example over land.

Third, SEL blurs out the values to the intended display range. It needs to flatten the unbounded grid values in our heat map to values that we can represent as an image. We use a modified Gaussian convolution matrix as an image filtering technique to spread the intensity values that exceed the chosen maximum (Fisher, Wolfart, and Wiley 1996). If the heat map is configured to contain e.g. max 10 intensity and the unbounded heat map has values of 50, that value should be smeared out over adjacent pixels (Figures 4 and 5).


Figure 3: Restriction map used in shipping rasterisation

0	0	0
0	50	0
0	0	0

→

0	10	0
10	10	10
0	10	0

Figure 4: Gaussian convolution filter example and how the intensity values are distributed to adjacent cells


Figure 5: Unbounded heat map (left) with Gaussian convolution filter applied (right)

To increase the accuracy of the Gaussian blur, our implementation takes into account the restriction mask from the previous step to know where it is allowed to put intensity. When a pixel is marked as unavailable in the mask, the blur kernel weights are adjusted to compensate, ensuring that we do not lose intensities from moving them around.

Key Performance Indicators

The key performance indicators that SEL calculates are derived from different data already present after calculating the routes. Three main KPI categories are calculated:

1. The number of ships a port produced in the simulated month. We derive the number of ships of a certain type that a port produces from the input data defined by the scenario. This value is the actual number of ships that are sent over a particular route to a destination, and provides an insight into port development.
2. Per-port routing efficiency percentage. For each port we examine each route, and divide the length of the route and by the rectilinear distance between origin and

destination. This single value fed back is the average efficiency of all routes from the particular port.

3. The amount of ships travelled over a shipping lane. For every lane we track which source geometry it belongs to. Source geometry is only defined for *explicit* shipping lanes. SEL goes over all available routes and all of the edges that make up that route. If an edge is an *explicit* shipping lane, then SEL adds the route intensity to it. The sums of these intensities are incorporated into each shipping lane's metadata.

PERFORMANCE, OPTIMISATION, VALIDATION

In this section we evaluate SEL and the challenges of keeping the simulation running as fast as possible while offering a wide variety of player options and maintaining accuracy to keep the simulation believable. To check and increase SEL's accuracy, we compared the SEL generated maps to shipping intensity data we acquired for three regions: Firth of Clyde, North Sea and Baltic Sea.

Firth of Clyde

The Firth of Clyde is a relatively small sea basin to the west of Glasgow, Scotland. Marine Scotland provided us with real-world AIS data and resulting heat maps for this marine region. In Figure 6 we compare the Marine Scotland heat map to SEL's generated heat map as viewed from within MSP Challenge. The MSP Challenge play area for the Firth of Clyde is just the Clyde estuary, so all the shipping intensity outside was not considered for the simulation, but it is possible to clearly observe the similarities.

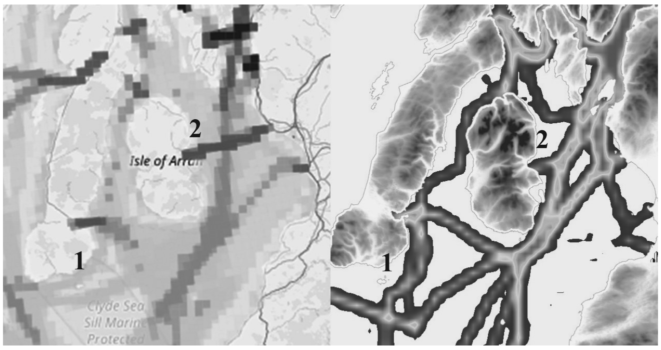

Figure 6: Firth of Clyde - Marine Scotland left, SEL right

We note the following differences between the two maps:

- There is a small island (1 in Figure 6) in the lower left quadrant of the image where the real-world data shows shipping intensity going north of the island, while SEL outputs the intensity further south. This is a result of the resolution of the pathing graph.
- The real-world data shows there is a line between the Isle of Arran (2 in Figure 6) and the Scottish mainland which is very busy. In SEL, this line is completely absent. Real-world maps show there is a ferry route at that exact intensity line. This is because the provided source data did not include this particular ferry route's number of ships per month or geometry.
- SEL noticeably distributes shipping over more separate lines than we see on the real-world heat map. This is a result of our pathing algorithm implementation.

Over the course of 2017, several Marine Scotland MSP professionals were involved in this implementation and evaluated the developments and final results. They deemed the results close enough to reality and representative enough for the region to render it useful for MSP Challenge sessions oriented towards education, training and stakeholder engagement. In early 2018, the simulation was applied in two MSP Challenge sessions with a total of 21 participants, both successful in their respective objectives of stakeholder engagement and higher education.

North Sea

The North Sea is a much larger sea basin in Europe, known for its heavy traffic. We acquired total shipping intensity data and heat maps for this sea basin concerning the period July 2016 - July 2017 from the Havbase website. In Figure 7 we compare the Havbase heat map to SEL's generated heat map as viewed from within MSP Challenge.

Figure 7: North Sea - Havbase left, SEL right

We note the following differences between the two maps:

- Due to the large amount of energy facilities, SEL generates a large amount of maintenance ships travelling to and from them. This is particularly visible to the right of Scotland, with its many oil and gas installations (1 in Figure 7). In this case this seems to be quite similar to the real world. However, SEL also does this for wind farms. While this is in itself realistic, we do not see these maintenance intensities in the same manner and with the same ports of origin in the real-world data.
- In the northern part of the Netherlands, at the Den Helder port (2 in Figure 7), there is a routing error that causes ships to go around the island of Texel before faring into the sea basin. This is an issue caused by the way that we treat bathymetry-based cost penalties.
- Like in the Firth of Clyde, SEL distributes shipping intensities much more on separate lines than can be seen on the real-world map (3 in Figure 7). This is again a result of our pathing algorithm implementation.

Over the course of 2017 and 2018, we worked extensively with several maritime professionals within the NorthSEE partnership to get to this implementation. They generally deem the results valuable, although they also tend to quickly point out the aforementioned differences with the real world. Since 2018 we have applied the simulation in 15 MSP Challenge sessions with a total of 215 participants, all successful in their objectives of stakeholder engagement, planning support and higher education.

Baltic Sea

The Baltic Sea is an even larger sea basin in North-Eastern Europe, officially spanning the Kattegat, Baltic Proper, Bothnian Sea and Gulf, Gulf of Riga, and Gulf of Finland marine regions. In this case the Baltic Marine Environment Protection Commission - Helsinki Commission (HELCOM) provided us with real-world AIS data and resulting heat maps concerning every month of 2016, split over our (and more) ship types. This was as yet the most complete dataset we were able to obtain for an area this large. In Figure 8, we compare the HELCOM heat map to SEL's generated heat map as viewed from within MSP Challenge.

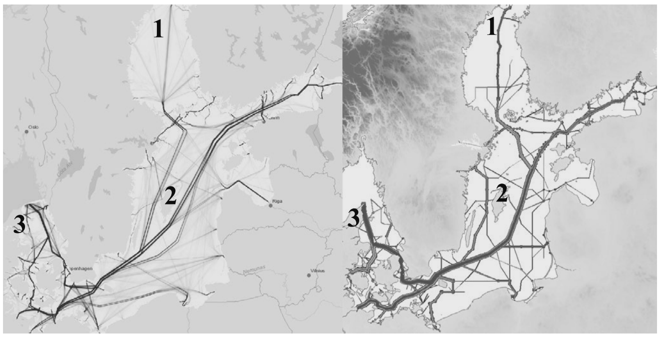

Figure 8: Baltic Sea - HELCOM left, SEL right

We note the following differences between the two maps:

- Missing data from ports in the northern Bothnian Gulf (1 in Figure 8) created a noticeable difference in ship dispersion.
- The large island of Gotland (roughly in the middle of image) allows ships to go past both sides of the islands (2 in Figure 8). To avoid congested areas, most of the times ships will take alternative routes in the real world. In this case, that would mean they divert to going north of the island to avoid congested areas at a cost of a (slightly) longer route. In our model alternative routes are not considered when the area reaches a specific density threshold.
- There is a significant difference in the position of shipping intensity at the Kattegat entry and exit area in the west (3 in Figure 8). Again this is attributed to the fact that our model does not include congestion.

Over the course of 2017 and 2018, we worked extensively with several maritime professionals from HELCOM within the context of the Baltic LINes partnership to get to this implementation. Similar to the NorthSEE partners, they generally deem the results valuable, although they also tend to quickly point out the aforementioned differences. Since 2018 we have applied the simulation in three MSP Challenge sessions with a total of 66 participants, all successful in their objectives of stakeholder engagement and planning support again.

SEL Performance

As can be imagined, the processing times of the three implementations differ highly. For the Firth of Clyde dataset (Figure 6) it takes around 4 seconds for the initial processing step to complete, resulting in around 3,000 routes. For the heaviest Baltic Sea dataset (Figure 8) this initial processing step takes around 17 seconds to complete, resulting in over 33,000 routes.

Each simulation step after the first takes less time since we can re-use data that we previously processed. When one or more of the *dynamic* data layers have been changed that SEL uses, subsequent steps for the Firth of Clyde take around 2.5 seconds and for the Baltic Sea around 13 seconds to complete. When no data layers are changed, the simulation time is reduced to less than one second for any of the three implementations.

The two parts of the simulation that currently take up the most time are building the pathing graph and calculating the routes, approximately 30% and 50% of total processing time, respectively when measured on the Baltic Sea data. The simulation time that we are currently able to achieve is slightly above MSP Challenge targets. Optimising the simulation, while keeping the same level of accuracy for the results, is an ongoing task. Still, the performance we are able to achieve is currently not a bottleneck in any of the sessions we are running.

A performance improvement that we can still implement with the current SEL architecture is to locally rebuild the pathing graph. Currently, if a data layer changes that SEL uses the entire graph is discarded and rebuilt. When we use the new graph, all paths are recalculated to ensure the routes are still valid. Instead of rebuilding the entire graph, we could invalidate a smaller portion of the graph and only rebuild that area. This has the potential to drastically lower the rebuild times of the graphs. Determining which paths need to be rebuilt is still a challenge as a change might open up shortcuts that were not possible before.

CONCLUSION

In this paper we have discussed how we created a convincing shipping simulation which can perform its calculations in an acceptable period of time. The simulation we outlined works with, and responds to, dynamic data fed into it by the MSP Challenge server. We presented how we approached the implementation of shipping simulation as a graph-based pathfinding problem, and how we rasterised this graph data to build a convincing heat map by use of several techniques, including a modified Gaussian blur, for use within the game environment.

We implemented this simulation in three marine regions: Firth of Clyde, North Sea, and Baltic Sea. In all three implementations we worked with shipping professionals to understand the shipping logic and obtain shipping intensity data from the region. We determined that with these three highly diverse regions the simulation is able to run within a small timeframe (4 - 17 seconds per simulated month).

We applied all three regions in a total of 20 formal MSP Challenge sessions successfully reaching their objectives of stakeholder engagement, planning support, and higher education. We incorporated feedback obtained during each application to improve the simulations for later sessions, and identify even further improvement potential. The provided outputs are nonetheless very suitable for representing ship navigation behaviour, and for keeping players engaged and thinking about (in)direct impacts of their plans on shipping.

We continue to improve the accuracy of the simulation. The first improvement that we will address concerns how we treat bathymetry cost penalties. Currently the cost penalties are incurred every time a ship crosses the line between deep and shallow water, while no cost is incurred as long as the ship is within shallow water. This is a naive approach, but works well for a large part of the shipping routes. There are a couple of routes where this nonetheless leads to unrealistic path segments. If we constantly incur cost penalties while a ship is within shallow waters, this might improve the accuracy of the paths. Moreover, as seen in the Baltic Sea shipping intensity data, there are several inaccuracies that can be attributed to our simulation not taking congestion into account. Implementing congestion into the simulation to have ships avoid congested areas is another accuracy improvement with high potential.

ACKNOWLEDGEMENTS

The research leading to these results acknowledges the contribution of the NorthSEE (2016-2020), Baltic LINes (2016-2019), SIMCelt (2016-2018) and SP-MSP (2016-2019) projects, which received funding from the European Union's Interreg North Sea Region (NSR) and Baltic Sea Region (BSR) programmes (both part of the European Regional Development Fund), the European Maritime and Fisheries Fund (EMFF), and the Erasmus+ programme, respectively.

REFERENCES

Bresenham, J. E. 1965. "Algorithm for Computer Control of a Digital Plotter." *IBM Systems Journal* 4, No. 1, 25-30.

European Union. 2014. "Directive 2014/89/EU of the European Parliament and of the Council of 23 July 2014 Establishing a Framework for Maritime Spatial Planning." *Official Journal of the European Union* 2014, 135-45.

Ferreira, D.C.; R.C. Marques; and M.I. Pedro. 2018. "Explanatory Variables Driving the Technical Efficiency of European Seaports: An Order-α Approach Dealing with Imperfect Knowledge." *Transportation Research Part E: Logistics and Transportation Review* 119, 41–62.

Fisher, R.B.; E. Wolfart; and J. Wiley. 1996. "Hypermedia Image Processing Reference." In *Hypermedia Image Processing Reference*, 156–60. John Wiley & Sons Ltd.

Hart, P.E.; N.J. Nilsson; and B. Raphael. 1968. "A Formal Basis for the Heuristic Determination of Minimum Cost Paths." *IEEE Transactions on Systems Science and Cybernetics* 4, No. 2, 100-107.

Hutchinson, K.; H.J.G. Warmelink; W. Boode; C. Santos; and I.S. Mayer. 2018. "An Offshore Energy Simulation through Flow Networks: CEL within the MSP Challenge 2050 Simulation Game Platform." In *The 32nd Annual European Simulation and Modelling Conference*. Ghent, Belgium: EUROSIS.

Itoh, H.; M. Numano; and E. Pedersen. 2003. "Modelling and Simulation of Sea Traffic and a Visualization-Based Collision Avoidance Support System." In *Modsim 2003: International Congress on Modelling and Simulation*, edited by D. A. Post, 1996–2001. Modelling and Simulation Society of Australia and New Zealand.

Mayer, I.S.; Q. Zhou; X. Keijser; and L. Abspoel. 2014. "Gaming the Future of the Ocean: The Marine Spatial Planning Challenge 2050." In *Serious Games Development and Applications*, 150–62. Lecture Notes in Computer Science. Springer, Cham.

Mayer, I.S.; Q. Zhou; J. Lo; L. Abspoel; X. Keijser; E. Olsen; E. Nixon; and A. Kannen. 2013. "Integrated, Ecosystem-Based Marine Spatial Planning: Design and Results of a Game-Based, Quasi-Experiment." *Ocean & Coastal Management* 82 (Supplement C): 7-26.

Nilsson, H.; J. van Overloop; R.A. Mehdi; and J. Pålsson. 2018. "Transnational Maritime Spatial Planning in the North Sea: The Shipping Context." Research Report NorthSEE Project.

Steenbeek, J.; G. Romagnoni; J. Bentley; J.J. Heymans; N. Serpetti, M. Gonçalves; C. Santos, et al. 2019, under review "Combining Ecosystem Modelling and Serious Gaming to Aid Transnational Management of Marine Space."

Stolte, W.; X. Keijser; A. de Kluijver; and A. J. Nolte. 2013. "A Conceptual Pressure-Response and a Simple Food Web Model for the Maritime Spatial Planning Challenge 2050 Serious Game." 1208167-000-ZKS-0007. Delft: Deltares.

Wolfrum, R. 2008. "Freedom Of Navigation: New Challenges." In *Freedom of Seas, Passage Rights and the 1982 Law of the Sea Convention*, edited by Myron H. Nordquist, Tommy Koh, and John Norton Moore, 79-102. Brill.

WEB REFERENCES

www.balticlines.eu
www.havbase.no
www.imo.org/en/About/Conventions/ListOfConventions/Pages/default.aspx
maps.helcom.fi
www.marin.nl/web/Facilities-Tools/Simulators/Simulator-Sales/Vessel-Traffic-Service-VTS-Simulators.htm
marinescotland.atkinsgeospatial.com/nmpi
www.mspchallenge.info
www.northsee.eu
www.simcelt.eu
stmvalidation.eu/projects/stm-validation/

BIOGRAPHIES

PHIL DE GROOT, **WILCO BOODE** and **CARLOS SANTOS** are all part of the Breda University of Applied Sciences (BUas) games and media research and development team *Cradle*, functioning as core developer, lead designer, and technical manager of the MSP Challenge Simulation Platform, respectively. **HARALD WARMELINK** is a senior project leader at BUas for MSP Challenge and other games & media R&D projects. **IGOR MAYER** is professor of serious and applied gaming at BUas and functions as the MSP Challenge director.

AUTHOR LISTING

AUTHOR LISTING