Predictive Psychological Player Profiling

Ahmad Azadvar
Malmö University
Faculty of Technology and Society
Department of Computer Science and Media Technology

Studies in Computer Science
Faculty of Technology and Society
Malmö University
ABSTRACT

Video games have become the largest portion of the entertainment industry and everyday life of millions of players around the world. Considering games as cultural artifacts, it seems imperative to study both games and players to understand underlying psychological and behavioral implications of interacting with this medium, especially since video games are rich domains for occurrence of rich affective experiences annotated by and measurable via in-game behavior. This thesis is a presentation of a series of studies that attempt to model player perception and behavior as well as their psychosocial attributes in order to make sense of interrelations of these factors and implications the findings have for game designers and researchers. In separate studies including survey and in-game telemetry data of millions of players, we delve into reliable measures of player psychological need satisfaction, motivation and generational cohort and cross reference them with in-game behavioral patterns by presenting systemic frameworks for classification and regression. We introduce a measurement of perceived need satisfaction and discuss generational effects in playtime and motivation, present a robust prediction model for ordinally processed motivations and review classification techniques when it comes to playstyles derived from player choices. Additionally, social aspects of play, such as social influence and contagion as well as disruptive behavior, is discussed along with advanced statistical models to detect and explain them.

Keywords. Human-Computer Interaction, Affective Computing, Player Experience, User Research, Behavioral modeling, Psychology of play
LIST OF PUBLICATIONS

Included Papers


Related papers but not included in the thesis.


**Personal Contribution and Clarification**

PAPER I, is in review for proceedings of the CHIPLAY 2021 Conference.

In all publications, main author bore the responsibility of communication and planning of the paper’s authorship, although in PAPER III this task was equally divided between the first three contributors. PAPER I, IV and V are a result of coordinated efforts of several researchers and institutions and first two authors played the coordination role for those publications.
ACKNOWLEDGEMENT

I would like to acknowledge and express my gratitude to contributors, researchers and supervisors who made this collection possible.

I would like to extend my gratitude to my parents, friends, and family as well as game industry professionals, the University staff, players, and others who supported and aided me throughout this process.
# CONTENT

ABSTRACT ........................................................................................................ III

LIST OF PUBLICATIONS ................................................................................ IV

  Included Papers .................................................................................. IV

  Related papers but not included in the thesis. ................................ IV

  Personal Contribution and Clarification ......................................... V

ACKNOWLEDGEMENT .............................................................................. VI

CONTENT ................................................................. VII

  Acronyms ..................................................................................... XII

Part I. COMPREHENSIVE SUMMARY .................................................... 1

I INTRODUCTION .................................................................................. 2

  Problem Statement ......................................................................... 3

  Research Questions ....................................................................... 6

  Pronouns, Style, and Clarification .............................................. 8

II BACKGROUND ................................................................................... 9

  1. Measuring Experience ............................................................ 9

  2. Psychological Need Satisfaction ......................................... 12

  3. Profiling and Modelling Behavior .......................................... 14

  4. Social Influence ....................................................................... 20

  5. Disruptive Behavior .............................................................. 23

III RESEARCH METHODOLOGY ....................................................... 30

  1. Research Elements .............................................................. 30

  2. Methods Explanation ........................................................... 31

IV METHODS ....................................................................................... 36

  1. Correlations and Regression .................................................. 36

  2. Scale Reliability Measures .................................................... 37

  3. Factor Analysis ................................................................. 38

  4. Comparing Means .............................................................. 40

  5. Social Network Analysis (SNA) ............................................. 42
6. Clustering methods .......................................................... 46
7. Predictive Analysis ............................................................ 48
8. Ensemble models ............................................................... 49

V RESULTS ........................................................................ 50
1. Questionnaire Design and Validity ......................................... 50
2. Cohort Specifications .......................................................... 53
3. Modelling Motivation .......................................................... 55
4. Playstyle Classification ....................................................... 61
5. Social contagion and Influence ............................................. 67
6. Detecting disruptive behavior ................................................. 70

VI CONTRIBUTIONS .......................................................... 72
RQ1) What are the typical approaches of studying one’s placement in consumer society and its relationship to media consumption? .......................... 73
RQ2) How is the study of -predictive- behavioral models achieved? .......... 75
RQ3) What would be the interrelation and appropriate classification of historical, political, and developmental indices of a gamer and their respective behavior? ............................................. 76
RQ4) What additional information could -a longitudinal- modelling of player-behavior based on their psychosocial profile provide us? ........ ............ 77

VII CONCLUSIONS AND FUTURE WORK ................................. 78
1. Measuring Experience ......................................................... 78
2. Cohort differences ............................................................. 79
3. Modelling Motivation .......................................................... 82
4. Playstyle Classification ....................................................... 84
5. Social Contagion and Influence ............................................. 86
6. Disruptive Behavior ........................................................... 88

References ............................................................................. 93

PART II. PAPERS .................................................................. 1
PAPER I. For Honor, for Toxicity: Detecting Toxic Behavior through Gameplay-- 1
Abstract ................................................................................ 1
### Table of Contents

1. INTRODUCTION ................................................................. 3
2. RELATED WORK ............................................................... 8
3. FOR HONOR ................................................................. 11
4. METHODOLOGY .............................................................. 14
5. DETECTING TOXICITY ...................................................... 23
6. DISCUSSION ............................................................... 25
7. CONCLUSIONS ............................................................. 32

REFERENCES ................................................................. 32

---

PAPER II. Aging Agents: Cross generational analysis of behavior and need satisfaction among players of Tom Clancy’s the Division 2

ABSTRACT ................................................................. 1

1. Introduction ................................................................. 3
2. Background ................................................................. 4
3. Methods ................................................................. 4
4. Results ................................................................. 8
5. Discussion ................................................................. 13
6. Conclusions ............................................................. 18

Acknowledgements ...................................................... 19

References ............................................................. 20

Appendix ................................................................. 23

---

PAPER III. Categorical Clustering Applied to the Discovery of Character Builds in TCTD2: The BaT Approach.

ABSTRACT ................................................................. 1

I. INTRODUCTION ............................................................. 3
II. BACKGROUND .............................................................. 4
III. CLUSTERING ............................................................. 6
IV. RESULTS AND DISCUSSION ........................................... 12
V. CONCLUSIONS ........................................................... 19
VI. LIMITATIONS AND FUTURE WORK ............................... 20

ACKNOWLEDGMENT ................................................... 21
<table>
<thead>
<tr>
<th>REFERENCES</th>
<th>22</th>
</tr>
</thead>
<tbody>
<tr>
<td>PAPER IV. Your gameplay says it all: modelling motivation in Tom Clancy's the division.</td>
<td>1</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>1</td>
</tr>
<tr>
<td>I. INTRODUCTION</td>
<td>3</td>
</tr>
<tr>
<td>II. BACKGROUND: MEASURING AND MODELLING MOTIVATION</td>
<td>4</td>
</tr>
<tr>
<td>III. PREFERENCE LEARNING FOR MODELLING MOTIVATION</td>
<td>9</td>
</tr>
<tr>
<td>IV. THE GAME AND THE DATA</td>
<td>10</td>
</tr>
<tr>
<td>V. DESCRIPTIVE STATISTICAL ANALYSIS</td>
<td>14</td>
</tr>
<tr>
<td>VI. MODELS OF PLAYER MOTIVATION</td>
<td>15</td>
</tr>
<tr>
<td>VII. DISCUSSION</td>
<td>23</td>
</tr>
<tr>
<td>VIII. CONCLUSION</td>
<td>24</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>25</td>
</tr>
<tr>
<td>PAPER V. Influencers in multiplayer online shooters: Evidence of social contagion in playtime and social play.</td>
<td>1</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>1</td>
</tr>
<tr>
<td>1 INTRODUCTION</td>
<td>3</td>
</tr>
<tr>
<td>2 BACKGROUND</td>
<td>4</td>
</tr>
<tr>
<td>3 METHOD</td>
<td>8</td>
</tr>
<tr>
<td>4 RESULTS</td>
<td>14</td>
</tr>
<tr>
<td>5 DISCUSSION</td>
<td>18</td>
</tr>
<tr>
<td>6 CONCLUSIONS</td>
<td>23</td>
</tr>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>23</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>23</td>
</tr>
<tr>
<td>PAPER VI. UPEQ: Ubisoft perceived experience questionnaire: a self-determination evaluation tool for video games.</td>
<td>1</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>1</td>
</tr>
<tr>
<td>1 Introduction</td>
<td>3</td>
</tr>
<tr>
<td>2 Existing work</td>
<td>4</td>
</tr>
<tr>
<td>3 Theoretical Framework (SDT)</td>
<td>7</td>
</tr>
<tr>
<td>Chapter</td>
<td>Title</td>
</tr>
<tr>
<td>--------</td>
<td>-------------------------------------------</td>
</tr>
<tr>
<td>4</td>
<td>Overview of current study</td>
</tr>
<tr>
<td>5</td>
<td>Development study</td>
</tr>
<tr>
<td>6</td>
<td>Validation Study</td>
</tr>
<tr>
<td>7</td>
<td>Discussion</td>
</tr>
<tr>
<td>8</td>
<td>Limitations</td>
</tr>
<tr>
<td>9</td>
<td>Future Lines of Research</td>
</tr>
<tr>
<td></td>
<td>ACKNOWLEDGEMENTS</td>
</tr>
<tr>
<td></td>
<td>REFERENCES</td>
</tr>
</tbody>
</table>
### Acronyms

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADL</td>
<td>Anti-Defamation League</td>
</tr>
<tr>
<td>AMI</td>
<td>Adjusted Mutual Information</td>
</tr>
<tr>
<td>ANN</td>
<td>Artificial Neural Networks</td>
</tr>
<tr>
<td>ARI</td>
<td>Adjusted Random Index</td>
</tr>
<tr>
<td>BaT</td>
<td>Builds as Text</td>
</tr>
<tr>
<td>DBSCAN</td>
<td>Density-based spatial clustering of applications with noise</td>
</tr>
<tr>
<td>DotA2</td>
<td>Defense of the Ancient 2</td>
</tr>
<tr>
<td>FH</td>
<td>For Honor</td>
</tr>
<tr>
<td>GEQ</td>
<td>Game Experience Questionnaire</td>
</tr>
<tr>
<td>GUR</td>
<td>Game User Research</td>
</tr>
<tr>
<td>HCI</td>
<td>Human Computer Interaction</td>
</tr>
<tr>
<td>k-NN</td>
<td>k (quantity) of Nearest Neighbor</td>
</tr>
<tr>
<td>LDA</td>
<td>Latent Dirichlet Allocation</td>
</tr>
<tr>
<td>LoL</td>
<td>League of Legends</td>
</tr>
<tr>
<td>MOBA</td>
<td>Multiplayer Online Battle Arena</td>
</tr>
<tr>
<td>OhE</td>
<td>One-hot-Encoded</td>
</tr>
<tr>
<td>PCA</td>
<td>Principal Component Analysis</td>
</tr>
<tr>
<td>PENS</td>
<td>Player Experience of Need Satisfaction</td>
</tr>
<tr>
<td>RF</td>
<td>Random Forests</td>
</tr>
<tr>
<td>RPG</td>
<td>Role-Playing Game</td>
</tr>
<tr>
<td>SDT</td>
<td>Self Determination Theory</td>
</tr>
<tr>
<td>SNA</td>
<td>Social Network Analysis</td>
</tr>
<tr>
<td>SVM</td>
<td>Support Vector Machine</td>
</tr>
<tr>
<td>TCTD</td>
<td>Tom Clancy’s The Division</td>
</tr>
<tr>
<td>TCTD2</td>
<td>Tom Clancy’s The Division 2</td>
</tr>
<tr>
<td>TD</td>
<td>Tom Clancy’s The Division</td>
</tr>
<tr>
<td>TD2</td>
<td>Tom Clancy’s The Division 2</td>
</tr>
<tr>
<td>tSNE</td>
<td>t-distributed Stochastic Neighbor Embedding</td>
</tr>
<tr>
<td>tSVD</td>
<td>truncated Singular Value Decomposition</td>
</tr>
<tr>
<td>UPEQ</td>
<td>Ubisoft Perceived Experience Questionnaire</td>
</tr>
<tr>
<td>WoW</td>
<td>World of Warcraft</td>
</tr>
</tbody>
</table>
Part I.

COMPREHENSIVE SUMMARY
I INTRODUCTION

Entertainment media and video games in particular have become an inseparable part of the entertainment regiment of consumer subjects. Growth of the video game industry emphasizes this fact and consequently necessitates not only improvements in technical quality of the products and by association, the quality of the experience, but also a deeper understanding of the gamer and her entertainment needs. Therefore, the discipline under study, Game User Research (GUR), is defined as the study of the consumers’ experience with the entertainment product, through assessment of different aspects of the perceived quality, considering the subjects’ positioning in the consumer society [1].

To grasp the makeup of a habitual gamer it seems essential to study historical, political and societal factors and their psychological effects combined with markings of developmental stages reflected on players’ behavior. Moreover, routine gaming implies a re-occurrence of behavior both within and outside the gaming environment. Although monitoring and recording behavior have long been a common practice in the game industry, most of the instances are limited to performance metrics, detection of anomalies, spending habits and churn prediction to maximize profitability, avoid undesired interactions and providing in-game feedback [2-5]. Immediate and tangible turnover of these methods guarantees their popularity and advancement but also discourages scientists to pursue a deeper analysis on the causes and effects of recurring behavior of consumer subjects.

On the other hand, evaluation of quality, or as we would refer to, identification and measurement of all factors influencing the perceived quality of a game, focuses on the game as the stimulus machine [6]. Hence, every aspect of the stimulus machine, from its technical performance [7] to its interactive [8] and social-ideological value [6] could be the subject matter of this quality evaluation. Principally, game developers use measurement of these aspects to ensure that the design goals are met throughout the existence-cycle (announcement, marketing, release, and live consumer services) of their products [9].

In this body of work, we assume that markers of being a consumer subject are reflected in players’ interactions with the game and other players. Consequently, we explore the relationship between player factors like demographics, preferences, sociality, and gameplay behavior categories such as playstyles, play
time patterns and social conduct. Furthermore, we articulate a methodical framework to model these interrelated concepts.

**Problem Statement**

The connection between the study of the game and scrutiny of its consumer subject, provides perspective into players’ demographic information such as age and social position but also motivation to choose Games as a medium in the first place, defining aspects of the game that would in turn impact player satisfaction and habitual engagement, as well as the nuances of modes of communication and social space that facilitates how the game shapes the player and vice versa.

Considering gamers as consumer subjects, which entails their historical placement, relationship to power, societies, and family structures they belong to, and the mark these factors leave on habitual and recurrent behaviors of gamers have rarely been addressed [9]. Implications of the considerable contribution of video games in the day-to-day leisure activities of the consumer have been largely reduced to dependence [10-12] and relationship to violence [13-15]. Yet, there has been additional contributions such as study of the cognitive effects [16] and games as learning platforms [17]; but also, sparse examples in epistemology, ethics [18] and politics [19]. The common theme throughout this study is the
attributes of players who are critical for a game’s success [20-22] by showing self-motivation to return to the game and keep playing it, as players want to decide how characters grow, which activities to undertake and which areas to explore [23]. The amount and the variety of activities afforded by open world and sandbox games has brought forth the need to invest into understanding players and their motivations. Partially because it is not sufficient just to know if players are engaged or having fun, but it is necessary to infer which activities they prefer and why. This need for more granular evaluation of different aspects of these game worlds entails the ability to connect activities offered by games with players’ needs, motivations and desires. The central role of motivation for the design of games, and the experiences they elicit, has been highlighted by a growing number of studies which adopt psychological theories of motivation within games [24-28]. Such studies, however, follow a top-down integration of phenomenological models of motivation, which aim to identify and explain stereotypical player behavior. We review prominent top-down approaches of surveying player experience in Chapter II, Section1. Our first step is to introduce a questionnaire to survey player identification as a gamer and asking them about motivations behind engagement with this particular medium. This approach will help game developers evaluate their products during and after development and compare design intentions with player satisfaction. Ultimately, developers will have a framework to adjust gameplay features to specific needs of their defined demographic. To that end, it is necessary to tailor game titles based on players’ perceived experience.

Over the last decade, games user research and industry-based game testing has shifted its focus towards quantitative approaches based on player analytics [29], with the aim to shed more light onto the understanding of player behavior and experience. These approaches mainly focus on either clustering players based on their behavioral patterns or predicting objectively defined aspects of their gameplay behavior for monetization purposes (e.g., churn prediction) [21]. Classification of player data, other than player churn and retention behavior [30], could be used to increase player efficiency in gameplay based on the choices afforded by the game [31]. Such approaches could help game developers to improve gameplay features and help them personalize representation of options [86]. It may also diversify the understanding of players and their varied needs [32]. We contribute to this tradition by introducing categorical clustering of players based on their character builds and analyzing social aspects of gameplay through elements of in-game social media. As character builds are highly reflective of
playstyle, their categorization could be a proxy to categorization of player behavior. Therefore, performance of the model in view of behavioral metrics such as health, armor, skill power, offense, defense, utility, and overall playtime split is examined to show that the different discovered builds are indeed representative of specific character attributes and playtime behavior.

Although the majority of approaches that aim to capture aspects of player experience (such as engagement or motivation) based on player analytics remain qualitative, given the complexity of measuring subjective notions of user experience in games [29], we introduce a data-driven player modelling approach [33] by assuming there is an unknown underlying function between what a player does in the game behaviorally—as manifested through her gameplay data—and her motivation. Our hybrid modeling approach is motivated by the lack of quantitative studies on the relationship between motivation and play, including subjective aspects such as age and generational cohort but also behavioral indices of disruptive social behavior, playstyle and performance.

On the other hand, with the current paradigm shift of the game industry towards games as a service [75-77], player retention has become one of the most important design goals and metrics. Entertainment value is commonly measured in hours of playtime. To accommodate for these shifting values, commoditization strategies have begun revolving around subscription-based models, free to play games with premium content, and season passes. They often have a focus on online (Live) competitive multiplayer experiences such as Tom Clancy's The Division (TCTD, TD [34]) and the consecutive iteration Tom Clancy's the Division 2 (TCTD2, TD2 [35]) and For Honor (FH [36]), which are the bulk of the subject matter of our studies, but they can also include a variety of other types of game experiences. ‘Live’ refers to all the activities and interactions created for the game community including pre- and post-launch as well as regular updates, new content, and events both in-game and out-of-game, throughout the game’s lifespan [37].

As shown by the longevity of games such as World of Warcraft (WoW) [38] or League of Legends (LoL) [39], social connections foster prolonged retention. One of the most important tools that the industry uses to investigate social connections, especially in social and online games, is Social Network Analysis (SNA). Increasingly, social network analysis methods are being used in games [40-44]. Similar to the literature on online communities [45], it suggests that there are key members who contribute to keeping the community alive. In our analysis of
social behavior in games, we identify and examine socially constructive player connections but also review approaches to detect, mitigate and reprehend socially disruptive behavior as well.

Disruptive behavior has been identified as a persistent issue in online games, especially online competitive games such as *Overwatch* [46], *League of Legends*, Defense of the Ancients 2 (DotA 2) [47] and other multiplayer online battle arenas (MOBAs). According to a recent study by the Anti-Defamation League (ADL), examples of toxicity range from emotional abuse to blaming others for losses, offensive words, derogatory appellatives, unsportsmanlike behaviors, and selfish conduct [48]. The same study reveals that of the surveyed 733 US gamers aged 18–45 who played online multiplayer games, 81% reported some form of harassment related to their race/ethnicity, religion, ability, gender, or sexual orientation in the previous six months, with a stunning 68% of online multiplayer gamers experiencing more severe abuse, including physical threats, stalking, and sustained harassment. This toxicity has an impact on the game experience as well as the well-being of players: the study reports that 64% of gamers feel harassment is shaping their gaming experiences, and such that players perform less (see [49]), avoid and stop playing certain games, become less social and feel isolated, and, most concerning, have depressive or suicidal thoughts. It is thus not a surprise that toxicity also has an impact on the retention and Lifetime Value of games [50].

Game publishers, platform owners, online voice-chat applications, and even the police and national intelligence and security services are aware of these issues and are working to confront them, but entangled with freedom of speech issues, technical difficulties, and a lack of chargeable offenses on the legal side make toxic elements a challenge to extinguish [90]. Additionally, players themselves do not tend to report offenses: fewer than half of respondents of the ADL study said they reported toxicity using in-game tools. This happens for several reasons, including the effort required in the reporting process, reports not being effective or taken seriously, or toxicity being a normalized part of the play experience [48]. That is why it is important to develop automatic toxicity detection strategies to help community managers.

**Research Questions**

Understanding gamers as explained in this section helps game developers to perform their creativity alongside the knowledge of their audiences’ psychological demands and expectation of quality, which includes reinforcement of positive social growth and mitigation of toxic and disruptive behavior on the
other end. Successful attempts at focusing on enhancing players’ well-being by addressing their psychosocial needs, therefore, should also bring about increased retention and revenue as a nice-to-have side effect. In summary, there is a need for continuous and longitudinal monitoring of players’ behavior, not limited to actions and performance in game, but expanded to fundamental characteristics of the threats to and need of the player as a consumer subject. More specifically, the following research questions need to be answered:

RQ1) What are the typical approaches of studying one’s placement in consumer society and its relationship to media (video game) consumption?

While few researchers have delved into the epistemological and political implications of video games as a media [6, 18], even fewer have investigated psychological prerequisites of play [2] rather than its relationship with violence and dependence [12,14]. Therefore, we are focusing on attributes of a player as a consumer subject that show themselves in patterns of behavior in-game, and through the study of player demographics and psychological needs.

We argue that the agency exercised by the player, in variety of playstyles, in quality and quantity of social interactions and even in the type and sequence of activities they undertake is a marker of player behavior and interconnections of these markers with player perception and psychology are worth exploring.

RQ2) How is the study of -predictive- behavioral models achieved?

Classification and prediction of behavior in video games typically is limited to conventional statistical methods and profit-oriented game performance indicators such as churn [30]. However, alternative methods of parsing through subjective player data, classifying categories of in-game behavior based on their context, unique social interaction analysis techniques, and detection of toxic behavior only through gameplay, are areas that we will examine and scout in the following sections.

RQ3) What would be the interrelation and appropriate classification of historical, political, and developmental indices of a gamer and their respective behavior? (psychosocial factors as vectors of behavior)

As most studies of player characteristics is done qualitatively, on small samples of students and classical validation techniques are rampant [32], the flip side of the coin is purely numerical approaches to model behavior and factors of it [9]. We aim to present hybrid approaches of modeling subjective and behavioral data together with comparing their performance with conventional methods as well as
extensive investigation into the nature of the input data and what the most efficient techniques are for analyzing that specific type of data.

**RQ4) What additional information could a longitudinal modelling of player-behavior based on their psychosocial profile provide us?**

The need for dynamic models of player behavior has been emphasized by researchers and industry professionals, specifically when it comes to analyzing time series data [30]. Although examples of application of long-term data are rare and far between, we approach our investigations with a sensibility towards long-term implications of our findings and how they could mature.

In Summary and as outlined, we start by proposing a survey of player psychological need satisfaction, to measure and ultimately model player motivation. In the next step, we analyze motivational and behavioral factors connected to age and generational cohort. Then, we introduce a framework for predicting player motivation based on gameplay data, which includes reconsidering the approach to process subjective player data as well as dynamic classification of player behavioral data. We continue with other novel approaches to player classification through character builds and in-game social media to further study player interactions with the game and other players, in our final step through discussion around and detection of disruptive social behavior in an online game.

**Pronouns, Style, and Clarification**

In this work, the pronoun “we” is often used to refer to researchers and authors of the presented work.

‘Players’, ‘Gamers’ and occasionally ‘participants’ is used to describe consumers of digital video games who are the subject matter of the majority of the studies in this body of work.
II BACKGROUND

This chapter provides an overview of the existing literature surrounding the concepts discussed in this thesis. Namely, a review of prominent game evaluation and motivation assessment tools is offered. Then, game motivation and profiling and their connection to players’ age and intergenerational cohort will be reviewed. Section 3.4 connects player behavioral modeling approaches and appropriate methods of analyzing player subjective data. Finally, social aspects of play, such as social contagion, disruptive and toxic behavior as well as their sign and effects will be explored.

1. Measuring Experience

Beside video games’ outstanding market outreach, they provide unique conditions for attention modulation and involvement of the player. This increased involvement and attention leads to reduced awareness of the self and the surroundings. It is an essential task to examine both the physiological and psychological processes that subjects undergo during an instance of such inhabitation in video games. At the same time, it is also imperative to investigate what are the characteristics of a game that provides conditions for experience of absorption also known as immersion, state of flow, and engagement; for an analysis of definitions and varying degrees of this phenomenon see [51]). Much of existing game evaluation tools are concerned only with quality of subjects’ experience (marked in Table.1 as “subject oriented). However, as Calleja suggests, a more comprehensive approach would entail simultaneous study of player assimilation into the game world as well as systemic acknowledgement and representation of player by the game [51].

Numerous other behavioral, physiological, and subjective experience evaluation tools have been developed and used to measure different aspects of player experience (see [52] for a review). In this section, theoretical background, and formulation of three prominent models, namely Game Engagement Questionnaire (GEQ) [53], BrainHex [54] and Player Experience of Need Satisfaction (PENS) [55] is being reviewed. (See Table.1 for a summary).
<table>
<thead>
<tr>
<th>Table 1: Comparing Measures of Game Experience</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Variables</strong></td>
</tr>
<tr>
<td><strong>STEPP</strong></td>
</tr>
<tr>
<td><strong>Presence + Flow</strong></td>
</tr>
<tr>
<td><strong>Perception of Improvement</strong></td>
</tr>
<tr>
<td><strong>Evaluations</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Methodology</strong></th>
<th><strong>Lower Bound</strong></th>
<th><strong>Upper Bound</strong></th>
<th><strong>Number of Surveys</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Partial Least Squares</strong></td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td><strong>Structural Equation Modeling</strong></td>
<td>21</td>
<td>21</td>
<td>21</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Theoretical Framework</strong></th>
<th><strong>Lower Bound</strong></th>
<th><strong>Upper Bound</strong></th>
<th><strong>Number of Surveys</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Theory of Action</strong></td>
<td>21</td>
<td>21</td>
<td>21</td>
</tr>
<tr>
<td><strong>Flow Theory</strong></td>
<td>21</td>
<td>21</td>
<td>21</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Results</strong></th>
<th><strong>Lower Bound</strong></th>
<th><strong>Upper Bound</strong></th>
<th><strong>Number of Surveys</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Results</strong></td>
<td>21</td>
<td>21</td>
<td>21</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Controls</strong></th>
<th><strong>Lower Bound</strong></th>
<th><strong>Upper Bound</strong></th>
<th><strong>Number of Surveys</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Frequency</strong></td>
<td>21</td>
<td>21</td>
<td>21</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Conclusion</strong></th>
<th><strong>Lower Bound</strong></th>
<th><strong>Upper Bound</strong></th>
<th><strong>Number of Surveys</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Conclusion</strong></td>
<td>21</td>
<td>21</td>
<td>21</td>
</tr>
</tbody>
</table>
As a theoretically established alternative to subject-oriented measurement tools, Self-determination theory (SDT) argues that players are intrinsically motivated to engage in an activity that satisfies their inherent tendency for psychological growth and well-being. SDT assumes that these basic psychological needs and their pursuit is universal and independent of culture and developmental stage [56]. Need satisfaction, therefore, refers to the perceived degree of satisfaction of the basic needs for autonomy competence and relatedness. SDT has been applied to a variety of academic and industrial fields including video games [55]. Deci and Ryan have shown how the self-perception of satisfaction of the needs identified by SDT, is a correlate of performance, self-esteem, and general well-being [57]. In context of video games, Ryan showed that enjoyment and future gameplay are correlates of satisfaction of autonomy, and competence in development of PENS survey [55]. These findings suggests that subjective perception of need satisfaction could be used as a game-oriented alternative to subject-oriented measures. In that, instead of surveying players about the fleeting idea of fun or retrospective investigation of player’s situational awareness, satisfaction of basic psychological needs is being evaluated. This evaluation is applicable not only to the game as a whole, but also each activity within the game as its own agent of need satisfaction.

Other game-oriented measures (BrainHex [54] and Yee [58]) provide a taxonomy of gameplay features that a player have experienced in other games and is likely to perform in the future. Instead of a list of preferred features, evaluation of need satisfaction also leaves room for a more creative approach for improving already existing game elements. Additionally, a prevalent limitation of existing models often includes a small and specific sampling of validation population in the developmental stages of the model (often the sample is limited to few classes of graduate students). Small and specific samples are a significant threat to the external validity of these models; for this reason, game developers could be skeptical in adopting existing framework and models to their practice.

Finally, since the identified needs of Autonomy, Competence and Relatedness are broadly defined (as discussed in the following section), there is a certain interconnection between the three rather abstract constructs. This is perhaps the reason why the evaluation of need satisfaction in video games is often accompanied by other metrics such as intuitive controls, presence, immersion and measures of extrinsic motivation [55].
2. Psychological Need Satisfaction

Introduced as a metatheory, self-determination theory primarily assumes there are individual tendencies toward a unified sense of self (consciousness), and that these tendencies are by nature constructive. It also argues that there are social-contextual factors that support or thwart this innate tendency [56]. Therefore, the theory introduces three universal and basic psychological needs that are essential for psychological well-being. Several measures of positive affect and mental health have been used as evidence for eudemonic well-being [59].

According to self-determination theory, the need for **Autonomy** refers to the experience of volition, through providing choices or anticipation of opportunities, whether illusory or real, that are perceived fair and equally potent. In an Open-world simulation, the sense of autonomy could be described as control over which type of activities to engage with and the agency to perform them in one’s preferred playstyle as well as game’s differentiated and emancipating support for these choices. Subfactors of Autonomy in games could be the sense of freedom and having options (**Playstyle**) and the impact that player choices have on the course of gameplay (**Agency**). The second need, **Competence**, emphasizes the need to feel capable and effective while constantly improving oneself through challenges. In a video game, competence can be induced by progression through repetition. In other words, activities that requires skillful performance of combinatory logic of the game and is usually reinforced by empowering feedback cues (e.g., an explosion as audio/visual feedback to shooting) are competence inducing. Subfactors of Competence in games are Growth and Mastery. Finally, **Relatedness** is the concept of social belonging (**Closeness**) and being hailed as a social construct (**Interdependence**). In context of open-world games, positive relatedness includes any interaction with other players or Non-player characters that promotes autonomy and/or competence of the player [55].

These psychological needs are defined as parallel to physical needs, hence psychological needs are distinguished from motivations and desires. While desires and motivations fueled by these needs not always complement psychological well-being and growth, SDT argues that the three basic needs naturally presuppose a positive outcome [56]. At the same time, being self-determined means that these needs function at an individual level and are closely tied to one’s perception of their satisfaction. Effects of subject’s sensitivity to each of the differentiated needs as well as a factor of importance or priority for them is another theoretical challenge of SDT [57].
On the other hand, as mentioned by Deci and Ryan [55], one activity can cover more than one need, but it could also satisfy one need and thwart another. For example, in context of video games, acquiring a superior tool or weapon is expected to satisfy the need for competence but it could, by making other options practically obsolete, thwart the need for autonomy. The definitions of the three basic psychological needs, as already explained, is rather broad. This inclusivity presents both an opportunity and a challenge: it is in fact possible to address the definitions creatively, but it also leads to pointless discussions on the confines of the definitions. Therefore, the semantic overlap of the constructs of SDT is both logical and lexical, and it refers to:

(1) the subjectivity of the perception of priority and importance of needs and
(2) interconnection of meaning between needs.

2.1 Generational Effect

Previous studies of intergenerational gaming have largely focused on the intergenerational interaction that games enable and its beneficial impacts, such as increased understanding of other generations and reduced social anxiousness [98]. Reviewing the factors and benefits of designing intergenerational games, De la Hera and colleagues argue that it is necessary to take a closer look into the effects of age and gender on benefits of playing digital games.

Furthermore, while some research has focused specifically on a single generation in order to identify factors such as gaming habits and reasons to play (e.g. [60-62]), the current study will expand on this topic by investigating similarities and differences between generations and their perception of need satisfaction and building linear and non-linear models to understand importance factor of game play and motivation of different generations.
3. Profiling and Modelling Behavior

Player modelling is a field of games user research specialized in understanding and simulating [63] videogame play through AI techniques [21]. The two main areas within this field focus on behavioral profile aggregation and predictive modelling.

3.1 Behavioral Profiling

Player profile aggregation usually relies on clustering algorithms and traditional user analytics [29] to expose and visualize underlying patterns in player behavior and experience. This line of research applies different forms of aggregation to create an abstract, higher-level representation of the game. Common computational approaches use $k$-means clustering, self-organizing maps [64], matrix factorization [65], archetypal analysis [66-67], and sequence mining [68-70]. The primary aim of these studies is to acquire a static profile of general behavioral and psychological patterns which can categorize the player population. In other words, the coin of measuring player experience has a flip side of player actions and choices inside the game, which by our assumption are proxies of player behavior.

As video games provide the unique conditions of interactivity as well as affective simulation, the dynamic nature of player behavioral data could be utilized to gain insight into player preferences [21], understand the underlying motivations for play [32], quality of the experience [71], and to make adaptive games [72]. Other than business applications of player behavioral modeling conventionally focused on player churn behavior and spending habits [72], and experimental modeling for method development [73], design oriented clustering studies on large scale samples of players of popular commercial video games are mostly focused on numerical clustering methods such as $k$-means [74] and are rarely conducted by industry practitioners [75] with context dependent, high dimensional dataset such as the various datasets introduced in this body of work [67].

Centroid-based clustering methods (e.g. $k$-means) have been successfully used to classify player behavior [76] due to their ease of use for large scale numerical values and low time complexity [77]. However, previous research [66] showed that they may not produce reliable results. Density based clustering methods, although employed less frequently, have also proven useful in modelling player
behavior with reduced dimensionality [78]. But as we will discuss in Chapter. IV Section 6.2, their dependence on the availability of pairwise distances between data points makes them impractical for handling large amounts of high dimensional data. Therefore, treatment of special data types has led practitioners and researchers to employ alternative methods. On one occasion [70] a sequence clustering method was used to identify and categorize game loops based on the frequency of repeated consecutive action types in a video game. In another example, [30] treatment of contextual time-series data was conducted with Dynamic Time Warping and a series of other refined algorithms for evaluation of game events. Other studies [79-80], also used methods similar to those of the current study when dealing with text-based data, although applied to player in-game chat logs or tweets about a game, in order to categorize and uncover player communications.

Regardless of data type, the common thread in clustering studies is the emphasis on the importance of context not only in mining and interpretation of data [66] but also in the methods that make sense of relationships between these data points.

3.2 Predictive Modelling

In contrast to aggregate player profiling, predictive affective and psychological modelling and profiling, takes a more dynamic approach and aims to actively predict certain behaviors or emotional shifts of the player experience [21]. This type of modelling often uses supervised learning and relies on gameplay data instead of aggregation and abstraction. Notable applications of player modelling include predicting player behavior [81] such as churn [82-83], playtime [2], or player experience [84-85]. While the several of these studies rely solely on gameplay data [9, 86], some of the studies focus on multimodal player data that fuse gameplay with physiological data [87-89] or data from the video streams of players [90]. While these multimodal signals are often shown to increase modelling accuracy, collecting physiological data is currently too expensive and intrusive to be feasibly applied in large-scale industry studies.

Motivation is a crucial element of player experience and research, be it player profiling or modelling. Understanding what drives players can help game designers and advise the development of adaptive and generative systems [92]. While some studies have incorporated motivational survey data into their user models to predict other gameplay outcomes, such as churn or enjoyment [93-95], relatively few focused on motivation as the target output of their models. As a rare example Canossa et al. applied regression to model player drives [24]
based on Reiss’ Motivational Profile [96]. In our inquiry, we used preference learning to model factors of Self-Determination in players of Ubisoft’s Tom Clancy’s The Division [9]. This section also extends that work significantly by using the same dataset but examining 8 additional motivation factors based on Nick Yee’s Model of Player Motivation [58]. Moreover, the two frameworks are compared in terms of their robustness and generalizability as well as their capacity to be predicted solely from gameplay characteristics.

3.3 From Theory to Measures of Motivation

We focus on two different popular surveys that capture aspects of player motivation. While both rely on survey data, they use fundamentally different approaches to process this data and derive relevant dimensions of the player’s motivation. The Ubisoft Perceived Experience Questionnaire (UPEQ) [32] uses a theory-driven framework to acquire aggregated scores of four factors of self-determination following the theory of Deci et al. [97], Ryan et al. [55], and Chen et al. [98]. The Model of Player Motivation, instead, is a data-driven method which relies on factor analysis that derives the underlying patterns of player motivation following the work of Yee et al. [58]. In this section we review the inner workings of both questionnaires used.

As discussed in Section 2, self-determination theory (SDT) is a well-established positive psychology theory of the facilitation of motivation based on the work of Deci and Ryan [99], which has been adopted to a wide variety of domains, including education [97], job satisfaction [100], parenting [101], health and exercise, [102], and videogames [20, 55, 103-104]. The core theory was developed to contrast earlier frameworks of motivation as a unitary concept [105-106], by focusing on the dichotomy of the intrinsic and extrinsic locus of causality behind motivation [107]. The latter is facilitated by external or internal rewards, pressures, and expectations, while the former is based on the intrinsic properties of the activity itself, namely how well it can support the three basic psychological needs of competence, autonomy, and relatedness. Videogames include a fair number of pressures and rewards which can promote extrinsic motivation [108], and yet they are generally regarded as good facilitators of intrinsic motivation [103]. Even when short-term shifts in motivation are observed during gameplay, games support the necessary psychological needs for the facilitation of intrinsic motivation on a higher level [104]. In the context of videogames, Ryan et al. [55] describe the basic psychological needs underlying intrinsic motivation as:

1) **Competence** or a sense of accomplishment and a desire for the mastery of an action, which manifests through the proximal and distal goals of the players.
This need is generally tied to self-efficacy and a sense of meaningful progression. It is supported through the interactions the players ought to master in order to complete the game but not completion in itself.

2) **Autonomy** or a sense of control and a desire for self-determined action, which manifests through meaningful choices, tactics, and strategic decisions the players can take. It is supported through rule systems and different game mechanics that both structure the play experience but allow for a high degree of freedom and meaningfully different outcomes.

3) **Relatedness** or a sense of belonging and a desire to connect and interact with others, which manifests through interactions with other players and believable computer agents. It is supported by multilayer interactions, believable and rich non-player characters, narrative design, and even interactions with other players outside the games as well.

4) **Presence** or the feeling of a mediated experience is a main facilitator of both competence and autonomy, and can be viewed as having physical, emotional, and narrative components [20,55]. Indeed, the feeling of presence or the pursuit of immersion can be a driving force behind the motivation of gameplay [104, 109-110]. Based on the strong relationship between STD and presence, both the Player Experience of Need Satisfaction Questionnaire [55] and UPEQ [32] use it to measure a level of involvement with the game which can facilitate other positive psychological needs.

It is important to note that the above factors are not contributing equally to the formulation of intrinsic motivation; while competence or relatedness are regarded as the core catalysts, autonomy generally plays a supporting role in the facilitation of motivation. Nevertheless, in absence of autonomy, motivation can only be considered introjected or compulsive [111]. Within games the main drive of intrinsic motivation is generally competence because of how the activity is structured, while relatedness contributes to enhancing the experience [20]. In this paper we rely on SDT and recent advances on measurement tools to quantify the four above-mentioned aspects of motivation. For that purpose, we use UPEQ, a game-tailored questionnaire designed to measure the factors of SDT as affected by the gameplay experience. UPEQ was developed [32] specifically to predict gameplay outcomes relevant for industry designers and stakeholders. Earlier work [32] has demonstrated that UPEQ is able to predict playtime, money spent on the game, and group playtime based on measured factors of SDT. Beyond its utility, UPEQ also addresses the limitations of prior domain-specific SDT questionnaires,
such as the Game Engagement Questionnaire [53], BrainHex [50], and the Player Experience of Need Satisfaction [55], while focusing on the adaptation of the Basic Need Satisfaction Scale(s) [98] into a survey specific to videogame play. The result is a reliable and consistent assessment tool with a strong theoretical foundation in SDT.

1) **Ubisoft Perceived Experience Questionnaire:**

UPEQ relies on Self Determination Theory (SDT) [97], which is a popular theory of motivation describing the phenomenon through the perceived locus of causality of the motivation and the satisfaction of basic psychological needs. The locus of causality of the motivation is either extrinsic or intrinsic [99]. The former type of motivation is facilitated through external drives or rewards, while the latter relies on the aforementioned psychological needs. These psychological needs are competence, a need for mastery and completion, autonomy, a need for meaningful choices and self-determined action, and relatedness, a need for connection with other humans— and in the case of videogames, expressive non-player agents [20]. UPEQ extends these factors with presence, which is the illusion of an unmediated experience that occurs when players stop perceiving the existence of the game’s medium [104,109]. UPEQ uses 5-point Likert-scales to measure these dimensions through a 24-item survey (with 7 items for each of competence, autonomy, and relatedness, and 3 items for presence). The survey was developed based on the Basic Need Satisfaction Scale(s) [98], addressing limitations of contemporary surveys for measuring SDT in games [53-55]. UPEQ has been shown to be a reliable measure and has been used in studies.

(a) as input to predict solo playtime, group playtime, and money spent [32] and
(b) as output to model motivation based on gameplay [9].

The main strength of this approach is its reliance on a well-established theory which led to many critical observations in multiple domains, proving the validity of the framework. The limitation of the method, however, also stems from its heavy reliance on a top-down framework. Although theory-driven approaches excel at abstracting observations for explanatory analysis, predictive modelling of these constructs may often be challenged due to the deviation of such constructs from the ground truth.

2) **A Game-Specific Model of Player Motivation:** The second approach for capturing motivation examined in this study relies on the work of Yee [58, 93, 99] and derives game-specific categories of player motivation from survey data. Yee et al. use a 66 item, 5-point Likert scale survey, which collects data on player preferences. These 66 questions can be organized into 6 main dimensions
(action, social, mastery, achievement, immersion, and creativity) [58]. To build this model, Yee et al. used factor analysis on survey data collected originally from 250,000 players. This model has grown over the years and the current version used by Quantic Foundry uses data from more than 400,000 gamers. Although the model used by Quantic Foundry is well-tested and robust, higher reliability can be achieved on smaller datasets by tuning the model to the specific game and player population at hand through reapplying the same analytics method and deriving data-specific factors. The presented study applies this methodology to find the items and dimensions of the Player Motivation survey of Yee et al. [58] that are most relevant to our gameplay data. Through a process of elimination and aggregation (see Chapter V Section 3), the tuned model arrives at the eight motivational factors. The main strength of this method is the game-specific nature of the motivational factors, which are tailored to the given player population. While this makes PM models very efficient in the context of the given dataset, it can be expected that the motivational factors identified do not generalize well to unseen games or populations due to their reliance on the topology of the current data.

3.4 Ordinal Player Modelling

Given the above theoretical framework on the ordinal nature of experience and the large body of recent empirical evidence on the benefits of the ordinal modelling approach [84], in this paper we view player motivation as an emotional construct [20] with ordinal properties. As a result, we compare player feedback on relative grounds and use PL to model the ranking between the levels of reported motivation in players as measured by the factors of UPEQ. We consider the UPEQ scores as the underlying ground truth we need to approximate. After acquiring a general score for all the measured factors for each participant, we return to analyzing and modelling the data as ordinal values, thereby following a second-order modelling approach [84].

We rely on ordinal data processing and modelling, in accordance with contemporary research that highlights the ordinal nature of human emotions and cognitive processes [84]. While traditional models of player experience rely on absolute and unified scales to overcome individual differences, they also skew the underlying—inherently ordinal—ground truth which is subject to adaptation-bias [112] and anchoring-bias [113-114] and diminishing returns from habituation [115] and recency effects [116]. Based on the numerous limitations of handling absolute values of subjective phenomena this study

1 www.quanticfoundry.com
instead focuses on the relative relationships of the data and thus uses ordinal data processing and preference learning [84]. A growing body of research is dedicated to the ordinal processing and modelling of emotions—not merely in game research but in affective computing as a whole [84, 117-118]. These studies show that beyond a first-order ordinal representation (where datapoints are already captured in a relative fashion), a second-order processing (where datapoints captured as absolute values are translated into an ordinal representation) improves the reliability and validity of the derived models [84]. Given the benefits of ordinal modelling in this paper we follow the second-order approach to convert the absolute Likert scores of the UPEQ and PM dimensions into ordinal values prior to modelling them via preference learning.

4. Social Influence

In the following sections, we will describe the work that is done on SNA and social contagion in the context of games. Then, we will turn to what we know about influencers in general before we discuss the other side of social influence, namely when it disrupts other players experience.

4.1 Influence in Social Networks

Social Network Analysis (SNA) is a family of methods for formally describing and analyzing relations between people as graphs with nodes (people) and edges (relations), with broad applications in offline and online social networks [119]. A major topic of SNA research is social influence, as expressed for instance in behavioral and social contagion theory [120-122]: behaviors (like physical activity or prosocial behavior) and their consequences (like obesity or happiness) cluster and spread within networks [e.g., 120, 123]. Methodologically, social influence is often hard to disentangle from homophily, namely where similarity is the primary cause for connections [120, 124]. Still, there is now good evidence for contagion processes in social networks via social-psychological mechanisms such as modeling or norm-setting [125-126]. Put differently, Not only do similar behaviors attract connections, being connected causes more similar behavior. As online social networks have become major means of communication, social influence has become subject to intense interest in communication and marketing as well as computer science and human-computer interaction (HCI) communities, especially computer-supported collaborative work and learning, Internet research, or informatics [127-129]. Practitioners have been chiefly concerned with finding ways to maximize the spread of desired information and
behaviors through networks, and to reliably measure the impact of specific actions and actors [130].

4.2 Social Network Analysis in Games

The volume of work in SNA is substantial, and due to limitations of space we here focus on previous work in games, directly related to the current project. Previous research on social behavior in games suggests that social interaction influence the user experience and forms an important motivational driver for play [131-132], giving the games industry a direct interest in social network analysis [133]. SNA has been employed as a method for investigating social interaction between players primarily since the introduction of social network games in the mid-late 2000s [134]. Social networks in games have been investigated using qualitative methods and ethnographic approaches [151], as well as using quantitative SNA [132, 135]. The available SNA work is mainly focused on massively multiplayer online games (MMOGs), using in-game social features such as friend lists to construct networks. Ducheneaut et al. [131] and Shen [136] examined social interactions in these types of games. Surveys have also been used as a method for collecting information about the social connections of players, e.g., Shen and Chen [137]. Szell and Thurner [138], studied the structure of friend-, enemy-, and communication networks, noting that friend and enemy networks were different topologically.

Player-generated structures such as guilds have also been investigated, e.g., by Ducheneaut et al. [139] and Chen et al. [140] who used SNA metrics such as density and centrality to map and characterize the properties of player guilds in World of Warcraft. More limited attention is given to other game genres. One exception is Iosup et al. [141] that looked at social networks in DOTA 2 using matchmaking as the baseline for building edges between players. Rattinger et al. [142] used similar connections between players in Destiny to build networks. The authors noted that the most engaged players were characterized by having large social networks. Following up, Schiller et al. [143] analyzed a social matchmaking service for Destiny players operating outside the game itself. Summarizing, SNA as applied to games has been focused on the associations that form between players during and around the playing activity [131, 135, 142]. There is more limited work on social structures formed around games [143], not only for external services, but also distribution platforms such as Steam and Uplay. The work presented here forms a concrete extension of previous work applying SNA in games contexts, not only by integrating information about social connections from the Ubisoft distribution platform Uplay, but also in its continuation of the
work by e.g., Rattinger et al. [142] on using SNA to identify players with specific properties across in-game behavior and network behavior.

### 4.3 Social Contagion in Games

With respect to social contagion, there has been some evidence in online games, such as generosity (gifting in-game money) [144-145], purchasing of in-game goods [146], and cheating such as bot usage [147], including initial exploratory attempts at identifying “spreaders” or influencers with an out-sized impact on cheating behavior [148]. However, research suggests that online in-game interaction network structures and dynamics are context-sensitive, meaning different kinds of interactions and relations (friending, trading, messaging, etc.) show different structures and dynamics [149]. Thus, the existence of social contagion for gift-giving does not immediately generalize to e.g., team play, as different kinds of interactions have different strategic and other utilities and thus bring in different considerations and social-psychological mechanisms [149].

### 4.4 Influencers

There is no agreement on what an influential person is [150]. However, two types of influencers can be distinguished in previous work:

1. an individual who impacts the spread of information or behavior, people who influence people [151]; and
2. an individual who exhibits some combinations of desirable attributes such as trustworthiness and expertise or network attributes (connectivity or centrality) [152].

The first group of influencers are often referred to as opinion leaders [153], prestigious innovators [154], key-players [155] and spreaders [156]. The second group of influencers are often referred to as celebrities [157], evangelists [158] or experts [152]. Here we focus on measuring and quantifying the influence of an influencer of the first type, for two reasons. First, because they may touch a large scale of audience with a very small marketing cost [159-161]. Second, because their tendency to spread desirable behavior may be key to keep healthy communities alive for a longer time [154, 156].

Centrality measures have been proven to be relevant indicators in the analysis and comprehension of influencers in a social network [162-163]. The most utilized measures of centrality are in- and out-degree, betweenness, eigenvector and closeness; they are all measures of an actor's prominence in a network [164].
Valente et al. [165] investigated correlations between these most common measures of centrality. The researchers found that there are strong but varied correlations among the centrality measures presented here. The average of the correlations was 0.53 with a standard deviation of 0.14, indicating these measures are distinct, yet conceptually related. Since the centrality measures examined are not mutually excluding members but have a slightly different selection criteria, in order to identify the players with most influence we will utilize all the centrality measures and select only players that are ranked at the top for each measure.

5. Disruptive Behavior

In this section, we outline the challenges in anonymous social space of video games that disrupt other players’ experience and the work that has been done in defining such behavior, what kind of online games provide the criteria for emergence of such behavior and reprehensive precautions taken by the industry to address this phenomenon.

5.1 Definition of Toxicity

There is no standard definition of toxic behavior; however, in the context of games it is generally defined as behavior that intentionally disturbs other player’s experience and well-being. The definition of toxic behavior in each game can vary, but, among others, it includes cyber-bullying, “flaming”, acting nosy, cheating, and illegal behaviors [166-168]. The Anti-Defamation League (ADL) defines it as “disruptive behavior” such as trolling/griefing, personally embarrassing another online player, calling offensive names, threatening with physical violence, harassing for a sustained period of time, stalking, sexually harassing, discriminating against by a stranger, or doxing [48].

In an effort to more systematically understand and capture toxic behavior in games, Kowert [167] proposes a categorization of toxicity based on performance type (verbal or behavioral) and impact type (transient or strategic). While the proposed categorization might not always be pertinent, the paper points to a crucial aspect of toxic behaviors in general, namely the fact that toxic behaviors are often culturally defined, both in a specific game’s culture and according to the countries players are from: behaviors that are considered toxic in one situation might not be considered toxic in another. It is precisely the culturally relativistic nature of toxic behaviors that renders automatized efforts of detecting toxicity so difficult. Because of this, in our study we rely on human-curated set of labels provided by Ubisoft community managers for the game For Honor (FH).
5.2 Victims, Perpetrators and Environments
Toxicity in games has been a considerable problem for years. Therefore, the academic community has researched this topic thoroughly, from very different domains ranging from social studies to computer science. Although the present section does not claim to offer an exhaustive overview of the field, it is necessary to briefly outline the extent of existing research efforts. Existing research can be grouped in several distinct areas: (1) studies on victims of toxicity, (2) studies on toxic players, (3) studies on toxic games.

5.2.1 Studies on Victims of Toxicity
This kind of research, by far the most prolific domain, is focused at identifying common socio-demographic traits of the players that are most frequently victimized by anti-social behavior in online games as well as assessing the impact of the harassment. The ADL report [48] shows how toxicity is not restricted to but strongly tied to gender, race/ethnicity, and other player demographics. Türkay et al. [168] investigated how players define, experience, and deal with toxicity and found that players often rationalize such toxicity as a normal part of gaming. Hayday et al. [169] explored current experiences of identity and Esport community membership focusing on the ideological grounding, current practices and tensions present within the communities. Kuznekoff et al. [170] set out to determine how gamers’ reactions to male voices differ from reactions to female voices and they found that “the female voice received three times as many negative comments as the male voice or no voice. In addition, the female voice received more queries and more messages from other gamers than the male voice or no voice”. McLean et al. [171] explored female experiences of social support while playing online video games and they suggest that “a lack of social support and harassment frequently led to female gamers playing alone, playing anonymously, and moving groups regularly. The female gamers reported experiencing anxiety and loneliness due to this lack of social support, and for many, this was mirrored in their experiences of social support outside of gaming”.

This research proves that toxicity is more harmful to women, not only with respect to psychological well-being but also because of certain coping mechanisms such as not using voice chat or hiding their gender. The previous study was also confirmed by Eriksson et al. [172]. The authors, besides confirming that women are more affected than men, also showed how toxicity puts women at a disadvantage within the game itself when trying to achieve higher ranks, compared to men. An additional insight comes from Fox et al. [173], the authors
showed that harassment in general predicts women’s withdrawal from online games. Fordham et al. [174] demonstrated how exposure to gender stereotypes within games potentially causes negative attitudes about women in other stereotyped domains, such as Science, technology, engineering, and mathematics fields.

In our work, we do not focus on the victims of toxicity, but rather on detecting the toxic players, which is the next distinct area we discuss. However, we advocate for including the victims of toxicity in helping to further define and mitigate toxicity.

5.2.2 Studies on Toxic Players

Another area where academic research has focused on is studying the perpetrators, which involves trying to identify the socio-demographic markers as well as profiling and predicting their behavior both in-game and in physical life. Lemercier-Dugarin et al. [175] examined the relationship between toxicity and several potential predictors such as personality traits, emotion reactivity, and motivations to play. They found that younger age, being male, spending a lot of time playing per week, and being highly achieving increased the likelihood of resorting to toxicity. High emotional reactivity and being high in two dimensions of impulsivity (negative urgency and sensation seeking) increased the likelihood of toxic behavior. Shen et al. [176] examined individual and team-level predictors of toxicity in games using longitudinal behavioral data and they found that experienced and skillful players are more likely to commit toxic behaviors than newcomers, while losing teams and teams with high internal skill disparity among their members tend to breed toxicity. But the most interesting finding is that toxicity is somewhat contagious: exposure in previous games has been shown to increase the likelihood that a player will commit toxic acts in future games. Märtens et al. [177] employed a novel natural language processing framework to detect profanity in chat-logs and developed a method to classify toxic remarks, showing how toxicity is non-trivially linked to game success. This study was expanded by Traas [178]: he found that toxic teams lose more matches if they were already losing and win less matches if they were already winning. Additionally, Verschoor [179] showed how in-game events such as the number of times that a player has died in the last minute, or the number of times that a player’s team mates have died in the last minute, can predict toxicity in chat. Rodriguez [180] investigated how machine learning models would perform in automatically detecting toxicity, reaching an accuracy of instances classified
correctly of 81.92%. Unfortunately, the study is based on a very limited sample, hence the results cannot be considered representative of the population. Nevertheless, it shows that detecting toxicity automatically is a feasible task.

Thus, considerable efforts have been made, through surveys or behavioral data, to either understand or detect toxic players. Our work fits into the latter, with the important distinction that we use gameplay data. Additionally, the context of our work is on For Honor, which is a relatively less studied type of toxic game.

5.2.3 Studies on Toxic Games

The last line of research efforts is concerned with the affordances that seem to be necessary to allow some virtual environments and online games to breed toxicity. In particular, Kordyaka et al. [181] investigated why toxic behavior occurs by disjunctively testing three different theoretical approaches (social cognitive theory, theory of planned behavior, and online disinhibition effect). They propose a unified theory of toxic behavior, underpinned by the evidence that anonymity and behavior normalization fuel toxicity in a number of different contexts. Kou and Nardie [67] report that anti-social behavior is pervasive and problematic in many online venues and suggest regulating anti-social behaviors by examining the efforts of the game developer Riot Games, namely the “Tribunal System” that empowers players to judge misbehavior. Relatedly, Kou and Gui [182] looked at the practices of community members to report (or flag) toxic behavior in LoL. They find that players (1) distrust the flagging system, (2) use the system beyond its intended use for toxicity, and (3) use it socially (e.g., team members discuss and “gang up” to flag another member).

The present work considers, with the input from community managers and help of game designers, what game features should be considered in predicting toxic players. More importantly, our work may complement efforts reported to regulate or mitigate anti-social behaviors described above.

5.3 The Case for ‘For Honor’

According to [168], the most prominent features that make certain online multiplayer games outlets for toxic behavior are:
a) competitiveness: games where players compete with others, victory is paramount, and it feels like the game is not fun if it is not won;

b) anonymity: as players use nicknames and most likely will not meet directly, they feel free to say anything or act like there are no consequences;

c) counterfactual thinking: a psychological phenomenon to imagine possible alternatives to what actually happened, which in online multiplayer games means that players tend to blame others for unwanted events; and

d) negative social culture: as players spend time in communities where there is no empathy and having fun watching other people suffer is normalized, it is a matter of time to adopt anti-social behaviors.

MOBA games are “notoriously toxic” [183], most likely because they have all these aforementioned prominent features: they are highly competitive, depend on team-based efforts of anonymous players, and have developed a negative social culture over time. Therefore such games have been of focal interest in studying toxicity in games [198]. As for the game FH specifically, it has been known for its toxicity with some players even saying that “this game has the most toxic player base I’ve ever encountered in a video game” (Havoc1003 on Gamespot)\(^2\) or, similarly:

“The game has the most toxic community I’ve ever seen like everyone just starts insulting you in chat if you beat them despite the fact they’re just light spamming or if someone beats you they just keep calling you trash and insulating you out of nowhere. Or people start ganging up in brawls for no reason.” –KAAMG1 on r/forhonor\(^3\)

Or

“I was told once that he hoped id catch aids and malaria... hoped id get a death sentence but spend some time in jail to be raped and beaten and that when i finally decide to kill myself that id have hitler and the devil gang banging my body in the depths of hell. All because i through him off the ledge for 2v1 me.” –TrowserShnake on r/forhonor\(^4\)

\(^2\)https://gamefaqs.gamespot.com/boards/168620-for-honor/75998673

\(^3\) https://www.reddit.com/r/forhonor/comments/gsqoj0/toxicity_in_for_honor/

\(^4\) https://www.reddit.com/r/forhonor/comments/64xboe/how_toxic_do_you_think_the_for_honor_community_is/
Similar to other MOBAs, the prevalent toxicity in FH offers the opportunity to study this phenomenon in this particular game. We were able to do this by working with a unique dataset that includes the full behavioral data of nearly 1,800 ‘sanctioned’ players and comparing their behavior with unsanctioned players. Sanctioned here means that a player has been reported to the community managers by other players and they had been confirmed guilty of a breach of the game’s code of conduct and received a sanction. Sanctions differ in their degree of severity (i.e., the type of punishment, typically a warning, temporary ban, permanent ban, etc.). The opportunity we considered here is to study if we can behaviorally distinguish these sanctioned players from unsanctioned players.

5.4 Game Industry and Toxicity
As for “identifying a mechanism to combat toxic elements” [50], the industry has been focused for some time on devising strategies to curb toxicity in games [184-186]. Some of the largest gaming companies such as Blizzard and Riot have started creating systems to combat the unpleasantness in gaming communities, for example, players that are reported for racism or profanity can be temporarily muted. Most of the time these measures are insufficient as players can evade them easily by omitting letters or adding numbers or special characters [188]. Players can also report toxic players using in-game menus or the reporting options offered by Xbox Live and PlayStation Network. These reports can lead to banning abusive players. For example, Blizzard recently banned more than 18,000 Overwatch accounts for toxic behavior [50]. Riot Games has been studying and trying to reduce toxicity for many years by adding pro-social in-game tips to encourage positive interactions as well as implementing a system in which players are rewarded with in-game goods for sportsmanship and virtuous behavior. That brought down verbal abuse by about 6% and offensive language by 11% [187]. Blizzard’s Jeff Kaplan (Overwatch Lead Designer) reported a decrease of more than 25% in both players being abusive and matches containing abuse after adding features that encourage positive comments and allowing players to create filters for their online matchmaking [188]. Ubisoft, where toxicity management is a priority, started a task force with the end goal to track negative player behavior, manage players that behave poorly, and implement features that will encourage players to improve their behavior such as chat improvements, and team kill tracking [189]. Overall, the various efforts from the industry seem to be concerned with implementing rapid and solid reporting systems and engineering pro-social behaviors.
There have been calls, however, both in academia and in industry, to leverage machine learning [21] to help detect toxicity as this might be “a rising tide” [50] or at the very least another mechanism to combat toxicity. For example, in examining toxicity among eSports players Türkay et al. [168] concluded that:

“Game companies may also need to investigate systems to combat toxicity that do not rely on player reporting. With the rise of machine learning, perhaps in the future, game companies will not need active reporting to target toxicity.”

The limited efforts thus far have focused on Natural Language Processing and text data, and thus verbal actions [177, 190]. These efforts are similar to how toxicity issues are addressed in other media, such as YouTube [191] and Twitter [192]. However, games are different from such media as users do not only demonstrate toxicity through verbal actions but also through behavioral actions [167]. While we include chat actions in our approach (i.e., not the content, just the behavioral act of chatting), our work in detecting toxic behavior is focused on such behavioral actions, as a form of player modeling [21]. To our knowledge, this is the first study that attempts to detect toxicity through gameplay on a large scale. We do not advocate for the complete automatization of toxicity detection; we merely propose to supplement the manual efforts of the community managers in terms of proactive flagging of toxic players. We recommend that a final human verification is necessary to close the loop and impose a sanction.
III RESEARCH METHODOLOGY

This section describes the overarching research methodology, the methods used and how they were applied, and a methodological reflection and discussion of other methodologies that were considered and are as valid for this thesis.

Based on the definitions provided by Mathiassen et al. [193] this section outlines the scope and research methods employed to address our research questions motivating fitness in criteria for Action Research.

1. Research Elements

Given the unique placement of the described area of concern we can conclude that it involves the practice of making games (design for entertaining interaction), the practice of maintaining games (continuous service and updates) and a multitude of theories that attempt to make sense of the psychosomatic experience of interacting with the game as well as the extent and nature of historical ramifications affecting it. This rather broad and interdisciplinary area requires a diverse set of Practice and Theory driven premises however, study of games goes far beyond the scope of the present study. With player experience and behavioral modelling at its center, this enquiry would include aspects of game design, player data collection, treatment, and analysis as well as theories of affect, psychological development and motivation insofar as they involve the quality of player experience.

The problem setting then, would consist of investigation into player modelling and practical challenges it imposes on the process of data collection and treatment (what type of data? What type of treatment? What type of analysis?), but also addresses the theoretical gaps of the relationship between player demographics, affective state (from emotions to motivations) and player action markers. Therefore, this study mainly employs a deductive inference style, relying on large volumes of behavioral and self-reported data for confirmation even occasionally when inductions are made. Adapting psychoneurological concepts and presumed equivalency of virtual stimulus, may be considered as rare examples of abductive inference (retroduction) utilized in this study.
2. Methods Explanation

Although framed above according to [193], as types of Action Research, nuances of current study could also be framed as Research for design [194], Case study [195] and questionnaire design [196], depending on the settled viewpoint of the observer.

Design research in game studies may be mediated by Player motivation as a central aspect of player experience which, in turn, is the key underlying objective of game design. Understanding the mechanisms of player motivation in a computational fashion would, arguably, allow the design of captivating games that foster long-term engagement. Although the literature is rich on psychological frameworks of player motivation [20, 24, 25, 104] and surveys that attempt to measure aspects of motivation [32, 53, 55, 58], the notion of predicting motivation quantitatively through the means of machine learning has only been introduced very recently [9] as a part of this thesis.

2.1 Questionnaire Design

Questionnaire design [196] within Player Motivation has been explored in two distinct mindsets both relying on survey data. Though they use fundamentally different approaches to process this data and derive relevant dimensions of the player’s motivation; The Ubisoft Perceived Experience Questionnaire (UPEQ; PAPER VI) [32], a representative of game evaluation type, uses a theory-driven framework to acquire aggregated scores of four factors of self-determination following the theory of Deci et al. [97], Ryan et al. [20], and Chen et al. [98]. UPEQ relies on Self Determination Theory (SDT) [97], which is a popular theory of motivation describing the phenomenon through the perceived satisfaction of psychological needs. These psychological needs are competence, a need for mastery and completion, autonomy, a need for meaningful choices and self-determined action, and relatedness, a need for connection with other humans—and in the case of video games, expressive non-player agents [20]. UPEQ extends these factors with presence, which is the illusion of an unmediated experience that occurs when players stop perceiving the existence of the game’s medium [104, 109]. UPEQ uses 5-point Likert-scales to measure these dimensions through a 24-item survey (with 7 items for each of competence, autonomy, and relatedness, and 3 items for presence).

The survey was developed based on the Basic Need Satisfaction Scale(s) [98], addressing limitations of contemporary surveys for measuring SDT in games [55], [53-54]. UPEQ has been shown to be a reliable measure and has been used in
studies (a) as input to predict solo playtime, group playtime, and money spent [32] and (b) as output to model motivation based on gameplay [9]. The main strength of this approach is its reliance on a well-established theory which led to many critical observations in multiple domains, proving the validity of the framework. The limitation of the method, however, also stems from its heavy reliance on a top-down framework. Although theory-driven approaches excel at abstracting observations for explanatory analysis, predictive modelling of these constructs may often be challenged due to the deviation of such constructs from the ground truth. On the other hand, and on the side of player preference methods, The Model of Player Motivation, instead, is a data-driven method which relies on factor analysis that derives the underlying patterns of player motivation following the work of Yee et al. [58]. Although the model used by Yee is well-tested and robust, higher reliability can be achieved on smaller datasets by tuning the model to the specific game and player population at hand through reapplying the same analytics method and deriving data-specific factors.

2.2 Design Research
If we consider PAPER IV [9] as design research and the game as design artifact, then evaluation of the intended design through the proposed assessment of our artifact would be applied in the environment of design processes and strategies but also informs game studies knowledge base by introducing new techniques and instruments according to the research framework suggested by [194]. Motivation is a crucial element of player experience and research, be it player profiling or modelling. Understanding what drives players can help game designers and advise the development of adaptive and generative systems [92]. While some studies have incorporated motivational survey data into their user models to predict other gameplay outcomes, such as churn or enjoyment [93-95], relatively few focused on motivation as the target output of their models. As a rare example Canossa et al. applied regression to model player drives [24] based on Reiss’ Motivational Profile [96]. The methodology followed also appears to be generalizable across any game as long as gameplay and motivation survey scores are available.

In the proposed methodology for PAPER IV, we opt for preference learning (PL) methods and ranked based correlations because of the strong connection between this ordinal machine learning paradigm and how player experience operates in games. In essence, PL models certain psychological processes by focusing on the differences between occurrences instead of their absolute values [84]. This approach falls much closer to the players’ cognitive processes—e.g.,
anchoring-bias [113-114], adaptation [112], habituation [115], and other recency-effects [116]—that help them evaluate their own experience internally.

During recent years there is growing evidence supporting the strength of preference learning for modelling emotions and user experiences both on a conceptual [84] and a technical basis [117-118, 197-198]. Conceptually, treating subjective-defined ground truth data as ordinal variables bring the representation of data closer to the players' underlying true attitudes [84]. This is based on the observation that one's affective state is always relative to a certain adaptation level, i.e., emotions have a shifting baseline based on previous experiences. Even though traditionally psychological data is treated as nominal categories [199], by now there is ample evidence suggesting higher validity and reliability when labels of emotion or experience are treated as ordinal [197-201]. Indicatively, Martinez et al. [197] compared classification and PL across a number of modelling tasks found that ranking yields more robust models than classification. A number of other studies compared the processing of affective annotations as both ratings (e.g., Likert items) and rankings and found that:

1) first-order data processing (i.e., ranks) yields higher reliability and inter-rater agreement and

2) second-order processing of the absolute rating values was also beneficial with regards to both reliability and validity [84, 117,118].

Recently Camilleri et al. [89] applied PL for obtaining general models of players across games and Melhart et al. [198] showed that ranking Support Vector Machines (SVMs) are more robust than support vector classifiers in predicting arousal across dissimilar affective corpora.

2.3 Case Studies
Whether it is an exploration of motivational effects of age or validation of design choices through player actions, many of the work that we present, have initialized as Case Studies [195], although our goal is to present frameworks on how such case studies could be conducted. PAPER II, for example, explores the relationships between age and generation groups and players' attributes such as game ratings, motivation, and in-game behaviors of the players. By including a large age span and making comparisons between different generational categories, that aims to expand on the approach of the previous research and contribute with new insights to factors that affect playtime patterns of players based on their motivations. While operationalizing playtime is straightforward, social behaviors are less so. Because the clearest form of exhibiting social behaviors is playing together, in PAPER V we decided to operationalize this aspect as social play,
specifically the percentage to which one plays a multiplayer game collaboratively with others versus alone. Put as a research question: Do influencers in online games affect connected players’ playtime and social play more strongly than average? And are disruptive behavior and their characteristics (type, severity) are predictable by scrutiny of in-game behavior? Answer to these questions may solidify our general models of motivation and would enable the assessment of previously unseen games providing great value to both industry stakeholders and academic researchers working towards general player modelling [21]. 

PAPER I & PAPER III make structural proposals for detection of toxic behavior and character equipment sets (playstyles) respectively. Such studies could be used to evaluate other theoretical frameworks in the context of game behavioral data via field studies and employment of these methods may inversely influence game design strategies and procedures which grants the opportunity for theory and data informed design research.

PAPER I, aims to establish a framework for prediction of toxic behavior through a case study [195] on For Honor (FH) an online multiplayer battle arena (MOBA) by (1) appropriately sampling players for comparison, (2) comparing multiple methods of categorization and feature selection, and then (3) using machine learning (in particular, random forests) with the purpose of predicting toxicity for a large sample of data from the game For Honor FH, based on labels derived from the community managers’ sanction procedures. We compared behavioral aspects of gameplay such as match performance, chat actions, and playtime patterns to predict not only the binary outcome of being sanctioned by community managers for involvement with toxic behavior but also the type of toxic behavior and severity of sanctions imposed on players. In other words, our aim was not only to determine if we can behaviorally distinguish ‘sanctioned players’ vs. ‘unsanctioned players’ but also the degree to which it is possible to identify in-game behaviors for players that have been labelled as toxic.

In PAPER III, [202], an exploratory, descriptive, and interpretive case study [195], we introduce the BaT (Build as Text) approach to categorize player character configurations in Tom Clancy’s the Division 2 (TCTD2), an online multiplayer shooter. We explain the principle of our approach, the data encoding process, the algorithm, and the final entropy-based allocation scheme of data points to clusters. As character builds affect character attributes and are highly reflective of playstyle, categorization of character builds can be seen as a proxy to categorization of player behavior that follows these characteristics. Therefore, performance of the model in comparison of behavioral descriptors such as health,
armor, skill power, offense, defense, utility, and overall playtime split is examined to show that the different discovered builds are indeed leading to specific character attributes and playtime behavior. We also show that the BaT approach is efficient in discovering both known (defined) and unknown (emergent) builds. Finding better ways to detect relative stability in character configurations, and study builds evolution in time per player deserves a full research work. Alternatively, monitoring the evolution of successive clustering sets after different game updates could also lead to valuable timely insights concerning the game’s meta evolution, i.e. how players adapt their builds to in-game mechanics changes, new items/skills, and existing items modifications for balancing. We expect to be able to engage in design and implementation of live player models in controlled experiments for adaptive content and individually trained recommender systems [203].
IV METHODS

This section describes the methods, algorithms and approaches relevant to the work presented in this thesis.

1. Correlations and Regression

Correlation is a marker of dependency and relationship between two variables. The common approach to calculate the average correlations is the Pearson’s technique outlined as follows:

\[
r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}
\]

- \(r\) = correlation coefficient
- \(x_i\) = values of the x-variable in a sample
- \(\bar{x}\) = mean of the values of the x-variable
- \(y_i\) = values of the y-variable in a sample
- \(\bar{y}\) = mean of the values of the y-variable

However, because we deal with multiple rank-based measures in our experiments as explained in Chapter III, Section 2.2, our main reported correlation statistics is Kendall’s tau, which is calculated so [204]:

\[
\tau_B = \frac{n_c - n_d}{\sqrt{(n_0 - n_1)(n_0 - n_2)}}
\]

- \(n_0 = \frac{n(n - 1)}{2}\), where \(n\) is data size
- \(n_c\) = number of concordant \((x,y)\) pairs
- \(n_d\) = discordant pairs
- \(n_1 = \sum_j \frac{t_j(t_j - 1)}{2}\) (\(t_j\) = number x values tied at jth value)
- \(n_2 = \sum_k \frac{u_k(u_k - 1)}{2}\) (\(u_k\) = number y values tied at kth value)
Linear regression was also occasionally used to model the constant relationship between non-ordinal variables. Equation below shows the basics of a linear regression equation:

![Linear regression equation](image)

2. Scale Reliability Measures

Statistical reliability of a test is measured in a myriad of approaches and algorithms. It is worth noting that while scales can be statistically reliable, they may not be valid, meaning that they may robustly measure a construct that is not the intention of the scale in the first place. We discussed that limitation of gaming-questionnaires by pointing out the importance of theory and clear definitions of constructs in Chapter II, Section 1.

For our test reliability measure, we selected two measures of internal consistency, one popular measure that overestimates reliability (Cronbach’s alpha) [205], and another which underestimates it (Rasch reliability) [206].

**Cronbach’s alpha** is an Average of inter-item correlation in the questionnaire that is the average of correlations between each two items. To calculate average item-to-total correlation, first we create a “total” item by adding the values of all items, compute the correlations between this total item and each of the six individual items, and then, average all the correlations. Cronbach’s alpha, a reliability measure designed by Lee Cronbach in 1951, factors in scale size in reliability estimation, calculated using the following formula:

\[
\alpha = \frac{K}{K-1} \left( 1 - \frac{\sum_{i=1}^{K} \sigma_{Yi}^2}{\sigma_Y^2} \right)
\]

where \(K\) is the number of items in the measure, \(\sigma_X^2\) is the variance (square of standard deviation) of the observed total scores, and \(\sigma_Y^2\) is the observed variance
for item $i$. The standardized Cronbach’s alpha can be computed using a simpler formula:

$$\alpha_{\text{standardized}} = \frac{K\bar{r}}{1 + (K - 1)\bar{r}}$$

where $K$ is the number of items, $\bar{r}$ is the average inter-item correlation, i.e., the mean of $K(K-1)/2$ coefficients in the upper triangular (or lower triangular) correlation matrix.

**Rasch reliability** also known as ‘person reliability’ on the other hand, is a size independent measure that examines the criterion for successful measurement by looking into observed variances and squared standard errors of the measured items, as such [207]:

$$R_\kappa = 1 - \frac{\sum(\text{Measure Standard Error}^2 / N)}{\text{Variance of Observed Measures}}$$

### 3. Factor Analysis

Factor analysis is used to reduce the dimensionality of the data at hand in order to look at the variability to be explained using fewer variables. For conducting **principal component analysis**, which aims to derive the minimum number of factors that can explain the most variability of the sampled data using unsupervised rotations, there are additional tests that determine whether the dataset is suitable for such an operation.

The first common test of factorability is to pass a minimum inter-item correlation, recommended to be above .3 for a good fit. Second, the Kaiser-Meyer-Olkin, a measure of sampling adequacy for factor analysis, measures the proportion of variance among variables that may be common, suggesting that lower proportions are better suited for factorability.

KMO test is formulated as follows:

$$KMO_j = \frac{\sum_{i \neq j} r_{ij}^2}{\sum_{i \neq j} r_{ij}^2 + \sum_{i \neq j} u_{ij}^2}$$
Where $R = [r_{ij}]$ is the correlation matrix and $U = [u_{ij}]$ is the partial covariance matrix [208].

Bartlett’s test of sphericity, used to verify homogeneity of variances via comparing correlation matrix of items with their identity matrix (or unit matrix, a $n=N$ matrix with diagonal 1s and zero for other values). If there are $k$ samples with sizes $n_i$ and sample variances of $S_i^2$, then Bartlett’s test statistic is:

$$\chi^2 = \frac{(N - k) \ln(S_p^2) - \sum_{i=1}^{k} (n_i - 1) \ln(S_i^2)}{1 + \frac{1}{3(k-1)} \left( \sum_{i=1}^{k} \frac{1}{(n_i - 1)} - \frac{1}{N-k} \right)}$$

Where $N = \sum_{i=1}^{k} n_i$ and $S_p^2 = \frac{1}{N-k} \sum_i (n_i - 1) S_i^2$ is the pooled estimate for the variance [209].

Other factorability measures may include the diagonals of the anti-image correlation matrix, and the communalities to further confirm that each item shares some common variance with other items but not showing an excessive overlap. Given these overall indicators, factor analysis can be deemed suitable.

Once we have shown that the dataset is ready for component analysis, then method of estimation and rotation are the next steps in performing component analysis. Method of estimation refers to the method that is used to obtain the extracted solution, and throughout this study, Maximum Likelihood is selected for that purpose. Rotation technique is the how different iterations of the factor analysis vary and, in this thesis, we used an orthogonal approach known as Varimax rotation.

For selecting the number of factors, Eigenvalues and Scree plots are used to determine the variance explained by the extracted components. In both methods, the goal is to select fewer factors that explain the most variability. For variables $K1$ to $Kn$, a principal component analysis syntax for IBM’s SPSS [210] with Maximum Likelihood and Varimax rotation (of $i$ iterations) were used.
4. Comparing Means

In multiple studies, we identify samples of players that may be different in their behavioral and self-reported measures. Sometimes, even the identification itself is achieved by showing that the selected sample is indeed different from the control group or the rest of the population. Therefore, comparing means may apply to testing re-testability of a scale by comparing means of random samples, or comparing differences in gameplay behavior of two consecutive generations or showing the effects of pairing with power users if a game a year after joining their inner circle. Whether we use One way (Analysis of Variance) ANOVA, Effect size [211], Paired-samples t-tests [212], or The Kruskal-Wallis H test [213], the goal is to show the significance of difference between the sample means.

ANOVA is appropriate for finding differences or inconsistencies among sample means assuming each group fulfills criteria for normal distribution and a constant variability within the sample. Numerically, one way ANOVA is a generalization of the two-sample t test. The F statistic compares the variability between the groups to the variability within the groups:

\[ F = \frac{MST}{MSE} \]

\[ MST = \frac{\sum_{i=1}^{k} \left( T_i^2/n_i \right) - G^2/n}{k - 1} \]

\[ MSE = \frac{\sum_{i=1}^{k} \sum_{j=1}^{n_i} Y_{ij}^2 - \sum_{i=1}^{k} \left( T_i^2/n_i \right)}{n - k} \]

where \( F \) is the variance ratio for the overall test, \( MST \) is the mean square due to treatments/groups (between groups), \( MSE \) is the mean square due to error (within groups, residual mean square), \( Y_{ij} \) is an observation, \( T_i \) is a group total, \( G \) is the grand total of all observations, \( n_i \) is the number in group \( i \) and \( n \) is the total number of observations. Another approach using the same assumptions, is paired sample t-test; The test statistic is calculated as:
where \( \bar{d} \) is the mean difference, \( s^2 \) is the sample variance, \( n \) is the sample size and \( t \) is a Student’s \( t \) quantile with \( n-1 \) degrees of freedom [212].

When comparing several independent random samples, we will use **The Kruskal-Wallis H test** for non-parametric alternative to One way ANOVA. The Kruskal-Wallis test statistic for \( k \) samples, each of size \( n_i \) is:

\[
T = \frac{1}{s^2} \left( \sum_{i=1}^{k} \frac{R_i}{n_i} - N \frac{(N + 1)^2}{4} \right)
\]

where \( N \) is the total number (all \( n_i \)) and \( R_i \) is the sum of the ranks (from all samples pooled) for the \( i^{th} \) sample and:

\[
S^2 = \frac{1}{N - 1} \left[ \sum_{\text{all}} R_{ij}^2 - N \frac{(N + 1)^2}{4} \right]
\]

The null hypothesis of the test is that all \( k \) distribution functions are equal. The alternative hypothesis is that at least one of the populations tends to yield larger values than at least one of the other populations [213].

Finally, as a measure of Effect Size, which describes mean difference of as effect, we employ **Cohen’s \( d_s \)**, a measure of mean difference between two groups of independent observations for the sample:

\[
d_s = \frac{X_1 - X_2}{\sqrt{\frac{(n_1-1)SD_1^2 + (n_2-1)SD_2^2}{n_1+n_2-2}}}
\]
Where the numerator is the difference between means of the two groups of observations and the denominator is the pooled standard deviation [214].

5. Social Network Analysis (SNA)

Social network analysis is the practice of network modeling to establish dynamics of a social network, connections, and evolution of them [215]. In our work, it refers to graph theory-based measures that was used to identify influencers in the in-game social network of the game, Tom Clancy’s the Division. We used conventional SNA techniques to identify influencers in our data set, a practice that has been done before in the context of both social networks at large [156] and multiplayer games specifically [216]. Given that there is no agreement on which individual measure to utilize when identifying influencers, we used six different measures of centrality: closeness, betweenness, eigenvector, in-degree, out-degree, and PageRank. All the sets of players identified by each centrality measure are intersected with each other to identify the players that are considered central for each of the six measures. In this work, we define influencers as players that satisfy all these six conditions.

We then plotted the resulting influencers onto a network graph where the nodes represent the players, and the color of a node indicates the community (module) the node belongs to. The resulting super-graph is depicted in Figure 2. The size of the nodes is proportional to the importance of a player; hence influencers display a much bigger size than normal players. Details of our method are described below.
Fig 2. The 49 identified influencers mapped on the super-graph using conventional SNA techniques.

Identifying Most Central Influencers. We first computed centrality measures, which aim to quantify the “influence” of a particular node within a network. Our aim was to identify within each community which player may be influential. To accomplish this, we considered the following measures:

1) **Closeness centrality**: how easily accessible a node is to all other players, represented as the length of the shortest path. The speed by which a player accesses all other players ranges between 0 and 1. We selected all nodes with values > 0, resulting in 182 players with the following formula:

normalized closeness centrality(node) = (number of nodes - 1) / sum(distance from node to all other nodes)

\[ C(x) = \frac{N}{\sum_y d(y, x)} \]
where \( d(y, x) \) is the distance between vertices \( x \) and \( y \) and \( N \) is the number of nodes.

(2) **Betweenness centrality**: it represents the number of shortest paths to other players, or how likely a player is the most direct route between two other players. The range fell between 0 and 168. We selected all nodes with values > 0; resulting in 78 players with the following equation:

\[
C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}}
\]

Which represents dividing total number of shortest path from node \( s \) to \( t \) through node \( v \) by total number of short paths from \( s \) to \( t \).

(3) **Eigenvector centrality**: while degree centrality counts all connected nodes equally, eigenvector centrality treats connected nodes differently based on their “im- portance,” or how well a player is connected to others. The range is between 0 and 1. We selected all nodes with values > 0.05; resulting in 198 players. For a given graph \( G := (V, E) \) with \( |V| \) vertices let \( A = (a_{v,t}) \) be the adjacency matrix, i.e. \( a_{v,t} = 1 \) if vertex \( v \) is linked to vertex \( t \), and \( a_{v,t} = 0 \) otherwise. The relative centrality score of vertex \( v \) can be defined as:

\[
x_v = \frac{1}{\lambda} \sum_{t \in M(v)} x_t = \frac{1}{\lambda} \sum_{t \in G} a_{v,t} x_t
\]

where \( M(v) \) is a set of the neighbors of \( v \) and \( \lambda \) is a constant. With a small rearrangement this can be rewritten in vector notation as the eigenvector equation.

(4) **In-Degree (prestige)**: number of connections to a node from others. These are players invited most often to groups. Range is between 0 and 5 and we selected all nodes with values => 2, resulting in 371 players.

(5) **Out-Degree**: number of connections from a node to other nodes. These are group creators that frequently invite other players. Values range between 0 and 630 and we selected all nodes with values > 0, resulting in 165 players. In- and out-degree together tell us how many players a certain player can reach directly.

For a vertex, the number of head ends adjacent to a vertex is called the **indegree** of the vertex and the number of tail ends adjacent to a vertex is its **outdegree**. Let \( G = \)
\((V, A)\) and \(v \in V\). The indegree of \(v\) is denoted \(\text{deg}^- (v)\) and its outdegree is denoted \(\text{deg}^+ (v)\). A vertex with \(\text{deg}^- (v) = 0\) is called a source, as it is the origin of each of its outcoming arrows. Similarly, a vertex with \(\text{deg}^+ (v) = 0\) is called a sink, since it is the end of each of its incoming arrows. The degree sum formula states that, for a directed graph:

\[
\sum_{v \in V} \text{deg}^- (v) = \sum_{v \in V} \text{deg}^+ (v) = |A|.
\]

If for every vertex \(v \in V\), \(\text{deg}^+ (v) = \text{deg}^- (v)\), the graph is called a balanced directed graph.

\(6\) **PageRank**: what fraction of players can be reached via directed paths. It uses links as a measure of importance. Each node is assigned a score based on its number of incoming links (its “in-degree”). These links are also weighted depending on the relative score of its originating node. The result is that nodes with many incoming links are influential, and nodes to which they are connected share some of that influence. The scores range between 0.000063 and 0.000059. We selected all nodes with values > 0.00006, resulting in 178 players.

The intersection between the 182 players with highest closeness centrality, the 78 players with the highest betweenness centrality, the 198 players with the highest eigenvector centrality, the 371 players with the highest in-degree centrality, the 165 players with the highest out-degree centrality, and the 174 players with the highest PageRank score returned 49 players. These 49 players will be referred to as influencers from now on. It is important to note how intersecting across the six measures of centrality gives us a very conservative selection of players since in order to be considered influencers they must satisfy all six criteria. Furthermore, as Fig. 2 shows, these 49 players map to a very large extent onto the sub-communities that form the heart of the network.
6. Clustering methods

In this section we introduce the clustering methods used in various experiments presented in this thesis.

Centroid-based clustering
Centroid based clustering represents clusters by a central vector which might not be a member of the dataset, objects then, are assigned to the clusters based on the minimum proximity calculated by squared distance from the central vector. K-means clustering (k-Nearest Neighbors or k-NN) is one of the most popular and a non-parametric classification algorithm; its input consists of the k closest training examples in the dataset whereas its output determines the class membership. For example, \( k = 1 \) implies that the data object (the input of 1-NN) is assigned to the class of that single nearest neighbor.

Density-based clustering:
Density bade spatial clustering applications with noise (DBSCAN), is a popular clustering algorithm that groups variables based on their distance (usually Euclidean) and number of points close to one another. Epsilon value in DBSCAN refers to the appointed distance that would be the threshold for considering two data points as neighbors and ‘minimum points’ is the minimum number of points with qualifying epsilon distance that are eligible to form a dense region and be considered a cluster [217].

The inherent limitations of numerical clustering methods lie in either the necessity for encoding categorical features in a numerical vector space (e.g., for k-means) or the need for specialized distance metrics that can deal with both numerical and categorical features - such as the Gower distance [218] for DBSCAN - which hardly preserve semantic distance between the original objects. Additionally, it is known that the usual Minkowski distances (e.g., Euclidean) used to compute similarity matrices do not behave well in high dimensional spaces, a phenomenon commonly referred to as the curse of dimensionality [219]. A Python algorithm sample for DBSCAN, with epsilon value of 0.5 and 5 as minimum number of members, is presented below:

```python
from sklearn.cluster import DBSCAN
DBSCAN(eps=0.5, *, min_samples=5, metric='euclidean', metric_params=None, algorithm='auto', leaf_size=30, p=None, n_jobs=None)
```
Categorical Clustering

Categorical clustering is used when attributes of data objects are not inherently comparable. One such method is Latent Dirichlet Allocation (LDA) algorithm, which is a classical text clustering algorithm [220]. It is suitable for clustering a mix of text and numerical data since the clusters it produces - also called topics - are provided as an ordered list of keywords found in the dataset. The order of keywords for a given topic is defined by their relative importance in said topic. This output is extremely convenient for interpretation, since any non-data-science-proficient person having a good knowledge level of the domain can make sense of these results and interpret each ordered keyword list as a category.

Another advantage of this algorithm is that it produces soft clusters instead of hard partitions of the data. For example, for each character configuration in a game, we obtain a probability distribution over all discovered topics (in contrast to just being assigned a single topic). In our context, this is advantageous because it enables us to obtain the most representative character configuration for each cluster/build. Examining the most representative configuration helps analysts and game designers understand and interpret the build much better than when looking at centroids obtained by numerical clustering approaches, which may not represent valid character configurations.

Furthermore, although classical topic modelling approaches do not always create semantically meaningful topics, in this work we apply them to keywords data and not natural language. Application of LDA to this type of data is particularly relevant, due to its underlying assumptions of probabilistic distributions over words and topics, considered under the 'bag-of-words' paradigm.

<table>
<thead>
<tr>
<th>Hyper-parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>alpha</td>
<td>'asymmetric'1</td>
</tr>
<tr>
<td>eta</td>
<td>'auto'</td>
</tr>
<tr>
<td>chunk_size</td>
<td>1000</td>
</tr>
<tr>
<td>passes</td>
<td>10</td>
</tr>
<tr>
<td>workers</td>
<td>12</td>
</tr>
</tbody>
</table>

1 alpha='auto' not available when using Distributed=True.

Several implementations of LDA are available in the open-source ecosystem, most of them having excellent performances in terms of computing time and memory. In this paper, we use Python Gensim library [221], but other ones (such as Spark
MLLib) would yield similar results. Used values for hyper-parameters can be found in Table II. We use the distributed variant of LDA to take advantage of multi-core computers at our disposal and control the random seed for more reproducible results.

7. Predictive Analysis

In multiple studies presented in this body of work, machine learning methods were employed in order to examine the importance of collected variables in predicting cohort information of players, their motivation or the type of activities that indicates potential for disruptive behavior. In this section we introduce specific models and their properties.

Artificial Neural Networks

Trying to mimic the way that biological neurons signal to one another, Artificial Neural Networks (ANNs) are deep learning algorithms comprised of a node layer, one or more hidden layers, and an output layer. Each node has a weight and a threshold for activation and sending data to the next layer. Therefore, the abundance of training data over time will increase the probability of activation of nodes and increased accuracy. Here, we focus on a feed forward, multilayer perceptron neural network, which is a logistic regression classifier where the input is first transformed using a learnt non-linear transformation. This transformation projects the input data into a space where it becomes linearly separable. This intermediate layer is referred to as the hidden layer [222].

Support Vector Machines

While support vector machines are usually used for classification problems by defining a hyperplane and decision boundaries of each class (and kernel definition for lowering computational cost in higher dimensions), they can be also used for predicting which side of the hyperplane does the new data point fall or optimizing the hyperplane to include maximum number of data objects within its decision boundary [223]. We used SVM algorithms to predict Generational class from gameplay data or ranking relationship between subjects in an ordinal dataset for prediction of Motivations from gameplay metrics.

Preference Learning

Learning order relations on a series of objects in order to predict preferences is the task of preference learning algorithms. As we explained in Chapter II, Section

---

the ordinal nature of motivational data we use throughout, it is suitable to transform one datapoint per player structure of our motivational datasets, since individual feature vectors used for the preference learning task are independent. This means that during the preparation of the Preference Learning (PL) experiments, each datapoint is compared to every other point during the pairwise transformation of the dataset. This transformation applies a preference threshold (Pt) parameter, which controls the margin of significance under which two datapoints are considered equal. The purpose of a threshold Pt is to counter the noise in the ground truth data which can skew modelling results. Additionally, to translate the relationship of datapoints into preference relations, this step also creates new datapoints for the machine learning task. Furthermore, because each pairwise comparison creates two new datapoints—describing the preference relation in both directions—the transformation balances the baseline of the classification task to 50% accuracy.

8. **Ensemble models**

In a couple of occasions, we employ ensemble models to improve our prediction-classification tasks. Ensemble models take a weighted vote of different predictors to learn to classify new data points. Whether it is an exploratory combination of classifiers for predicting generational classes or stacking randomly generated decision trees to form Random Forests, the goal is to achieve more robust predictors.

**Random Forests**

Decision trees are non-parametric algorithms that split a node into two (or more) sub nodes to increase the homogeneity of data by recursively splitting our training samples. A Random Forest (RF) then, is an ensemble learning method, which operates by constructing a number of randomly initialized decision trees and uses the mode of their independent predictions as its output. Decision trees are simple learning algorithms, which operate through an acyclical network of nodes that split the decision process along smaller feature sets and model the prediction as a tree of decisions. In our application, we used the RF implementation available at the randomForest R library.

---

8 [https://cran.r-project.org/web/packages/randomForest/](https://cran.r-project.org/web/packages/randomForest/)
V RESULTS

This chapter is a presentation of summary of the results achieved by studies in topics discussed in Chapter II. In the first section, we review the results of our proposed questionnaire for measuring experience and psychological need satisfaction. In the next section, we summarize cohort specifications of motivation and playtime patterns via the generational analyses. Third section presents the difference in modelling motivation by comparing results from two different approaches but also covers a classification of player actions which will be expanded by adding an outline of categorical clustering approach in section four. Effects of Social play and outsized participation in online communities in short and long-term are shown by comparison in section five though the final section focuses on the results of detection models for identifying type and severity of disruptive social behavior.

1. Questionnaire Design and Validity

In this section, we present the development of a concise Self Determination Theory (SDT) based questionnaire specific to the evaluation of video games. This model, UPEQ, includes only basic needs of autonomy, competence and relatedness. These constructs regardless of their limitations, are correlates of sustainable positive interaction with the game [20]. Hence, they could benefit game developers in offering them feedback that is not only game-oriented and actionable but also does not hinder the creative process of game development.

Through two large-scale studies we have investigated the internal consistency of the model, demographic and individual effects and correlates of need satisfaction, and sense of physical transportation into the game. In the first study, we examined the development and scale reliability evaluation of survey items applied to more than 20 popular video game titles released by November 2015. The second study investigated the coherence of self-reported need satisfaction of players with their recorded in-game behavior such as playtime and money spent in the game. It also shows the feeling of physical transportation into the game and its relationship with need satisfaction. In conjunction, these studies aim to present a reliable alternative for video-game evaluation, which is not only theoretically well-established but also effective in practice.

As explained in Chapter, III Section 2.1, Cronbach’s alpha and Rasch reliability test, as two measures of assessment of scale reliability, [205-206], indicated that UPEQ fulfills the criteria for the internal consistency of a scale. Then, factorability tests
and subsequently factor analysis of 21 items of UPEQ confirmed the representation of the theoretical structure within the factors. Table III shows the resulting factor loadings.
<table>
<thead>
<tr>
<th>Factor</th>
<th>Item</th>
<th>Correlation within factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Autonomy</td>
<td>I was free to decide how I wanted to [play].</td>
<td>.715</td>
</tr>
<tr>
<td>Autonomy</td>
<td>I could approach [the game] in my own way.</td>
<td>.722</td>
</tr>
<tr>
<td>Autonomy</td>
<td>The game allowed me to [play] the way I wanted to.</td>
<td>.694</td>
</tr>
<tr>
<td>Autonomy</td>
<td>I had important decisions to make when [playing].</td>
<td>.777</td>
</tr>
<tr>
<td>Autonomy</td>
<td>The choices I made while [playing] influenced what happened.</td>
<td>.798</td>
</tr>
<tr>
<td>Autonomy</td>
<td>My actions had an impact on the [game].</td>
<td>.722</td>
</tr>
<tr>
<td>Competence</td>
<td>With time, I became better at [playing].</td>
<td>.655</td>
</tr>
<tr>
<td>Competence</td>
<td>My [gaming] abilities have improved since the beginning.</td>
<td>.645</td>
</tr>
<tr>
<td>Competence</td>
<td>My mastery of the [game] improved with practice.</td>
<td>.655</td>
</tr>
<tr>
<td>Competence</td>
<td>I was good at [playing].</td>
<td>.683</td>
</tr>
<tr>
<td>Competence</td>
<td>I felt competent at [playing].</td>
<td>.660</td>
</tr>
<tr>
<td>Competence</td>
<td>I felt very capable and effective when [playing].</td>
<td>.684</td>
</tr>
<tr>
<td>Relatedness</td>
<td>I really like the people I play with.</td>
<td>.601</td>
</tr>
<tr>
<td>Relatedness</td>
<td>I consider players I regularly interact with to be my friends.</td>
<td>.546</td>
</tr>
<tr>
<td>Relatedness</td>
<td>Other players are friendly towards me.</td>
<td>.531</td>
</tr>
<tr>
<td>Relatedness</td>
<td>What other players did in the game had an impact on my actions.</td>
<td>.737</td>
</tr>
<tr>
<td>Relatedness</td>
<td>I had to adapt my actions to other players' actions.</td>
<td>.808</td>
</tr>
<tr>
<td>Relatedness</td>
<td>I was paying attention to other players' actions.</td>
<td>.831</td>
</tr>
<tr>
<td>Relatedness</td>
<td>I felt close to some of the characters.</td>
<td>.625</td>
</tr>
<tr>
<td>Relatedness</td>
<td>I was bonding with some of the characters.</td>
<td>.636</td>
</tr>
<tr>
<td>Relatedness</td>
<td>I cared about what happens to some of the characters.</td>
<td>.539</td>
</tr>
</tbody>
</table>

Table III  Factorial structure of UPEQ items
Correlates of UPEQ’s Autonomy, Competence and Relatedness were examined using Kendall’s tau, and it was shown that self-reported hours spent in the game correlated with UPEQ’s measures of Competence and Relatedness independently. All subscales of UPEQ significantly correlated with the likelihood of the players recommending a game and players purchasing the game as well as players Rating the game in a 5-point scale.

The second study also explored correlates of subscales of UPEQ and physical presence with other measures of game evaluation and behavioral tracking. It was found that UPEQ’s Autonomy, Competence and Relatedness are significant correlates of Physical Presence, Self-reported rating of the game as well as in-game measures such as number of days that the player has played the game, money spent on the game and group playtime percent. Physical presence also showed significant correlations with the Rating of the game, number of daysplayed and money spent on the game.

2. Cohort Specifications

In this independent study on April 2019, a month after the release of Tom Clancy’s The Division, The UPEQ questionnaire was distributed among players via an online survey, and it showed reliability and all its main and subfactors significantly correlated with the Rating of the game. Although rating itself showed a negative correlation with ‘playtime hrs’ and ‘daysplayed’, it is assumed that higher rating as well as higher scores on factors and subfactors of UPEQ indicate a higher satisfaction with the game in general.

From over 200 pairs of Kruskal-Wallis H tests, the ones that reject the null hypothesis, generally show a trend for Baby Boomers to score higher in measures of Need satisfaction and specifically Agency (Meaningful choices & impact on the game) and Growth (Progression & increasing challenges).

The variety of the Behavioral measures were more prominent, such as ‘daysplayed’, ‘playtime hrs’, ‘prelvl30 playtime’, rejecting the null hypothesis for all six pairwise comparisons. Means and standard deviations for these measures are presented in Table IV.
Table IV. Behavioral measures, means and standard deviation (SD)

<table>
<thead>
<tr>
<th></th>
<th>Boomers</th>
<th>Gen X</th>
<th>Millennials</th>
<th>Gen Z</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>daysplayed</td>
<td>102.23</td>
<td>81.13</td>
<td>172.38</td>
<td>107.06</td>
<td>124.52</td>
</tr>
<tr>
<td>playtime_hrs</td>
<td>298.60</td>
<td>229.92</td>
<td>196.40</td>
<td>212.53</td>
<td>209.27</td>
</tr>
<tr>
<td>prelvl30_playtime_hrs</td>
<td>416.61</td>
<td>371.15</td>
<td>180.13</td>
<td>212.53</td>
<td>209.27</td>
</tr>
<tr>
<td>days_in_group</td>
<td>62.26</td>
<td>52.60</td>
<td>35.96</td>
<td>34.19</td>
<td>35.29</td>
</tr>
<tr>
<td>playtimegroup_hrs</td>
<td>139.18</td>
<td>114.99</td>
<td>79.92</td>
<td>51.72</td>
<td>81.02</td>
</tr>
<tr>
<td>prelvl30_playtimegroup_hr</td>
<td>146.14</td>
<td>120.94</td>
<td>86.85</td>
<td>429.74</td>
<td>147.11</td>
</tr>
<tr>
<td>max_gearscore</td>
<td>460.62</td>
<td>455.19</td>
<td>427.30</td>
<td>426.15</td>
<td>150.66</td>
</tr>
<tr>
<td>max_level</td>
<td>29.54</td>
<td>29.42</td>
<td>29.06</td>
<td>27.93</td>
<td>5.43</td>
</tr>
</tbody>
</table>

Note: Means of the top three variables in this table are distinct at all 6 degrees of comparison, while for the next four rows, Baby Boomers and Generation X means are not significantly different. The ‘max level’ mean for Baby Boomers is not significantly different from both Generation X and Millennials.

Based on the self-reported measures, Rating of the game only differed between Baby Boomers who rated higher than Millennials. Players self-reported the number of hours per week which were spent on playing video games, and the sole difference identified for this measure was for Generation Z (Median = 21-30 hours per week) compared to the other generations (Median = 11-20 hours per week).

Additionally, analyses of self-reporting based on the Money spent on video games per month resulted in a decreasing trend of reporting with age, displaying the lowest amount for Baby Boomers (Median = 11-30 $) in comparison with Generation Z (Median = 31-60 $).

Kendall’s tau for non-parametric correlations with age and generation index (categorical number associated with each generation) were particularly high. Among those, strongest significant correlation was observed between generation index and ‘prelvl30 playtime hrs’, the number of days played (‘daysplayed’) and the amount of time (hrs) played (‘playtime hrs’), as well as ‘days in group’. Age showed similar power in correlations.

We employed a series of machine learning techniques to further examine the significance of behavioral and self-reported measures in prediction of Age and generational index. As this testing conditions include 64 classes of age in 4 classes of generations, a random attempt at predicting age and generation will have baseline accuracies of 1.56% and 25% respectively.
In this experiment, a multilayer perceptron, a support vector machine regression module, a linear model with a forward stepwise selection and a multimodal ensemble found ‘playtime hrs’, Competence, Agency, ‘prelvl30 playtime hrs’ and ‘daysplayed’ were the most important factors in predicting age and Generational index.

3. Modelling Motivation

This section explores the performance of models trained on different features of the dataset when predicting a motivation dimension of UPEQ or Yee’s Player Motivation scale (PM). Figure 3 shows the test accuracy of a 10-fold cross-validation process, with inputs chosen among the features of the dataset listed in Table VII. Results clearly show that non-linear RBF models are better predictors of all motivation dimensions.
Observing the predictive capacity of different input features, it is clear that combining all features leads to higher accuracies. When comparing the accuracy using All Features versus the smaller groupings of Table V. (excluding All Gameplay), accuracy is significantly higher for All Features in 64 out of 72 cases.
In terms of the actual motivation dimensions, the most robust models reach a very high accuracy. The best UPEQ models reach 81% average (94% maximum) accuracy, while PM models have 79% average (93% maximum) accuracy. The two frameworks show comparable levels of performance on their respective tasks and the presented study showcase the external and ecological validity of both tools independently.

<table>
<thead>
<tr>
<th>Category</th>
<th>Metrics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Playtime</strong></td>
<td>Days Played, Days in Groups, Days in the Dark Zone, Sessions, Playtime, Group Playtime, Dark Zone Playtime, Playtime as Rogue</td>
</tr>
<tr>
<td><strong>Completion</strong></td>
<td>Non-Daily Missions, Daily Missions, Side Missions, Days with Incursions, Incursions</td>
</tr>
<tr>
<td><strong>Progression</strong></td>
<td>Gear-Score, Dark Zone Rank, Level, Early Level 30 (i.e. maximum level was reached earlier than the average), Reached Level 30</td>
</tr>
<tr>
<td><strong>Early Play</strong></td>
<td>Level Below 30, Early Playtime, Early Group Playtime, Early Dark Zone Playtime, Early Playtime as Rogue</td>
</tr>
<tr>
<td><strong>DLC</strong></td>
<td>Underground Playtime, Survival Playtime, Season-Pass</td>
</tr>
<tr>
<td><strong>Play Styles</strong></td>
<td><em>Adventurer</em> (focusing on solo missions), <em>Elite</em> (having high gearscore), <em>PvE All-Rounder</em> (engaging in cooperative activities), and <em>Social Dark Zone Player</em> (engaging in PvP)</td>
</tr>
<tr>
<td><strong>All Gameplay</strong></td>
<td>All metrics except playstyles, 26 in total.</td>
</tr>
<tr>
<td><strong>All Features</strong></td>
<td>All metrics including playstyles, 30 in total.</td>
</tr>
</tbody>
</table>
The best models based on these experiments of are good predictors on the specific motivation dimension they are trained on. Now, we wish to explore the degree to which our models are able to generalize and predict other dimensions of motivation within the same survey or even dimensions captured in other motivation surveys. Such an experiment would uncover the generality level of individual measures and the overlap between the two approaches (UPEQ and PM). We expect higher accuracy when predicting features from another framework which strongly correlate with a given training feature (see Fig. 4), and a lower accuracy when predicting other features within the same framework, as these constructs are meant to measure distinct factors of motivation. Fig. 5 shows the accuracy of models trained on UPEQ and PM dimensions and tested on other dimensions of the same framework or even on dimensions of the other framework. Reported accuracy is the average test accuracy via a 10-fold cross-validation, where the outputs in the test fold are swapped with a different motivation dimension. It is evident that testing models
trained on UPEQ dimensions and tested on other UPEQ dimensions yields predictions significantly above the baseline in more cases (8 out of 12) than when models trained on PM dimensions are tested on other PM dimensions (27 out of 56). Looking at how models trained on PM dimensions predict UPEQ dimensions, we note that each model can predict one PM dimension highly: autonomy for the fantasy model, competence for the growth model, relatedness —unsurprisingly—for the social model, and presence for the fantasy model. The same relationships exist on models trained on UPEQ dimensions, but the accuracies are slightly lower (on average 56.8% when predicting PM from UPEQ models versus 57.4% when predicting UPEQ from PM models). The highest overlap in the same framework (based on prediction of one dimension by a model trained on another dimension) is with presence and autonomy in the UPEQ model, while competence also has lower but significantly above the baseline accuracies when predicting all other UPEQ dimensions (and all PM dimensions).
The data suggests that while both frameworks of motivation offer some unique strengths, there is enough overlap between the two that application of one framework over the other can be prioritized depending on project needs, while still retaining some indication of the other's dimensions. The overall good results of individual modelling—and the observed overlap, predicted by both statistical analysis and predictive modelling—confirms the reliability and validity of both measuring tools.

---

<table>
<thead>
<tr>
<th>Training Dimensions</th>
<th>Autonomy</th>
<th>62%</th>
<th>61%</th>
<th>50%</th>
<th>64%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Competence</td>
<td>63%</td>
<td>82%</td>
<td>61%</td>
<td>60%</td>
</tr>
<tr>
<td></td>
<td>Relatedness</td>
<td>51%</td>
<td>59%</td>
<td>78%</td>
<td>48%</td>
</tr>
<tr>
<td></td>
<td>Presence</td>
<td>70%</td>
<td>59%</td>
<td>48%</td>
<td>81%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Testing Dimensions</th>
<th>Autonomy</th>
<th>53%</th>
<th>68%</th>
<th>50%</th>
<th>50%</th>
<th>61%</th>
<th>58%</th>
<th>53%</th>
<th>61%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Competence</td>
<td>59%</td>
<td>58%</td>
<td>57%</td>
<td>63%</td>
<td>62%</td>
<td>69%</td>
<td>53%</td>
<td>57%</td>
</tr>
<tr>
<td></td>
<td>Relatedness</td>
<td>69%</td>
<td>44%</td>
<td>52%</td>
<td>57%</td>
<td>51%</td>
<td>60%</td>
<td>49%</td>
<td>48%</td>
</tr>
<tr>
<td></td>
<td>Presence</td>
<td>51%</td>
<td>73%</td>
<td>56%</td>
<td>54%</td>
<td>63%</td>
<td>63%</td>
<td>54%</td>
<td>55%</td>
</tr>
</tbody>
</table>

### Fig. 5
Average cross-dimension prediction accuracy of the best models. The y-axis shows the dimension the model was trained on. The x-axis shows the dimension the model was tested on. Coloured cells indicate an average performance significantly above the 50% baseline at \( p < 0.05 \). Grey cells show lack of significance.
4. Playstyle Classification

In the previous section we used factor analysis to cluster player activities in the game, but not all measures gameplay behavior are numerical. In most role-playing games such as TCTD2, categories of player choices whether they are dialogue options, gear pieces or title of missions and areas most visited, are not numerically different. Therefore, as an initial approach for identifying character builds characterized by gear pieces equipped by a player, we tried using the more classical centroid-based and density-based clustering techniques on the character configurations. For k-means clustering, all raw character configurations are encoded as numeric feature vectors. A normalization step for numerical features and one-hot-encoded (OhE) for categorical values, dimension reduction using (tSVD), the silhouette coefficient [226] and the sum of squared displacements to determine the optimal number of clusters and finally, (tSNE)[218] for visualization were used.

Due to the necessity to precompute a custom distance matrix, we could only afford to perform density-based clustering on a small subset of the original data. Numerical features are first normalized, Gower distances are calculated for categorical values, grid search was done to maximize the silhouette coefficient and then visualized using tSNE embedding over a range of perplexity values.

As we reviewed in the last chapter, inherent limitations of the above methods, led us to use a categorical algorithm (LDA) to classify our build data.

The core idea behind the Build as Test (BaT) approach is that each character configuration can be seen as a text document, composed of a list of keywords describing how the character is customized by the player. Here, inventory type elements, including both weapons and gears, come together with skill elements as well as specializations, to be merged into a single list. Numeric features are transformed into categorical bins. The level of description for the keywords should be selected carefully, as it needs a good knowledge level of the core game mechanics. Choices in granularity level of the character configuration descriptors are driven by the concept of a build and are specific to the game. In this study, once base game mechanics have been well assimilated, the exercise of choosing the relevant level of granularity was intuitive and yielded comprehensive results.

Concretely, the list of keywords defining each character configuration is encoded using a classical text frequency data structure, and stored in a sparse matrix format, which makes it efficient for handling very large volumes of data as it is
common in Role Playing Games (RPGs) that equipping the character with several pieces of a same ‘set’ provides powerful bonuses as a result of their combination, which are clearly communicated to the player. Those granted bonuses participate in the build construction since they provide advantages to specific attributes, related to different playstyles.

We outlined the method we used for categorical clustering of the player of Tom Clancy's the Division 2, and the clusters found for 9 topics are listed in Table X. For each of them, the characterizing keywords are listed in decreasing order of relative importance. To keep verbosity low, we limited the information shown in this table to the top 5 keywords, but the game experts used up to top 15 keywords to interpret the clusters. They attributed names to each cluster based on this input and their knowledge of the game. This task was deemed to be surprisingly easy, compared to previous attempts to make sense of high level in-game data aggregations.

Next, we visualized how the discovered builds relate to one another in terms of similarity and size. For this purpose, we applied a Principal Component Analysis to the distribution of keywords for each build and plotted the projection onto the first two principal components (see Fig. 6). Although it is only an approximation, it provides a useful overview of the relative sizes and positions of clusters in the build space. This visualization also tends to confirm the results’ relevance, since neighbor builds 0,4,7,8 share conceptual elements of playstyle, while distant builds (e.g., 2) are indeed distinct in terms of use of gear, weapons, and other attributes.
In order to better understand the discovered builds and study behavioral differences between them, we computed in a post-modelling phase various statistics and playtime metrics.

Fig. 6  2D PCA projection of the nine-topic builds, labelled by build ID
<table>
<thead>
<tr>
<th>Keywords</th>
<th>Build Name</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Casual Joe</td>
</tr>
<tr>
<td></td>
<td>Underivable</td>
</tr>
<tr>
<td></td>
<td>Standard Rail</td>
</tr>
<tr>
<td></td>
<td>Gunshooter</td>
</tr>
<tr>
<td></td>
<td>Easy Exotic</td>
</tr>
<tr>
<td></td>
<td>Clinch</td>
</tr>
<tr>
<td></td>
<td>Sniper</td>
</tr>
<tr>
<td></td>
<td>Sniper-Wide</td>
</tr>
</tbody>
</table>

**Table VI.**

Clusters Found For 9 Topics - Ordered Keywords Lists
supplemented by the fraction of activity spent playing main mission, dark zone (DZ), and other in-game activities. The last three sum to 100%

Table VII:

<table>
<thead>
<tr>
<th>Activity (%)</th>
<th>DZ</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Playtime</td>
<td>4%</td>
<td>1%</td>
</tr>
<tr>
<td>Missions</td>
<td>3%</td>
<td>3%</td>
</tr>
<tr>
<td>Build Attributes</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Cluster Summary Statistics
Table VI. shows the number of character configurations allocated to each discovered build (i.e., the Size), as well as the fraction of the dataset it represents. From this table we can, for example, see that build 8 is the most popular one among players, whereas builds 4 and 5 are rarely utilized.

Table VII. shows the main, game-oriented statistics of each discovered build. The Build attributes columns of this table show the means and standard errors of the 6 considered attributes. The Playtime split columns show how players of each build spend their gaming time: first, in terms of the average fraction of time spent in a group vs. solo; And second, in terms of the fraction of time spent in main missions, dark zone (DZ), and other activities (exploration, side-missions etc.). The last three columns sum up to 100%. Group playtime is usually utilized to indicate the tendency for social play.

We could argue that each build’s playstyle comes with a different level of interdependence with other builds, consequently making it suitable for social play. The statistics obtained for the builds are in accordance with their function, gameplay style, and intended design.

To assess the relevancy of the proposed method, we tested whether BaT was able to recover at least a subset of builds expected to be present from external sources and game experts. Independently from BaT experiments, we performed a simple rules-based build assignment on the same dataset and compared the obtained results.

To do so, we manually defined 4 builds, well-known from online forums and internal playtests. Each of them could be assigned to a character configuration if and only if said data point satisfied all the predefined build’s conditions. Therefore, each data point could be allocated to at most one predefined build. When it was not allocated to any of them, the character configuration was labeled as “Undefined” See in Table VIII.
We also quantified the overall match between the rules-based clustering and BaT clustering results, without explicitly defining build pairs. For this we used the adjusted random index (ARI) [227] and the adjusted mutual information (AMI) [228] indicators. ARI and AMI can be used to compare the degree of match between two clustering results even when the number of clusters produced by the two methods are not the same, which is the case in this study. The obtained values of 0.21 and 0.16 for ARI and AMI respectively, indicate a match that is better than random, but not perfect (a score of 1.0 would indicate a perfect match). This suggests that the BaT approach is capable of recovering known builds to a good extent.

### 5. Social contagion and Influence

In this section, we show the differences between the influencers that we identified using graph theory and Social Network Analysis, as described in the methods section, and two other samples of players in general statistics from the game. We then look at the effects of joining the influencers’ social circle in short and long-term, including conversion to influencers and overall retention.

Table IX. shows an overview of three groups: influencers, powers users, and the total population. We compare them based on the lifetime of players to the time of data collection to show that our defined group of influencers are different from those two groups. On average, it turns out that the powers users are indeed the powers users we would expect with more sessions played, more daily playtime, but especially far more playtime, kills, skill kills (i.e., killing enemies with particular abilities), and items extracted compared to the influencers and the total population. Interestingly, both influencers and power users spent about two-thirds of their time in group play, whereas this is the reverse for the general
Another interesting observation is that for performative measures (e.g., kills, skill kills, and items extracted) the influencers are similar to the general population.

As for group play, our results show that power users are only marginally more engaged per hour than the average player and that influencers take far more initiative in creating groups.

Table IX. Comparison of the three populations: Influencers, power users, and total population.

<table>
<thead>
<tr>
<th></th>
<th>Influencers</th>
<th>Power Users</th>
<th>Total Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total # players</td>
<td>49</td>
<td>2,102</td>
<td>14,716,507</td>
</tr>
<tr>
<td># Sessions, M (SD)</td>
<td>178.27 (313.94)</td>
<td>213.71 (258.13)</td>
<td>44.54 (442.50)</td>
</tr>
<tr>
<td># Kills, M(SD)</td>
<td>7,353 (5,286)</td>
<td>26,937 (55,374)</td>
<td>6,849 (10,738)</td>
</tr>
<tr>
<td># Skill kills, M(SD)</td>
<td>1,172 (1,719)</td>
<td>5,385 (4,261)</td>
<td>1,041 (3,247)</td>
</tr>
<tr>
<td># Items extracted, M(SD)</td>
<td>437 (328)</td>
<td>1,561 (3,566)</td>
<td>513 (1,895)</td>
</tr>
<tr>
<td># Friends, M(SD)</td>
<td>208.07 (104.59)</td>
<td>26.51 (32.42)</td>
<td>8.60 (36.35)</td>
</tr>
<tr>
<td># Groups created, M(SD)</td>
<td>87.94 (90.82)</td>
<td>203.03 (301.12)</td>
<td>22.36 (138.46)</td>
</tr>
<tr>
<td># Groups joined, M(SD)</td>
<td>47.19 (148.42)</td>
<td>173.40 (137.21)</td>
<td>20.38 (52.17)</td>
</tr>
<tr>
<td># Players interacted with in group play, M (SD)</td>
<td>341.89 (229.47)</td>
<td>27.05 (274.69)</td>
<td>10.72 (306.43)</td>
</tr>
<tr>
<td>Total playtime in hrs, M (SD)</td>
<td>119.63 (98.51)</td>
<td>454 (172.37)</td>
<td>67 (217.38)</td>
</tr>
<tr>
<td>Daily playtime in hrs, M (SD)</td>
<td>2.56 (1.64)</td>
<td>3.39 (1.96)</td>
<td>1.56 (1.47)</td>
</tr>
<tr>
<td>Time spent in group–solo play</td>
<td>61%–39%</td>
<td>67%–33%</td>
<td>38%–62%</td>
</tr>
<tr>
<td>Time spent in coop–competitive play</td>
<td>53%–47%</td>
<td>46%–54%</td>
<td>49%–51%</td>
</tr>
</tbody>
</table>

In terms of group play, influencers and power user have difference in quality and quantity of their social interactions as well. Power users tend to play almost exclusively with their friends. At the same time, each influencer plays with a much larger number of other players compared to the already large number of their friends, indicating that influencers play in groups with considerably more players than just their friends. Therefore, while power users spend on average about equal amount of time in group play, they are more likely to play with friends rather than strangers.

To further show the effect influencers have on other players, we isolated all players that played with influencers and did the same for Power users and a random sample of players as well, then we compared the data from the two weeks before joining the community of the target group and then compared daily playtime and social play ratio of the two weeks before and two weeks after joining the communities of the influencers, power users, and random players.

Results shown in Table X. show a clear change and impact on the behavior of players that join the community of the influencers: the daily playtime increases considerably, from a number very close to the general population average to a number very close to the influencers themselves; the amount of time spent in
groups increases from the total population average to almost the same amount of the influencers. For playing with power users, in contrast, both the daily playtime and social play do not change drastically. The numbers are also similar to the powers users themselves, suggesting that not only do power users play together with their (limited) group of friends (see Table IX.), but they are also likely to play together with other power users. Engaging with random players does not change behavior and as expected here, the statistics are similar compared to the total population.

Table X. Changes in influencers’, Power Users’, and Random Players’ communities two weeks before and two weeks after engaging with those groups.

<table>
<thead>
<tr>
<th></th>
<th>Influencers</th>
<th>Power Users</th>
<th>Random Players</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daily playtime in hrs, M (SD)</td>
<td>1.78 (1.41)</td>
<td>2.21 (1.36)</td>
<td>3.39 (2.43)</td>
</tr>
<tr>
<td>Social play ratio</td>
<td>41%–59%</td>
<td>59%–41%</td>
<td>60%–40%</td>
</tr>
</tbody>
</table>

As our goal is to foster long lasting influence and study that effect, we first considered if players are still actively playing TCTD after a year and used the same variables of playtime and social play as measures because these can indirectly provide insights about retention: more engaged players and players engaged in the multiplayer aspects of a game tend to stick around longer. To calculate retention here, we looked at which of the players who joined the communities of the influencers, power users, and random players were still active after one year. Table XI. shows the retention results. After 1 year, 23% of the influencers’ community is still active, whereas this is 29% of the power users’ community and only 5% of the random players’ community.

Table XI. Players interacting with influencers, power users, and random players that continue to be active and have become influencers themselves after one year.

<table>
<thead>
<tr>
<th></th>
<th>Initial population</th>
<th>Active after 1 yr</th>
<th>Retention</th>
<th>Influencer conversion</th>
<th>Conversion rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Influencers’ community</td>
<td>16,742</td>
<td>3,901</td>
<td>23%</td>
<td>1,002</td>
<td>25.7%</td>
</tr>
<tr>
<td>Power users’ community</td>
<td>1,346</td>
<td>390</td>
<td>29%</td>
<td>22</td>
<td>6%</td>
</tr>
<tr>
<td>Random players’ community</td>
<td>560</td>
<td>28</td>
<td>5%</td>
<td>2</td>
<td>1%</td>
</tr>
</tbody>
</table>

The second consideration is more ambitious. We considered which of the players from these communities may have become an influencer themselves. As we have demonstrated, influencers have an impact on other players and so if these players are converted into influencers, they, on their turn, can influence others—and thereby keep the community alive, even if certain influencers decide to leave the game. For this analysis, we first applied the same method for identifying the
original 49 influencers except a year later (see Chapter IV Section 5). Then, we considered which of the newly identified influencers mapped onto the players from the initial population that the original influencers, power users, and random players engaged with.

Table XI. shows the conversion results, which are remarkable on influencers’ ability to convert their adjacent community to become influencers themselves by a rate of 26%.

6. Detecting disruptive behavior

In this section we present to which degree we are able to detect sanctioned players as a function of their playing behavior, whether we are able to model the severity of toxicity, and finally model the type of toxic action performed.

In our first attempt we view the task of modeling toxicity as a binary classification problem, and we used random forests to distinguish between sanctioned and unsanctioned players in our dataset. We first investigated the predictive capacity of the 13 selected features which hold substantial predictive capacity as the toxicity models reach accuracy of 82%, on average with the 95% confidence interval. By adding more data, to improve our testing accuracy, and we reach high accuracies of 90.7%, on average.

Given the very promising results we obtained in the binary (sanctioned vs. unsanctioned) experiments, our next step is to dive further into toxicity prediction and construct models that not only predict whether a player will be labelled as toxic but also the severity level of the toxic behavior. We test to which degree we can predict both toxic behavior and its severity by employing RF models that map between the game-related behavioral features and 3 classes of sanctioned, warned, and banned (the accuracy baseline is 33.3%). Our RF models reach accuracy of 85% on average. The model, however, seems to misclassify banned players as warned almost as often as it correctly classifies them as banned.

So far, our findings suggest that both toxic behavior and its severity can be predicted with sufficiently high accuracy through a small set of in-game behavioral features. We also examined the degree to which our RF models can predict the type of toxic behavior. Such a predictor is critical to community managers as it provides a more nuanced and detailed information for any of their data-informed decisions about a specific player and the game.
The two types of toxic actions considered are offensive behavior and unfair advantage. The former toxic action is related to any type of offensive behavior observed during gameplay whereas the latter is related to behaviors that lead to unfair play. Note that these two toxic actions are not inclusive of all possible action types for which a player could be sanctioned in For Honor; however, they cover the vast majority of toxic behavior in the game (approx. 99% of sanctioned players which makes them representative) and we assume they would be easier to predict based on in-game behavioral features. As we report, constructed RF models predict the 3 classes (unsanctioned, offensive behavior, and unfair advantage) with an accuracy of 87.6%. These results suggest that the type of toxic action can be predicted with a very high degree of accuracy.
VI CONTRIBUTIONS

This section summarizes the research contributions of the publications that this thesis compiles and unifies. First, contributions and publications are linked to different RQs. Then, each RQ is presented and discussed from through the contribution of relevant included papers.

PAPER (III) represents Research method development efforts also covered in (I), (IV) and (V).

PAPER (IV) will represent (I), (II) and (VI) field studies in topics of player modeling and motivation.

Finally, PAPER (VI) represents multimodality and data triangulation and problem-solving methods also used in (I) and (II).

Thus, the contribution type of this work is primarily problem solving and field study with occasional Research method development. Research methods were proposed to alter the process of data collection and treatment of both behavioral and evaluative self-reported measures to replace reductionist approaches to game telemetry and experience evaluation (Methods guiding the problem-solving cycle [193]). The Ubisoft Perceived Experience Questionnaire (UPEQ) [32] was developed for example, to address the problem of often singular and unfounded game evaluation tools, aiming to emphasize the multifaceted essence of modern gaming. This approach will help game developers evaluate their products during and after development and compare design intentions with player satisfaction. Ultimately, developers will have a framework to adjust gameplay features to specific needs of their defined demographic or discourage players to participate in toxic behavior. In that regard, it is necessary to tailor game titles based on players’ perceived experience. In other cases, especially when it comes to affect and motivation, Field study contributions were prompted to enrich self-reported features of the player feedback dimension by adapting and incorporating existing theories to the gaming environment (contributions to the specific concept found in other literature for the area of concern [193]). This type of contribution enables researchers to argue that expressing and eliciting emotions through video games is not only limited to the players experience of positively valanced emotions [229], but also refers to the conditions that a game designer would create for a certain -negative or neutral- emotion to arise. For example, if the designer has provided the conditions for the players to feel sad about a particular event in the game, successful execution of the game would
include eliciting a negatively valanced emotion that could potentially be desirable for the player seeking a depth of experience. In a couple of instances, the context of the game or the type of collected data led to development and generalization of -combination of- methods suitable for similar disciplines and applications (Methods guiding research cycle [193]). In PAPER III clustering of high dimensions of categorical data is demonstrated in a step-by-step method that encompasses the full process and application of the proposed method, from data collection and preparation to performance evaluation in context and against external data (e.g. socially emergent character builds from game forums are compared with text-based clustering of context-heavy categories), which may be applied to similar types of data within games (Role playing games and their choices) or outside the area of concern. Similarly, in PAPER I, the need for facilitation of detection and reprehensive action against toxic and disruptive behavior, led the method development for feature selections and prediction of severity and type of toxic behavior based only on player data.

| Table XII: Relationship Between the Different Research Questions and the Publications |
|-----------------------------|-----------------------------|
| **RQ** | **Papers** |
| RQ1 | I, II, V, VI |
| RQ2 | I, II, III, IV |
| RQ3 | II, III, IV, VI |
| RQ4 | III, IV, V |

**RQ1) What are the typical approaches of studying one’s placement in consumer society and its relationship to media consumption?**

Researchers have used a range of methods and approaches for addressing what happens in a player’s personal life and its effects on media and specifically games’ consumption. By placement in consumer society, we mean those attributes of players as consumers which show themselves in patterns of consumption (e.g. total playtime, number of days played, number of sessions per days played, social vs solo play, spending habits, playstyles and in-game progression) as well as demographic and psychological characteristics (emotions, mood, motivation, age, gender, preferences, etc.) that we explore and report their effects on game
consumption. In our contributions we aimed at mixed method and innovative approaches to make sure that players’ self-reported measures (typically surveys) were studied in their relevant context and are cross-referenced with player behavior.

As discussed in **PAPER VI**, we take on the approach of the typical game surveys and show that they mostly rely on small samples and linear intercorrelations between self-reported measures. Our contribution is to present a review of most frequently used game evaluation questionnaires in comparison to our proposed assessment tool based on Self-determination theory, that not only has the capacity to be applied to a game as a whole but also to more granular activities within the game. More importantly, independent subcategories of perceived need satisfaction, show linear regression with the number of days played, group playtime percent among other factors of player behavior in-game.

**PAPER V** investigates players influence on one another, and shows that the social contagion phenomenon, well established in human psychology and behavior, is expandable to the social network within the game and that the effects of it are observable in short- and longer periods of time. Although we employ typical approaches such as t-test for measuring these effects, use of graph theory to explore player consumption patterns in in-game social media use and extract the list of socially influential players and monitoring their effects on other player over time, is our contribution in pushing the limits of convention in study of gamers.

**PAPER II** is an investigation for generational cohorts of players and how they differ both from a behavioral and a motivational standpoint. Studies of difference in approach to and perception of a video game in an intergenerational context are very rare and often limited to qualitative assessment or participation of older adults in the design process. In this study we explore linear and non-linear relationships between generational index and markers of in-game behavior and self-reported measures such as motivation. We discuss implication of playtime behavior and motivation at length, providing contributions for researchers and game designers.

Reviewing how researchers investigated disruptive and toxic behavior in games, **PAPER I** offers an approach to detect toxic players by modeling their behavior in-game using a Random Forest algorithm and contributes to the existing literature by providing evidence that toxic behavior is not isolated to the behavior itself and other determinants such as perception of social placement and competence are involved in occurrence of disruptive acts, which can help game
designers and community managers in detecting and dealing with disruptive behavior.

**RQ2) How is the study of -predictive- behavioral models achieved?**

Prediction and modeling of player behavior and modeling typically happens surrounding concepts such as retention and churn. In these studies, k-means clustering, and collaborative filtering are popular and occasionally time series analysis is used for prediction. Our most notable contributions, however, concern nuances of behavior and player attributes that are specifically targeted towards researching players and providing insights for game designers.

**PAPER IV** presents one of the few examples of using gameplay behavior as input of a prediction model for player motivation, while most studies of game motivation are limited to linear relationships between factors of a questionnaire. We show that using Preference Learning for pairwise comparison of Likert scale data type improves the accuracy of the model.

The main contribution of **PAPER III** is the application of categorical clustering (Latent Dirichlet Allocation, LDA) to the text data of gears and equipment on a player’s character build as a foundation for prediction of playstyle and for recommender systems. This paper shows that emergent clusters are significantly different in their playstyle and approach to gameplay. We could also extract character builds and corresponding behavior that was known to game designers. The principles of the model are generalizable to any text-based game data such as loot, player inventory and narrative choices.

Similarly, **PAPER II** ventures into the uncharted territory of predicting player attributes, namely generational index, based on player behavior and by comparing multiple methods of establishing that relationship with the gameplay data. We employed a step forward linear model as well as neural network, support vector machine and an ensemble of all models to explore the ability to predict generational index and performance of each method.

Predictive behavioral modeling takes a different turn in **PAPER I**, targeting detection of disruptive behavior in two different classes and three levels of severity. We indicated that narrowing down the specifics of the disruptive behavior (type, severity) will increase model accuracy. We also showed how a hybrid method of feature selection is beneficial in selecting a few variables that cover varieties of toxic behavior.
RQ3) What would be the interrelation and appropriate classification of historical, political, and developmental indices of a gamer and their respective behavior?

Throughout this body of work, the goal is to challenge the norms of inferring relationships between players’ self-reported measures and their in-game behavior. In other instances, providing context and innovative methods to classify and parse through behavioral data, gave us unique insights into proficiency of the methods chosen as well as how different types of data could be treated for a specific purpose (e.g., using graph theory for the in-game social network).

In PAPER VI, we outline a framework for survey design and validation that relies on connection to game behavior. In addition, theoretical challenges, and limits of application of the other popular game evaluation and preference tools were reviewed. We showed that factors of player psychological need satisfaction in our proposed questionnaire, each independently correlate with time and money spent on the game as well as group playtime.

As a next step, and after showing the validity and linear behavioral correlates of the questionnaire designed and validated in PAPER VI, we propose a non-linear approach to classify and predict player motivation based on granular categories of gameplay in PAPER IV. Treatment and analysis of self-reported data via Preference Learning (transforming Likert scale data to ordinal values) as well as a method for classification of aggregates of behavioral metrics and using these categories to uncover relationships between player motivation and aggregate features of in-game behavior, is the contribution of this work.

In PAPER III, we propose LDA clustering method for data types that are inherently categorical and compare that approach to numerical clustering methods such as K-means or DBScan. Our contribution to this research question is confirming discovered categories by calculating their similarity and validity. Compared to expert players and community recommendations. We argue that the framework of this study in discovering and evaluating classification of behavior is applicable to all manner of categorical game data.

To classify and compare historical, psychological, and habitual attributes of different generations in an online game, PAPER II proposes a rank-based model for comparing independent sample means, Kendall’s tau for non-parametric correlations and a series of machine learning algorithms to indicate predictor importance of behavioral data for detecting players’ generational index. This
paper reports significant differences in performance and perception of different elements of the game among generational cohorts.

**RQ4) What additional information could a longitudinal-modelling of player-behavior based on their psychosocial profile provide us?**

The extensive focus of this body of work on implications of each study for game designers and game user researchers goes beyond establishing relationships and describing case-specific models and insists on generalizable approaches and long-term effects of the interventions and findings.

As such, **PAPER V**, while identifying a group of social influencers via multiple measures of graph theory, shows the effects of joining that group of players in increased playtime and group activity over a short period of time (within two weeks) but it also reports this effect on increased retention after a year of joining the social circle of influential players.

Similarly, in **PAPER IV**, we focus on in-game activities of players two years into their experience with the game, so that the aggregated data could support players’ general disposition towards the game captured via UPEQ (the questionnaire designed and validated in **PAPER VI**), which was established with a high degree of accuracy. While the survey responses give us a snapshot of player responses, we argue that long-term and dynamic classification of behavior will improve robustness and accuracy of predicting player psychological characteristics such as motivation, to a great extent.

The Categorical clustering method and its validation that we introduced in **PAPER III** is an example of practical application of dynamic clustering for behavioral data that is reinforced with historical data and dynamic validation of categories. Our contribution is the principle we establish for long term monitoring of playstyles and other player choices confirmed by the community of players which provides opportunities for expansion of predictive models and recommender systems in the game.
VII CONCLUSIONS AND FUTURE WORK

In this Chapter, we discuss conclusions and future line of research regarding the various studies we described in previous chapters. First, measuring player experience via UPEQ and cohort differences in game behavior and motivation is explained, then we discuss implications of our approach to classify and predict motivations, but also playstyles and player choices. Last two sections will focus on detection and impact of social influence and disruptive behavior that may occur during gameplay.

1. Measuring Experience

Ubisoft Perceived Experience Questionnaire, UPEQ, was developed as a game assessment tool that is theoretically founded and gives game developers a sensible idea of how the virtual worlds created by them accounts for players sense of agency and expression of sociality through constant acknowledgement of player competency.

An initial study was used to evaluate internal consistency of UPEQ items and its subscales as an actionable alternative to existing player experience questionnaires. As results suggested, regardless of the game chosen, UPEQ could be a reliable and consistent model of game assessment at least compared to other self-reporting measures, and it provided more granular feedback on players’ appreciation as we discuss in the next section it also demonstrated better predictability compared to subjective measures of experience. In another attempt to reaffirm internal consistency of the model in a larger and game specific sample, we examined significant correlates of UPEQ as and their relationship with recorded in-game behavior. Our analysis confirmed that each subscale of UPEQ independently predicts measures of engagement in game and are a reliable alternative for direct rating of player experience, although it does not specify nuances of need satisfaction and games’ success in accomplishing them. In addition, it partially supports measurements such as days and money spent in the game which are usually metrics used to gauge player engagement. We were also able to exhibit successfully that need satisfaction is indeed related to feeling of being transported to the game world and being incorporated therein.

Consequently, UPEQ has been used on several game titles within Ubisoft to evaluate state of the game and compare it to competition in terms of need satisfaction. It could also be beneficial to set production goals based on UPEQ scores in different iterations of the game during development phase.
As mentioned earlier, incorporating a self-determination theory survey for evaluation of video games is not free of limitations. On theoretical side, further investigation into overlap in definition and perception of constructs such as Autonomy, Competence, Relatedness and Physical Presence would be useful not only to have a better sense of identifying hindrances in need satisfaction of the player but also to clearly guide game developers on how to resolve those issues and enhance the experience of game assimilation. Beside theoretical limitations and assumptions of the theory [51], samples in both of our studies were targeting volunteer online respondents, which risks surveying of only a particularly vocal set of video game audiences, for example most respondents represented men who are residents of specific geographical regions (Europe and North America) skewing the data toward their preferences. Future studies may target regional, age and gender diverse groups.

2. Cohort differences

One of the avenues of exploration on our journey to evaluate contributing factors to players’ quality of experience has been age diversity and its impact on play behavior. As Chapter V. Section 2, has presented the results from a study focusing on two large scale surveys deployed to players of Tom Clancy’s The Division 2 [263], an online multiplayer shooter played by millions of players of different ages, this section provides insight into the various factors relating to four different generational cohorts: Baby Boomers, Generation X, Millennials and Generation Z.

In summary, older players feel more agentic in the game, less masterful, closer to non-playable characters, and feel more present in the narrative than the younger generations. Additionally, even if older players perceive the game to provide them with more meaningful and believable options, the feedback provided by the game does not make them feel competent and effective.

The ‘Playtime’ data from this study highly correlates with, and is the most important predictive factor of, age and generation in multiple methods. We showed that duration of play and playing in groups correlates positively with age. Playtime is among the most popular measures of engagement [231] as well as player churn prediction [232]. More playtime for older adults other than more available free time for the retired individuals, could be connected to the trend of lower perception of Mastery among older generations. Older generations may be compensating for their lower perception of Mastery by spending more time in
the game, specifically when they reportedly feel more agentic and present in the game narrative. The data analysis showed that the more mature audience generally engage with the game more frequently, spend more time in the game, take more time to complete the game (playtime per level) and stay in the game for additional content, consistent with findings of Salmon and colleagues [233].

Results of the current study confirm that Baby Boomers play significantly more and share with Generation X the highest frequency in purchasing additional content, though in self-reported measures they reported the most modest amount of money spent in a month on video games compared to the other generations (Millennials reported the highest amount spent). We suggested that this trend may be a sign or a consequence of more income and available leisure time for the aging population, according to the existing literature [61, 234].

For game developers, playtime and number of days played are not necessarily due to higher engagement, but perhaps the generational index of that population. Although, future research should consider further investigation relating to higher playtimes of both the Baby Boomers and Generation X cohort and their respective potential obsession with this medium as discussed by Przybylski and colleagues [235].

Connecting measures of Need satisfaction (UPEQ), playtime behavior and self-reported gaming habits of the players of different age groups, gives this study a unique perspective into generational differences in expectation, approach and perception that varies as video game audiences mature. This leads to various suggestions for the industry, academics and policy makers.

The results from this study have shown that Agency, Mastery and feeling close to fantasy elements of the game, are the most discriminating measures of need satisfaction scores among generations. As a Previous study [233] has shown, older adults prefer games that are easier to learn, challenging and single player. Our results confirm this finding by reporting a positive correlation between age and players’ perceived agency and closeness to non-playable characters and narrative presence in the game. Another point worthy of discussion the percentage of social play is decreasing with age. An explanation might be the lack of technological familiarity [236] or reluctance to engage with online communications due to safety risks and fear of toxic behavior [237] that we discuss in Section 6 of this chapter. Additionally, higher scores of narrative presence and NPC Closeness may be a signal that the older adults prefer story-based games which are usually played individually. Future studies may explore
this emerging pattern to examine its generalizability to other games and what it implies for game development targeting older adults. While this study provides insight into generational differences of players of a single commercial video game, investigation of between-and-within subjects as well as between-and-within video games could shed more light on the nature and causality of the explored correlations.

We also observed a mismatch between perceiving low competence but showing high in-game progression for older generations which could be a result of an age-related decay in game performance or one’s perception of it, but also designing games with the younger audience as a target and therefore missing nuances of how older adults interact with this medium [60]. Future studies could examine the relationship between game performance metrics, age, and perception of competency through interviews and comparative analyses.

For game designers, the results show that there are significant differences in perceived need satisfaction among generations, suggesting that certain generations may assess the game differently (which is not reflected in the single measure of rating) or have different aptitudes for subfactors of UPEQ. Future studies may explore multiple games between and within subjects.

For researchers and policy makers, the current study has established that age and generational index is indeed a factor of importance in the study and rapport of online gaming populations and their corresponding perception and behavior, and further investigation into socio-political factors related to this phenomenon is therefore advised.
3. Modelling Motivation

The common thread throughout this body of work is the motivational aspects of play in connection to gameplay metrics but also venturing into demographic differences and specific behavior detection or prediction. This section discusses how we can apply the results explained in Chapter V, Section 3, for predictive modelling, which examined two distinct approaches for representing player motivation.

More than 400 surveys were collected from players of Tom Clancy’s The Division using the Ubisoft Perceived Experience Questionnaire and the Model of Player Motivation surveys. The survey data was processed in an ordinal fashion and subsequently modelled through preference learning. Results showed high success in predicting all motivation dimensions based solely on high-level gameplay metrics. While the study showed the external reliability and validity of the measurement tools used, it also highlighted their unique strengths and weaknesses. The first framework—the Ubisoft Perceived Experience Questionnaire [32]—was designed by aggregating survey-based motivation feedback in a theory-driven manner as presented and explored in previous sections, while the second framework—Yee’s Model of Player Motivation [58]—applied a data-driven approach, deriving motivational dimensions from the survey data itself. Results show outstanding accuracy in predicting independent motivational factors and substantial success in transferring the predictive capacity of models’ cross dimensions and frameworks. The results also highlight the reliability and robustness of both frameworks and the underlying measurement tools. The models based on UPEQ dimensions, showed a larger conceptual range, generalizing well to predict other aspects of player motivation as well. This suggests a more robust and readily available tool, whereas models based on Yee’s player Motivation (PM) [58] require more preprocessing when it comes to acquiring motivational dimensions. The lower cross-dimension accuracies in models trained on PM dimensions, however, suggest a higher internal reliability which we addressed as conceptual overlap of the SDT constructs in Section 1. The PM framework can thus be tuned to measure and model motivational dimensions specific to the given game and player population. Independently of the framework used, the ordinal labelling and preference learning methodology used in this study proved to be robust and yielded models with 80% accuracy on average.
This study provides useful insight for academics, industry researchers and game designers alike. Firstly, the study uses a large amount of real-life data, aggregating behavioral patterns from over 400 players over a period of 2 years. This provides high ecological validity to the presented results, which hopefully paves the way towards an industrial application of predictive user modelling. Secondly, the study examines two strong and popular frameworks for measuring player motivation in a holistic manner.

Future work may develop methods of capturing player motivation in a continuous manner to address the snapshot style of conventional motivation surveys. Models based on more granular input can be used to create design tools and adaptive systems, which can facilitate both the design process and the gameplay experience. In terms of output features, the bounded nature of Likert-like surveys injects a certain level of bias into the data [84, 238]. The second-order representation of preference relations, followed here, is one way to deal with this issue; however, a more reliable and valid approach would implement a measuring tool which represents motivation ordinally following the first order representation path [84].

As noted, incorporating other modalities apart from gameplay metrics often increase the accuracy of predictive models. Large amounts of multi-modal datasets opens up future research avenues to large-scale multimodal modelling of audience motivation. It is important to also note that pairwise transformations yield large amounts of generated data, this approach is excellent for dealing with small sample sizes. This method can be a valuable tool for quality assurance departments of game studios, which deal with limited amounts of experience annotations. The technical limitation of the presented study, however, is its reliance on SVMs. While these algorithms are shown to reach outstanding performance on affective computing tasks, they scale poorly to large datasets with datapoints in the thousands. Datasets of this size will require the use of algorithms, such as deep preference learning [239-240], which are better equipped to approximate the unknown function of motivation of thousands or millions of players.

Moreover, future research should focus on the extensibility of the presented methodology to other games. The robustness of the obtained models, however, already suggests that they could potentially generalize to similar games given similar gameplay features. The methodology followed also appears to be generalizable across any game as long as gameplay and motivation survey scores
are available. Future studies should focus on validating the generality of the current motivation models and building more general ones using more diverse datasets. These general models of motivation would enable the assessment of previously unseen games providing great value to both industry stakeholders and academic researchers working towards general player modelling [21].

4. Playstyle Classification

In previous sections, we have introduced an instrument for measuring player motivation and need satisfaction and we showed how different generations perceive their experience and behave differently. Then, we examined predictability of the introduced instrument based on player behavioral measure and against another popular motivational model. In this section, we discuss adding more dimensions to player behavioral aspects through categorical clustering of player choices and in-game behavior. The case study we described for this purpose was in clustering inherently categorical game data, highly dependent on the context and we demonstrated a step-by-step method that encompasses the full process and application of the proposed method, from data collection and preparation to performance evaluation in context and against external data (e.g. socially emergent builds from game forums). We showed that considering Builds as Text (BaT method) is more suitable than k-means or DBSCAN in producing comprehensive clusters. Our description of player behavioral data reflected by playtime split and character attributes showed that meaningful clustering of character builds results in representation of different playstyles and behavioral profiles that forms around their affordances.

For data scientists, detailed illustration of our method may help in replication, performance improvement and incentive to experiment with less explored methods. Determination of significant behavioral differences between clusters also implies to game designers and scholars to study emergent character configurations beyond player performance metrics and encourage varied and personalized playstyles by monitoring player choices, behaviors incorporated with them and what they imply for player evolution. Furthermore, we suggest this method being applied to other RPG games that allow gathering similar type of data, in order to better understand how players customize their characters or make other choices in the game.

There are some limitations to this work, which open several leads to improve and continue it in future studies. First, choosing alternative ways to snapshot character configurations is a challenging task as players tend to continuously modify them,
whether for pure exploration (theory-crafting) or regular improvements (better gear found, levelling up). Finding better ways to detect relative stability in character configurations, and to study builds evolution in time and per player deserves a more rigorous research work. Alternatively, monitoring the evolution of successive clustering attempts after different game updates could also lead to interesting insights concerning the game’s overall evolution.

Second, the categorical nature of the data makes the proposed method quite sensitive to the artefacts that can appear during data pre-processing. Another limitation related to data encoding is that the keywords in the learning dataset all have equal weight, which is not necessarily the best option because of complex interactions between different categories of character features.

Future studies may explore the design process and implementation of accurate player in-game performance indicators, not only based on completion speed of main missions. Correlating such indicators with player choices would provide insights about that choice’s respective strength, and that could also be followed over time and per individual player. This would, in turn, provide relevant information to game designers concerning the practical efficiency of their designed options, and potentially help them to perform early and targeted game balancing adjustments.

Finally, future studies could use categorization of player choices as well as other dynamic measures of player performance, progression and social interactions and provide a continuous picture of player attributes for cross validation with perceptions, motivational and social aspects of play.
5. Social Contagion and Influence

We narrowed our focus to social aspects of player experience and increased engagement with the game through social interactions and outlined how using conventional Social Network Analysis (SNA) techniques can identify key members that are engaged with the game Tom Clancy’s The Division (TCTD) but also with other players with an outsized impact. We indicated that these so-called influencers really influence other player’s behavior, as measured in their playtime and social play, and their impact is more than other players. We summarize our findings on social contagion, retention, and player differences before explaining the limitations and implications of this study.

Our results highlight that the identified influencers seem to have a very tangible impact on the people they play with; these other players begin playing longer and spend more time in groups. We contrasted the impact of influencers with those of power users and average users and found that interacting with these users does not yield any significant differences in playtime and only a significant but with a smaller effect size on social play for power users. Therefore, a strong social contagion effect is unique to the influencers.

Retention as a key measure for success in the game industry, and we found that players who interacted with influencers and power users stay in the game for longer times. While the retention is higher for power users, it should be kept in mind that influencers are able to retain ten times the number of players and that power users tend to engage only with similar users, so their influence is more of a reinforcing feedback loop than having an impact on the community at large.

What is most striking, however, is that players who have interacted with influencers may become influencers themselves after a year by a 26% chance. Such influence is not as noticeable with power users or random players and suggests that the social contagion effect of influencers may go as far as converting a significant portion of the players they interact with into influencers. Because we did not (quasi-)experimentally test this or, observed whether these new influencers exhibit the same kind of impact, we cannot claim causality here neither can we fully illustrate what impact this has on the community. However, these results provide further evidence of the important role that influencers play in online game communities, especially with the issue of retention in mind. In fact, as the sustained lifetime of a game depends in large measure on a healthy, lively community of players engaged with the multiplayer aspects of the game, these players seem to form the invisible social backbone of a game community.
Although descriptive metrics highlight the differences between influencers vs. power users vs. average players, it is important to note that these metrics are not sufficient to identify influencers. What defines influencers in contrast to power users is that they have a wide-reaching and solid network of friends and an active engagement with the multiplayer aspects of a game rather than an elite performance in the game. When applying one or a combination of metrics, it was impossible to achieve the same result. Therefore, SNA seems to be required to identify influencers. The approach we have taken here is to define influencers on the basis of combining six centrality measures and then inspecting the results visually for verification.

Future research is needed to further refine this approach and examine how it generalizes to other contexts and games. Furthermore, neither the SNA nor the metrics explicitly define these influencers are and how it is possible that they can convert others into influencers. Because explanations such as that influencers are inherently social, a suggestion supported by evidence that there are similarities between virtual and real-world personalities and behavior [24-25, 241-242], or are a different type of player, more into social play [58, 243], cannot fully explain the results we observe here. It may be that those converted are socially inclined people and that interacting with another socially inclined person but already socially active player (i.e., an influencer) activates how they can and want to play. Regardless, our quantitative, hypothesis-testing approach is inherently limited in generating detailed explanatory portraits of players and the dynamics we observed. Future qualitative and mixed methods work, where influencers and the players they interact with are interviewed or closely followed over a period of time, can provide further evidence on understanding who these players are and why they have such influence on others beyond being socially active as described with the metrics here.

Our presented work has several limitations. The SNA work focused on a single game. Although there are differences between TCTD and other online games, the type of game, its mechanics, and especially group play management is similar to other online multiplayer games. The most important difference in terms of group play compared to other online games is that it is limited to up to four players at a time, meaning that player ties may be closer than in other games and that there are more loosely connected communities than in other games. More importantly, influencers who initiate group play may be more influential as they are the ones who make connections across the network, whereas in games where guilds and factions play a role there are other (social and cultural) mechanisms of how
players join groups. As future work will consider this phenomenon more closely, more robust standards and definitions will be established, and the results presented here can be (dis)affirmed.

This work has important implications for HCI researchers and industry. For HCI researchers, we have established here how social contagion or influence occurs in online, in-game networks, specifically in the context of short-term, small-sized pick-up groups. While we need to be cognizant of this particular context, this work advances the field at large. Where previous SNA work on influencers in games and social media focused on network features only [244], our work highlights how influencers act differently from others and the extent of their impact on others and the community over time. Additionally, as it remains an ongoing discussion on how to identify influencers in social networks using SNA metrics [150], our work suggests a combination is necessary and that measures of activity and popularity are not good metrics. This latter is relevant for the community management for every game or social media, which generally tend to be most interested in active and more vocal users. Our data suggests that for TCTD at least, the “most central” users may be the most important for engaging and retaining other users. Therefore, community managers should tap into SNA to identify such positive social influence, to leverage them for feedback or to reach a large part of the community and promote positive participation in the social space of the game.

6. Disruptive Behavior

While in the previous section we discussed positive social influence and contagion, in the current section we expand on our results from the study of players of For Honor which showed that we can behaviorally distinguish toxic players from other players and even distinguish among toxic players in terms of the level of severity as well as the type of their toxic behavior. We obtained these results by carefully selecting a sample of players, extracting and then selecting game features based on input from designers and statistical results, and deploying machine learning algorithms to predict toxic behavior. Altogether, this sums up our method and results for detecting toxic behavior through gameplay, which is scalable and generalizable to other MOBA games.

Following the tradition, we would discuss the implications and future avenues of research regarding disruptive player behavior. First point worth mentioning is that
typically, trying to predict more granular outcomes reduces the accuracy of a prediction models. Our models instead gain accuracy when trying to predict more precise outcomes, moving from an accuracy of 82% when predicting sanctioned and unsanctioned players to 85% when predicting the severity of the sanctions and to 87.6% when predicting the type of toxic behavior, suggests that for predicting toxicity more precise outcomes can be added, such as severity and type of toxic behavior, without losing much predictability but gaining information about players. More importantly, it provides evidence that not only are toxic players distinguishable among other players, but we can also even rather accurately distinguish between toxic players based on their in-game behaviors.

Furthermore, it is important to note that selecting random forests (RFs) as our machine learning algorithm in this experiment has certain advantages with regards to our aims. Random forests do not only offer robust predictive capacity across any dataset we tried; notably, they offer a white-box, expressive method that is transparent to any community manager of the game. RFs—being a selection of decision trees—can inform any relevant stakeholder about the features involved in distinguishing between toxic vs. non-toxic players and their corresponding importance. This is one of the reasons we chose RFs over other machine learning techniques such as support vector machines (SVMs), which achieved a similar performance.

In terms of misclassification, we find that there is less chance of misclassifying banned vs. warned players. This seems to suggest that banned players are behaviorally more distinct. As their behavior is more severe, this aspect of the toxicity detection is desirable. Additionally, we find that unsanctioned players are easy to separate from banned and warned players. Although the RF models presented reach high accuracies, misclassification can still occur in a few cases; therefore, it is strongly recommended that any automated effort to detect toxicity in the player community should not be deployed independently of human verification and a final confirmation that a certain player, classified automatically as toxic, did indeed break some of the rules stated in the code of conduct. Future work will need to examine the degree to which the method we propose can generalize across dissimilar, potentially larger, and more representative datasets within this game. Once toxicity labels are available for other games, we would be able to test to which degree we can identify general patterns of toxicity across games and game genres.
As we stated in Chapter II, Section 5.4, we propose this study as a blueprint to create a tool to support community managers, not to replace them. This tool would allow the community managers to be more proactive and avoid relying on players reporting offending individuals, which, as we have seen, happens in less than half of the cases [48] and is often not used as intended [182]. Specifically, it would provide more objective red flags and potentially help catch a much larger number of toxic players than what is usually the case by relying on players’ reports. We imagine that if a detection method as described in this work would become part of the toolbox of community managers, they can deploy this to:

(1) **Identify extent and type of toxic behavior and determine mitigating actions:** being able to more objectively ascertain the extent of and the type of toxic behavior in the community, it will provide community managers the ability to consider and discuss with the game designers how to mitigate this, for example through eliciting pro-social behaviors or suggesting changes in the match-making process.

(2) **Verify player reports on toxic behavior:** if the player reports match the outcomes from the detection method, then community managers can more rapidly respond to toxic behavior and more easily assess whether the player reports are accurate. While this mixed approach is a verification, we strongly recommend that community managers provide a final confirmation before sanctioning players.

(3) **Proactively identify toxic players:** instead of waiting for players reports, community managers can now actively monitor players that are likely to display toxic behaviors, until such behaviors are displayed.

Besides the implication of complementing the toolbox of community managers in their fight against toxicity, our results from FH provide direct insights into issues that may help alleviate toxicity. Our study implies that social game modes are used and players progress through them should be evaluated to decrease player frustration, which is a major contributing factor to toxic behavior. Future research focused on the impact of specific toxic actions on player experience and game communities may facilitate such understanding.

Additionally, the excessive chat behavior among sanctioned players shows a strong will for connection and communication among sanctioned players. While abusing this feature may not be the ideal manifestation of that will, game designers can include more communication options but also incorporate
frequency caps for messages sent. Future work should investigate the overlap between behavioral and verbal actions, i.e., are players who commit behavioral toxic actions also more inclined to take verbal toxic actions? Findings from such an effort may be able to shed further light on toxic player types, as well as what set of techniques are needed to comprehensively detect toxic players (e.g., combining random forest classifier on gameplay data with NLP classifier on chat data). We note that our work does include ‘chat actions’ but for this we only looked at the messages sent per minute of playtime and thus not the content of the messages. However, as stated, it is clear that toxic players make more excessive use of chat messages compared to other players.

The connection with enthusiasm and toxicity may be another promising avenue of future research, when we administered UPEQ [9] to the population of sanctioned players, we found correlations that could be meaningful for further investigation. Specifically, Positive correlation between For Honor Rating and the sanctioned status may point towards higher change of toxic behavior among people who are more engaged in the game and consequently assign a higher rating to For Honor; whether we can attribute this relationship towards offensive behavior (higher engagement leading to higher emotional response leading to higher change of disruptive behavior) or towards seeking unfair advantage (higher engagement leading to higher need to feel competent in the game leading to use of exploits and cheats). On the other hand, Positive correlation between Interdependence and the sanctioned status also seems a promising avenue for further research. Although we could say that the more players perceive that their performance depends on the input of other players, the more likely they are to commit offensive behavior when they consider that other players perform poorly. Contrarily, players can see that the outcomes heavily depend on others’ actions and decide to seek exploits and cheats as a way to tilt the outcome in their favor.

While our approach proved to be successful, the prediction results are only as good as the toxicity labels that were used. We believe the industry is in need of a more robust understanding of toxicity in order to address it better (through player reports and/or detection methods as presented here), which starts with clearly defining it. We suggest this needs to be done more systematically, and with the input and help of the player community, especially those who suffer the most from toxicity. By leveraging HCI work concerned with amplifying the voices of underrepresented and vulnerable communities [244-245], as well as leveraging inclusive participatory design [246] and Feminist HCI [247] practices, this will not
only help amplify the voices of underrepresented communities in online game communities on how to address toxicity but also contribute to designing and keeping a more safe, healthy environment. Last but not least, we advocate for more diverse and inclusive community managers, including underrepresented groups of players among their ranks.
References


A. Canossa, J. B. Martinez, and J. Togelius, “Give me a reason to dig” Minecraft and psychology of motivation. in Proceedings of the Conference on Computational Intelligence and Games (CIG), 2013, pp. 1–8.


Ubisoft Entertainment. (2016). Tom Clancy’s The Division, [Computer software].

Ubisoft Entertainment. (2019). Tom Clancy’s The Division 2, [Computer software].


Blizzard Entertainment. (2004). World of Warcraft, [Computer software].


Valve Corp. (2013). Defense of the Ancient (Dota) 2 [Computer software].


100


Edward Keller and Jonathan Berry. 2003. The influentials: One American in ten tells the other nine how to vote, where to eat, and what to buy. Simon and Schuster.


[189] [n.d.]. DEV BLOG: TOXICITY. .com/en-us/game/rainbow-six/siege/newsupdates/4mPLa195TWiXREafM4T4oC.


H. Martínez, G. Yannakakis, and J. Hallam, "Don't classify ratings of affect; rank them!" IEEE transactions on Affective Computing, no. 1, pp. 1–1, 2014.


[228] N. X. Vinh, J. Epps, and J. Bailey, "Information theoretic measures for clusterings comparison: is a correction for chance necessary?”, in ICML ’09: Proceedings of the 26th Annual


PART II. PAPERS