Extending Game User Experience - Exploring Player Feedback and Satisfaction
The Birth of the Playsona

BjörnStraåt

Academic dissertation for the Degree of Doctor of Philosophy in Man-Machine-Interaction (MMI) at Stockholm University to be publicly defended on Friday 1 December 2017 at 13.00 in L50, NOD-huset, Borgarfjordsgatan 12.

Abstract
Video games are experience-based products and user satisfaction is key for their popularity. To design for as strong an experience as possible, game developers incorporate evaluation methods that help to discover their users’ expectations and needs. Despite such efforts, problems still occur with the game design that lower the user experience. To counter these problems, the evaluation methods should be investigated and improved.

To address this need, I have explored various design tools and user experience theories. Applying these in a game evaluation context, I have analyzed user-created game reviews and conducted longitudinal user interviews and game diary studies in connection to playing a newly released game, in other words different methods to take advantage of users’ expectations, opinions, attitudes and experiences. One result of the analysis of the obtained data is a set of “slogans” that illustrate how and why users lose interest in a game. A second result is a method for extracting user attitudes from pre-produced user reviews and how this can be used in game development. Thirdly, I introduce an alternative model, aimed at game user experience development, the Playsona. The Playsona is a lightweight tool that introduces a variant of the Persona-method, specifically for video game design.

Keywords: video game design, user experience, game user experience, playsona, aspect based sentiment analysis, focused player diaries.

Stockholm 2017
http://urn.kb.se/resolve?urn=urn:nbn:se:diva-147770

ISBN 978-91-7797-035-4
ISBN 978-91-7797-036-1
ISSN 1101-8526

Department of Computer and Systems Sciences
Stockholm University, 164 07 Kista
EXTENDING GAME USER EXPERIENCE - EXPLORING PLAYER
FEEDBACK AND SATISFACTION

Björn Strååt
Extending Game User Experience -
Exploring Player Feedback and
Satisfaction
The Birth of the Playsona

Björn Strååt
For my loved ones.
Thank You.
Contents

Abbreviations ................................................................................................................................. iii
List of Tables and Figures .............................................................................................................. v
Acknowledgement .......................................................................................................................... vi
Introduction ................................................................................................................................. 1
  Purpose and Research Questions .............................................................................................. 2
  Research Design & Method Overview ...................................................................................... 4
  Contributions to the Field of Game UX .................................................................................... 5
  Thesis Outline ............................................................................................................................ 5
Background .................................................................................................................................... 7
  HCI, Usability and UX ................................................................................................................ 7
  HCI and UX in video game studies .......................................................................................... 9
Theoretical Framework .................................................................................................................. 13
  The Character of Things .......................................................................................................... 13
  Model of User Experience ....................................................................................................... 14
  User Engagement ...................................................................................................................... 16
  The Expanded Game Experience Model ................................................................................ 16
  Online User Reviews and Metacritic ...................................................................................... 17
Scientific Approach and Method ................................................................................................. 19
  Case Study Approach .............................................................................................................. 23
  Data Collections and Units of Analysis .................................................................................. 24
  Data Analysis Methods ........................................................................................................... 25
    Content Analysis .................................................................................................................... 25
    Aspect Based Sentiment Analysis ...................................................................................... 26
    Word frequency and selection ............................................................................................. 26
Ethical Considerations .................................................................................................................. 26
  Method Description per Paper ................................................................................................. 27
    Study 1:a ............................................................................................................................... 27
    Study 1:b ............................................................................................................................... 28
    Study 1:c ............................................................................................................................... 30
    Study 2:a ............................................................................................................................... 30
    Study 2:b ............................................................................................................................... 33
Results ........................................................................................................................................37

In what ways can existing UX theories be used within the context of game user research? ........................................................................................................................................38

Study 2:a..................................................................................................................................38

Study 2:b..................................................................................................................................41

In what ways can indirect user feedback be integrated into game user research? ........................................................................................................................................42

Study 1:b & 2:b .....................................................................................................................42

In what ways can longitudinal user feedback be integrated into game user research ........................................................................................................................................44

Pre-play Impressions........................................................................................................44

Graphical Representation of Attitudes...............................................................................45

Concluding the Results Chapter .......................................................................................47

Discussion and Conclusion ..............................................................................................49

The Birth of the Playsona..................................................................................................50

Creating the Playsona.........................................................................................................50

Future Work ..........................................................................................................................53

Papers ..................................................................................................................................55

Studies included in the dissertation ..................................................................................55

Study 1.....................................................................................................................................55

Study 2.....................................................................................................................................55

Study 1.....................................................................................................................................56

Study 1:a: ‘Does Game Quality Reflect Heuristic Evaluation? - Heuristic Evaluation of Games in Different Quality Strata’ .................................................................56

Study 1:b: ‘VOX POPULI - A Case Study of User Comments on Contemporary Video Games in Relation to Video Game Heuristics’ ..............57

Study 1:c: ‘Top Versus Bottom: Game Evaluation from an Expert or Player Perspective’ ......................................................................................................................58

Study 2.....................................................................................................................................58

Study 2:a: ‘Exploring Video Game Retention Mechanics – A Longitudinal Case Study’ .................................................................................................................................58

Study 2:b: ‘Probing User Opinions in an Indirect Way: An Aspect Based Sentiment Analysis of Game Reviews’ ..............................................................................59

Other publications by Björn Strååt .......................................................................................60

References ................................................................................................................................63
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSA</td>
<td>Aspect based sentiment analysis</td>
<td>26</td>
</tr>
<tr>
<td>ACIV</td>
<td>Assassins creed IV: Black Flag</td>
<td>29</td>
</tr>
<tr>
<td>APC</td>
<td>Apparent product character</td>
<td>15</td>
</tr>
<tr>
<td>AUE</td>
<td>Actual User’s Expectations</td>
<td>51</td>
</tr>
<tr>
<td>DA1..DA3</td>
<td>Dragon age 1, 2, and 3 respectively</td>
<td>33</td>
</tr>
<tr>
<td>F1..F3</td>
<td>Female participant 1, 2, and 3 respectively</td>
<td>38</td>
</tr>
<tr>
<td>GUX</td>
<td>Game user experience</td>
<td>10</td>
</tr>
<tr>
<td>HCI</td>
<td>Human computer interaction</td>
<td>1</td>
</tr>
<tr>
<td>ISO</td>
<td>International organization for standardization</td>
<td>8</td>
</tr>
<tr>
<td>IUE</td>
<td>Intended User’s Expectations</td>
<td>45</td>
</tr>
<tr>
<td>KWIC</td>
<td>Keyword in context</td>
<td>29</td>
</tr>
<tr>
<td>M1..M4</td>
<td>Male participant 1, 2, 3, and 4 respectively</td>
<td>38</td>
</tr>
<tr>
<td>RoC</td>
<td>Repertoire of character</td>
<td>13</td>
</tr>
<tr>
<td>RQ</td>
<td>Research question</td>
<td>3</td>
</tr>
<tr>
<td>RTHR</td>
<td>Ride to Hell: Retribution</td>
<td>29</td>
</tr>
<tr>
<td>SCI</td>
<td>Sensory-, challenge- and imaginative based immersion</td>
<td>15</td>
</tr>
<tr>
<td>TD</td>
<td>The Division</td>
<td>30</td>
</tr>
<tr>
<td>UCD</td>
<td>User centered design</td>
<td>8</td>
</tr>
<tr>
<td>UX</td>
<td>User experience</td>
<td>1</td>
</tr>
</tbody>
</table>
List of Tables and Figures

Tables
Table 1: Overview of papers in this dissertation Page 4
Table 2: Overview of scientific approach, purpose, unit of analysis, research questions, overall methods, and contribution of the two studies involved in the dissertation Page 20
Table 3: Alphabetical list of games used in Study 1:a Page 27
Table 4: Games used in Study 1:b Page 29
Table 5: Overview of the papers included in the dissertation. Cells marked with Ω indicate that the paper has a relevant contribution to the research question. Cells marked with α indicate that the Study in some sense contributes to the solution. Page 37
Table 6: "Slogans" from Study 2:a as a result of using the UX theory in game user research. Page 41
Table 7: Aspects distributed on review ratings, for all three games in the franchise. The numbers are from the evaluators’ sentiment analysis, and show the frequency of each entry. Page 43
Table 8: Chi-square values for each aspect, of all three games in the franchise Page 44

Figures
Fig. 1: The Model of User Experience Page 14
Fig. 2, 3, 4: The graphs show willingness to play more (top graph), changes in sense of excitement (second graph), and boredom (bottom graph) the participants feel when playing the game after 1, 5 and 20 hours Page 46
Fig. 5a: Figure 5a: The developer must ensure that the intended user’s expectation is in conformity with the actual user’s expectations. The Playsona process (Figure 5b) may be used to achieve this. Page 52
Fig. 5b: The Playsona is created from aspects derived from user reviews of similar games. The developer combines suitable aspects into a Playsona. The purpose is to help the developer to focus on the right things. Page 52
Acknowledgement

As so often happens when one goes on an adventure without a specific goal, the journey is more important than the goal itself. When I first set out on this adventure, or maybe endeavor would more correctly capture the essence of my work presented in this book, I had only a vague conception of where it would take me. As luck would have it, I was blessed with my dear friends and colleagues - way more gifted at this craft than I -, Henrik Warpefelt and Magnus Johansson. With their aid I was able to begin my humble contributions to the academic and scientific community. However, there would not have been a book for me to write these words in if I had not had my supervisor, mentor and good friend Harko Verhagen. With an inexhaustible energy, unmatched enthusiasm, and an almost uncanny ability to pick me up from my deepest despairs, Harko has been the pilot that I needed in order to set my course and goal for this journey. For this, I am eternally grateful my friend.

For all the rest of You, too many to mention, I would like to extend my deepest gratitude for Your understanding when I have been unable to attend social events, when I have glazed over in mid-conversation due to my mind being pre-occupied with study designs, when I have neglected, postponed, forgotten, ignored and flat out hid from you. It was never personal, I love you all.
Thank You for being my friends and family.

/B
Introduction

Shooting with a bow and arrow may look deceivingly simple. All you need to do is pick up the bow, put the arrow on the string, pull back and release. However, if you have any intention of hitting a certain mark, the art of archery becomes considerably more complicated. A coach specialized in archery will show you how to overcome many of the challenges along the way, such as selecting the correct bow for the task; selecting the correct arrow for the bow; correct position of the feet; correct grip with the bow hand; proper angle of the wrist; correct finger setting on the draw-hand; correct drawing of the string; aiming; breathing; release and follow through; each step must be thoroughly trained and practiced, before the archer can start hitting the aimed at target.

The archer analogy is true about game design. From a distance, it looks deceivingly simple: come up with an idea, code the gameplay, make graphics and just sell it. Coding and graphics design aside (both are vast technological and philosophical subjects in themselves, and will not be further touched upon in this thesis), the actual video game, and the final user experience that follows, depend on whether several design elements are correctly aligned in a delicate balance of entertainment, challenge, and usability. The proverbial archery coach that almost the entire software industry invokes is called User Experience (UX). The UX of an artifact is derived from the sense of satisfaction the use of the artifact provides. A high sense of UX is typically achieved by thorough usability and accessibility analysis during the development of an artifact. For several decades, Human Computer Interaction (HCI) has been an integral part of software development. The development of user friendly interfaces is a constantly ongoing process, and the HCI research field is vast and well established. Video game software is no exception; in the early 2000s (Desurvire, Caplan, & Toth, 2004; Desurvire & Wiberg 2009; Federoff, 2002) game researchers used HCI methods to create domain specific alternatives (Pinelle, Wong, & Stach, 2008), with the aim to enhance user interactions in video games, and other researchers (Laitinen, 2006) showed how generic HCI tools, developed for and around productivity software, had low compatibility with certain aspects of video games. Focusing on usability, however, is not enough. Being able to “use” the game is a bare minimum, as the users expect much more out of the experience of play (Zaman & Abeele, 2012).
UX has been researched in many ways in general HCI areas (Hassenzahl, Platz, Burmester, & Lehner, 2000). The concept of UX reaches beyond pragmatic usability, and into more affective issues, such as positive emotions towards an artifact. Video game research is no exception. UX in games is a large field in game research, and has been steadily growing for at least a decade.

The ability to design game mechanics that will keep the users interested and engaged is critical in achieving a long-term relationship with the user. A publisher’s reputation is highly dependent on user attitudes, and on a constantly expanding market, mistakes can be costly. The challenge for the developers lies in finding the pivotal points when a new and intriguing game feature may lose its charm, and to balance that content with retentive content. In this thesis, I will present the ways that I have explored these challenges, mostly by examining what users dislike rather than what they like.

Video games are an engaging and interactive type of entertainment, and the level of engagement is determined by a game’s potential to retain the user’s interest. The user is central in this context, and the ability to develop for the user is dependent on the user information that the developer can gather; this is true in almost any customer relation. Keeping constant contact with users can be costly and inconvenient. In this dissertation, I present an investigation of player attitude and retention; I present suggestions on how to derive intelligence from user created reviews and how key game aspects in online discussions can be used to improve game design. Furthermore, I will describe how the users’ perception of a video game changes over time and what causes this change. What follows is an analytical description of how to make users feel satisfied, or perhaps more accurately how to avoid making them dissatisfied. As we shall see, that is not always the same.

Purpose and Research Questions

In this work, I explore various video game design elements that may cause a negative user experience. The data is collected from user diaries, interviews, and from online user reviews concerning popular game series. My ambition is to show the different root causes that make users less inclined to stay engaged or even become bored and frustrated with a game, and thereby terminate their engagement entirely. Similar studies have been conducted about other media phenomenon (O’Brien & Toms, 2008) and about the user engagement cycle of video game players (Kultima & Stenros, 2010). Many scholars have written about UX, both in general terms (Hassenzahl, 2003; Hassenzahl et al., 2000; Hassenzahl & Tractinsky, 2006) and more specific for the video game domain (González Sánchez, et al., 2009; Nacke et al., 2010).
The studies presented in this dissertation overlap with these previous studies, but they also add additional perspectives. Users build expectations, often based on the game designer's previous reputation and games. The developer must be able to predict and prepare for these expectations to create a successful game. If the experience from playing the game deviates too much from these expectations, the game may be negatively received, which in turn can lead to poor reputation.

Many games, regardless of a popularity rating (Strååt, Johansson, & Rutz, 2015), contain issues that could easily have been averted during the design process. Furthermore, users tend to express experiences with issues that the developers were not even prepared to address (Strååt & Verhagen, 2014). Thus, the need to have access to UX throughout game development is a relevant issue. Game developers already use a variety of methods to measure the UX of video games, such as telemetry data collected directly from the game client, interviews, user observation studies, beta-testing and expert evaluation methods such as heuristic evaluations. Notwithstanding these efforts, games are released with issues that negatively affect the player experience. This fact makes it necessary to investigate if the previously mentioned methods can be further developed or complemented, and if new methods can be introduced. Furthermore, it is also compelling to investigate UX theories that explore non-game artifacts, and see if and how these can be used to enhance game UX methods. For instance, what can be found from the perspective of these theories, and how can it be used in game development? Users of computer games often share their experiences on web forums such as gamespot.com or metacritic.com. How can these written testimonies of experience be used? What can be learned from the users indirectly, by analyzing their own words? And in the case of a prolonged user engagement, what lessons can be learned from a longitudinal study of game UX?

In this context, the overall research question (RQ) that has guided this research is the following:

RQ 1: How can novel methods be used within the context of game user research?

1. In what ways can existing UX theories be used within the context of game user research?
2. In what ways can indirect user feedback be integrated into game user research?
3. In what ways can longitudinal user feedback be integrated in game user research?
Research Design & Method Overview

For a more in-depth description of the methods applied in this thesis and the articles, please see section Scientific Approach and Method, or the methods section in each separate paper. This dissertation comprises two studies, study 1 and study 2. Table 1 provides an overview of the methodology, approach and objectives of the studies. The first two papers involved, Study 1:a and Study 1:b, explore the domain of video game experience using evaluation tools (such as heuristics) and user expressions (as expressed in an online video game forum). Study 1:a and Study 1:b used the case study method for data collection with a deductive approach. The purpose was to explore the research subject to contribute to a new hypothesis. The results allowed me to identify problem areas with expert evaluation methods (Study 1:a), and to identify divergent user expressions in user created video game reviews (Study 1:b).

Study 1:c consists of a theoretical discussion, summarizing Study 1:a and Study 1:b, underscoring that design elements which disrupt the game experience are easy to find, but design elements that enhance the game experiences are more difficult, and that users rarely comment on interface issues, unless they found the interface to be disruptive to the game experience.

In Study 2:a, I conducted a longitudinal case study, using surveys, player diaries and interviews as data collection methods, to examine first hand where and why a change in attitude towards a video game happens. Finally, in Study 2:b, an elaboration of Study 1:b is provided where a manual aspect based sentiment analysis (Pontiki et al., 2016) of all user reviews from a major video game series is conducted.

Table 1: Overview of papers in this dissertation, their main scientific approach, the objective of research, methods, and contributions.

<table>
<thead>
<tr>
<th>Study 1</th>
<th>Study 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Design</strong></td>
<td>Exploratory</td>
</tr>
<tr>
<td><strong>Subject of research</strong></td>
<td>Tool evaluation</td>
</tr>
<tr>
<td><strong>Method</strong></td>
<td>Case study</td>
</tr>
<tr>
<td><strong>Contribution</strong></td>
<td>Problem identified</td>
</tr>
<tr>
<td></td>
<td>User expressions</td>
</tr>
<tr>
<td></td>
<td>Literature study</td>
</tr>
<tr>
<td></td>
<td>Conclusion study 1</td>
</tr>
</tbody>
</table>
Contributions to the Field of Game UX

This thesis contributes both to the video game research community and to the video game development industry. The contributions to the research community are the methods of gathering and analyzing user comments used in Study 2:a and Study 2:b., namely the Focused Diaries and the aspect-based sentiment analysis performed on user-created reviews.

The contribution to the video game industry is the Playsona model (described in the discussion chapter) and the list of retention mechanics from Study 2:a. Furthermore, the video game industry might benefit from the methods described in Study 2:b.

Thesis Outline

This thesis starts with an introductory chapter (the current chapter), which is intended to give the reader an overview of the topic and an understanding on the importance and impact of understanding UX in video games. Following this chapter is the background chapter, where relevant theories, models, and tools are introduced. These models and theories have either been an inspiration to me and my ideas, or have been key elements in my own research. I then present my scientific approach and arguments for the methods I have used. In the following Results & Analysis chapter, I present the results from my studies. The following chapter contains conclusions and discussion on the topic of this thesis, with some ideas and suggestions on future work. This is followed by a chapter describing the papers involved in this thesis in terms of purpose, method, and conclusions.
Background

My interest in usability and UX in computer games was awakened long before I started writing this thesis. Games have been my hobby for as long as I can remember, but I have always felt a certain degree of love-hate relationship towards them: why are some design elements so frustrating, and why did the developer decide to put them in the game?

To understand video game UX, and video game usability, it is important to be at least briefly acquainted with “regular” HCI/human machine interaction theories and practices. This chapter is an introduction to the concept of HCI, with a brief description of the most common characteristics of the topic, and how it relates to video games.

HCI, Usability and UX

*It must be borne in mind that the object being worked on is going to be ridden in, sat upon, looked at, talked into, activated, operated, or in some other way used by people individually or en masse. When the point of contact between the product and the people becomes a point of friction, then the industrial designer has failed. On the other hand, if people are made safer, more comfortable, more eager to purchase, more efficient – or just plain happier -by contact with the product, then the designer has succeeded.*


HCI studies the interaction between humans and computer interfaces. In its earliest day, HCI focused on pragmatic views of interaction, where the user should be able to accomplish the tasks s/he set out to perform: the interaction with the machine should be goal oriented, and could sometimes demand expert skills from the operator. For instance, Lucy Suchman (1987) compared the interaction between the user and the machine with communication between humans; in earlier times, when we used machines to perform our jobs (dig, drive, lift), the operator would start the machine and then control its tasks. In modern operation, we tend to communicate with the machine to do
more advanced tasks, a communication that sometimes can be compared to a
dialog: the machine asks for our input and gives us feedback, which leads to
our next input etc. until the task is completed.

The quality of this interaction is referred to as usability, a term that reflects
various quality measures, such as time taken to perform a specific task and
goal achievement (Benyon, Turner, & Turner, 2005). The concept of usability
is so fundamentally important that the International Organization for Stan-
ardization (ISO) developed a standard, ISO 9241-11, to define its meaning

The ISO standard defines usability as the “Extent to which a product can
be used by specified users to achieve specified goals with effectiveness, effi-
ciency and satisfaction in a specified context of use,” where effectiveness re-
flects accuracy of goal achievement, efficiency reflects resources in relation to
the accuracy and completeness of goal achievement, and satisfaction reflects
positive user attitudes towards the use of the product.

Efficiency and Effectiveness were in large part the major goals for early
HCI scholars, and are still the ones that can easily be measured in quantifiable
ways (for example through measuring time for task completion). As comput-
ers have developed from terminals and workplace machines to the ubiquitous
presence they hold today and users are no longer experts but anybody, de-
mands for more satisfying interaction have risen. Computer users started to
use computerized devices for entertainment and pleasurable past-times, and
the industry needed to cater to the new needs that came with new users. This
shift in focus made way for the idea of user centered design and more research
was being performed on the satisfaction-part of the usability definition.

User centered design (UCD) is the philosophy that all design should aim to
humanize the interaction with technology and keep the users’ needs in focus
(Sward & Macarthur, 2007). The interaction and the extended relationship
with an artifact is referred to as UX. Cockton (2006, p. 166) provides an in-
teresting description of this extended relationship: “Quality in use and fit to
context are qualities of user experience during interaction. Outcomes and
lasting impacts endure beyond interaction”. In other words, not only do the
designer need to improve the actual use of an artefact, but also make sure it
provides a lasting sense of wellbeing with the user. Other scholars agree; for
example Sward and Macarthur (2007, p. 36) stated: “[UX]…extends UCD to
incorporate all aspects of the end user’s interaction with the product or ser-
vice and the organization that supports it,” meaning that the experience lies
not only in the actual use of the artifact but all things surrounding this use.
Furthermore, Sward and Macarthur (2007, p. 36) say that “Experience begins
with an awareness of the product or service and includes all aspects of the
end user’s interaction,” which can be interpreted as our UX starts by simply
realizing an artifact exists. The UX of interaction has become the leading fo-
cus for usability design, and by extension HCI.
UX also has an ISO definition; ISO 9241-210, clause 2.15 defines UX as: “[A] person's perceptions and responses resulting from the use and/or anticipated use of a product, system or service.” (ISO 9241-210:2010(en), 2017)

As can be seen in both the ISO definition and the definition provided by Sward and Macarthur (2007), UX not only includes actual use, but also anticipated use. This anticipation is further described by Hassenzahl et al. (2000), who addresses this phenomenon as “users’ situation” (please see Figure 1) which can roughly be described as the state of mind that the user is in when encountering the artifact for the first time – a user that is familiar with the designer, or the concept of the artifact will have a different situation from a user that has never previously seen or interacted with a similar artifact. Hassenzahl’s (2000) situation is derived from Janlert and Stolterman’s (1997) concept of character and user repertoire of character (please see chapter Theoretical Framework).

Hassenzahl et al. (2000) discussed the qualities of interaction in terms of pragmatic and hedonic attributes, where pragmatic attributes concern the actual purpose of an artifact (a cellphone is used for phone calls for example) and hedonic attributes concern the user’s wish to be entertained or achieve a certain UX (a specific brand of cellphones will convey a special experience, such as a certain lifestyle). They suggest that the hedonic attributes are the main reason for a user to select one artifact over another, if the pragmatic attributes are equivalent. This implies that a highly hedonic product (such as a video game; video games are often built around the purpose of entertainment) may be well implemented in the pragmatic sense, but still be unsuccessful due to insufficient attention to the hedonic attributes.

In conclusion, UX is derived from an ambition to meet the user demands of an artifact, not only from the actual use, but from many aspects concerning the artifact, and that the UX is dependent on the users’ pre-cognition of an artifact. In the next section, I will describe how this is related to video game design.

**HCI and UX in video game studies**

Video game researchers and developers have followed the timeline of HCI, usability and UX development, but with domain-necessary alterations. Relatively early, video game researchers established that “regular” usability and HCI measures and tools were insufficient and sometimes even counter-productive in the video game domain (Isbister & Schaffer, 2015). The concept of UX development in productivity software, such as operative systems, mail clients, word processors and so on, is not fully transferable to video games, due to the interactive entertainment nature of video games: games are often designed around challenges and problem solving, while productivity software
should be essentially free of challenges. Federoff (2002), for example, performed on-site observational studies on video game studios to develop her set of video game specific heuristics – a list of 40 heuristics divided into three categories: Game Play, Game Mechanics, and Game Interface. Laitinen (2006) separated the evaluation of the actual game – the “fun” – from other types of interaction when evaluating a game during its development. For the game-parts, he used Koivisto and Korhonen’s playability set (2006), and for the other parts (interaction with game setup menus etc.) he used regular and more generic HCI evaluation tools, such as Norman’s design principles (2013) and Nielsen’s heuristics (1995). Korhonen and Koivisto (2006) developed a set of playability heuristics for mobile games, and can be viewed as an early approach to UX in games, as it is aimed at the “satisfaction” part of usability. The “Playability Heuristics for Mobile Games” by Korhonen and Koivisto (2006) separate usability heuristics from playability heuristics, where the latter are aimed at the users’ entertainment and the former are aimed at the users’ convenience of use.

This early approach to game UX was soon followed by others. Sánchez, Padilla Zea, and Gutiérrez Vela (2009, p. 357) had a similar take on playability. They defined it as: “a set of properties that describe the Player Experience using a specific game system whose main objective is to provide enjoyment and entertainment, by being credible and satisfying, when the player plays alone or in company”. Playability is, according to them, a way to measure to what degree a player manages to reach specific goals, considering factors such as efficiency, skill, satisfaction and pleasure in a game environment. They propose an evaluation framework for improving playability. This framework is originally based on the three cornerstones of usability: efficiency, effectiveness, satisfaction, but developed various sub categories aimed at gaming and entertainment (Sánchez et al., 2012). The players’ sense of experience evolves and changes over time as the player gets more familiar with the game. Design elements that are intriguing and challenging in early game play must develop over the course of the game, to keep the players’ interest high enough to stay engaged (Fabricatore, 2007; Iacovides, Cox, McAndrew, Aczel, & Scanlon, 2015). As the player learns to operate the game, and steadily becomes better at it by playing, the challenges must be balanced towards the learning curve of the player.

Many game researchers suggest more appropriate definitions for UX in games, such as Gameplay Experience (Nacke et al, 2010), which is the experience the player has during the interaction with the game, or Player eXperience, suggested by Sánchez et al. (2012) which focuses on the playability of a game. A holistic approach is provided by the Game User Experience (GUX) (Iacovides et al., 2015), which considers both in-game challenges, players’ sense of flow and players’ overall engagement with the game. For this thesis, I will use the GUX definition, as it suggests that the UX of a game transcends
the actual act of playing it, and takes other aspects of the users’ total game experience into consideration.
This chapter describes the various theories and models that I have used and been inspired by in my own work.

The Character of Things

When people say things like 'this is a reliable car,' or 'this is a tedious word processor,' they do not refer to any specific function—they are rather making a high-level description of the artifact. (Janlert & Stolterman, 1997, p. 297)

Janlert and Stolterman (1997) described user perception of artifacts as a repertoire of “character,” which we carry within us. This repertoire is based on our previous experiences with similar artifacts. All items have “character,” i.e., some aspects that the user will recognize at their very first encounter with the artifact. The designer of an artifact relies on the user’s ability to recognize the character of the item that they are designing, thus conveying to the user the feeling of an interesting, practical tool that they want. The challenge of the designer is to realize and clarify these character aspects into the product.

Every user has a Repertoire of Character (RoC) that has developed over time. We, the users, use our repertoire to evaluate new artifacts. Introducing new features in a design requires adherence to existing character-features, which the designer then can develop further. The users’ RoC is constantly expanding, shifting and evolving, and designers should stay alert to these changes to deliver interesting, appealing, and satisfying design (Janlert & Stolterman, 1997).

Whenever users first come in contact with a new artifact, they form preconceptions of its features, or character (Janlert & Stolterman, 1997). Users construct a personal vision of the product’s character, which in turn allows them to determine the appeal of the product (e.g., "It is good/bad"), emotional consequences, such as a sense of pleasure and satisfaction, and behavioral consequences, for example increased time spent with the product (Hassenzahl & Tractinsky, 2006). After a period of use, the user develops a mental model (Norman, 1983) of the system. This mental model is constantly evolving as the user’s skill and habit improves from consistent use, which in turn forces
the avant-garde designer to evolve their design in a constant cycle of innovative development. When designing something new, it is often wise to keep some parts that the user can identify from their own RoC. If too big a step is taken away from the RoC, there is a risk that the users will not understand the product and will perceive it as foreign and less willing to use it. By keeping relevant parts from a known design, and altering or adding elements to it, the designer can introduce new characteristics to the users’ RoC and thus make them accept and appreciate the new product.

**Figure 1: The Model of User Experience (Hassenzahl, 2003).**

Model of User Experience

The model of user experience (Hassenzahl, 2003) illustrates how UX is the result of each user’s individual understanding of an artifact based on their previous experience and attitude. In the model (Figure 1) the designer perspective (a) represents the different design features – content, presentation, functionality, interaction – that the designer sets up to deliver the Intended Product Character (IPC), with the aim to facilitate use of the product and to introduce different user strategies.

Meanwhile, designers may or may not be successful in rendering their intent. Upon first contact with a product, each individual user will take in the
product’s immediately discernible features — the *Apparent Product Character* (APC) — and construct their own *user perspective* (b) on the product’s character (Janlert & Stolterman, 1997). The APC and the user’s individual version of it will lead to *consequences*: judgments of the *appeal* of the product, and, eventually, behavioral consequences (extent of usage of the product) (Hassenzahl & Tractinsky, 2006).

As might be expected, features or characteristics (intended and/or apparent) are not the sole determinants of how a product will be perceived and experienced. According to Hassenzahl (2003), the user’s perspective and consequences are highly dependent on the user’s *situation*, which is shaped by previous experiences with similar products. From a video game perspective, the user’s situation can be viewed as genre preference (for example action games, role-playing games), previously played games from the same developer, or relationship to an intellectual property such as Tolkien’s Middle Earth or J.K. Rowling’s Harry Potter universe.

**Hedonic and Pragmatic Attributes**

According to Hassenzahl (2003; Hassenzahl et al., 2000; Hassenzahl & Tractinsky, 2006), users will assign attributes to any artifact that they interact with. In a broader sense, there are two major attribute categories *pragmatic* and *hedonic*. The pragmatic attribute of *manipulation* is essentially about the core handling of the artifact (basically a measure of usability of the artifact). The hedonic attributes regard the users’ experience from a UX point of view: *Stimulation* relates to an artifact’s ability to challenge the users’ senses, and create opportunities for personal development. The artifact should inspire exploration and show which experiences “lie ahead.” In a game context, it may be about the game’s ability to introduce challenges in new and exciting ways, and to encourage the player to explore the game world.

*Evocation* is an artifact’s ability to evoke memories of past experiences, or emotions and thoughts that are important to the user. In a game context, this could be the game’s ability to evoke emotions such as joy, fear, satisfaction etc., and to allow the player to experience exciting or terrifying situations that they do not dare to experience in real life.

*Identification* is the way individuals express themselves and their personality through accessories, such as clothes of a certain brand, or printed with a music group or favorite sports team, watches, cell phones etc. The choice of possessions shows the individual’s position for or against something. In a video game, this may be exemplified by the avatar’s appearance and name.

Stimulation, identification and evocation are very similar to the game research concept of immersion; challenge based immersion from the SCI model (sensory, challenge-based and imaginative immersion) by Ermi and
Mäyrä (2005) for example, is related to stimulation, and evocation is strongly related to sensory immersion from the same model. A designer’s goal should be to create a product with potential to appeal to, satisfy and please the user. Hence, it is important for designers to familiarize themselves not only with user perspectives but also with different user situations, to be able to predict and perhaps even influence the consequences from the user’s perspective.

User Engagement
O’Brien and Toms (2008) studied user interactions with various digital media, such as games, online shopping and other online services. They wanted to know what motivates a user to start, uphold, and terminate an engagement with a digital artifact. They used a random data sample of 17 individuals who were interviewed about their experiences when interacting with various media. Each participant could freely choose which media they wanted to discuss. The study resulted in a model for describing user engagement before, during, and after interaction with the selected media. O’Brien and Toms’ model proposes two categories of engagement termination: positive, with a sense of achievement, and negative with a sense of frustration. Motives such as uncertainty, information overload, boredom and frustration are mentioned as the main reasons for negative termination.

The Expanded Game Experience Model
Kultima and Stenros (2010) suggest that the game experience starts long before the actual gameplay. They describe the entire cycle of the game experience; the potential user searches and evaluates information about different games from a variety of sources such as online reviews, friends, and game advertisement and so on. The user then selects a game based on the analysis. In short, the users create their own image of the game that they want to play, based both on previous experiences and the information that has been gathered. If the actual game experience deviates too much from this image, the user is likely to be disappointed and dissatisfied, and will terminate the game engagement. Kultima and Stenros illustrate this process in a very fine-grained model; each step is well described, from the user’s first decision to play a game to the final decision to terminate the play activity. Their model, with its step-by-step process description, has inspired my research.
Online User Reviews and Metacritic

The use of video game reviews as a resource for game studies is still a fairly uncommon phenomenon. Most of the studies that have been performed explore professional reviews, and these studies have shown the merit of game reviews from various perspectives. Pinelle et al. (2008) used professional reviews as a source to find common video game issues, which they compiled into a set of design patterns. Zagal, Ladd and Johnson (2009) found that game reviews often include design suggestions and serious discussions on game designer’s intention and goals, and Zagal and Tomuro (2013) studied cultural differences and similarities in user created reviews from Japan and USA. The paper that is closest to my use of user created reviews (Study 1:b and Study 2:b) is Koehler, Arnold, Greenhalgh, Boltz and Burdell’s “A Taxonomy Approach to Studying How Gamers Review Games” (2017). They used an existing theoretical model, a video game taxonomy, and compared user submitted reviews with the categories of the taxonomy. They found that users to a certain degree used the same concepts as the taxonomy, and that there was a difference in use of the concepts depending on the game rating.

My studies have all used user created reviews from the Metacritic webpage. Metacritic.com is a site that aggregates professional reviewer scores from various online media review sources. Television shows, movies, music and video games (various platforms) are examples of media that are presented. Metacritic calculates an average score called Metascore, based on the various professional reviewers by converting the reviewers’ local score into a score ranging from 0 to 100 (e.g., a local score of 8 out of 10 renders a Metascore of 80). These scores are weighted (based on the quality and overall stature of the source) and finalized into a professional Metascore.

Regular non-professional users are also allowed to score the media on a scale of 0 to 10. The unweighted average of this score is presented by Metacritic as the Userscore. Non-professional users can also post their own reviews along with their score. The Userscore does not consider the length or quality of these reviews; a simple four-word comment, such as “this game is good,” is valued the same as an analytical 500-word essay. User reviews and scores are posted anonymously under a self-selected user name. The user score is divided into three tiers: Positive, Neutral and Negative, where Positive refers to ratings 8 to 10, Neutral refers to ratings 5 to 7, and Negative refers to ratings 0 to 4. The rating tiers are color coded in green for Positive, yellow for Neutral and red for Negative.

Metacritic has been the subject of many discussions. The validity and value of the professional reviews have been questioned in various video game blogs and online magazines (Metacritic Matters: How Review Scores Hurt Video Games, 2015; Time to kill Metacritic, 2014), and the site has been used in
game and social studies, e.g., as an examination and comparison of player experience vis-à-vis professional reviews (Johnson, Watling, Gardner, & Nacke, 2014), or as an important factor in assessing game value and quality (Greenwood-Ericksen, Poorman, & Papp, 2013). Most commonly, the discussion has been on professional reviews.
Scientific Approach and Method

This chapter describes the overall scientific approach applied and the methods used for data collection and analysis. The methods described in this chapter are more thoroughly described in each paper. I have not used the same method throughout my papers; instead, each method has been selected based on the purpose of each paper. Table 2 provides an overview of the two studies, and the papers involved, their approach, purpose, research questions, main methods and contributions. Following this the ethical considerations are introduced. The chapter ends with a detailed methods description for each paper.
Table 2: Overview of scientific approach, purpose, unit of analysis, research questions, overall methods, and contribution of the two studies involved in the dissertation

<table>
<thead>
<tr>
<th>Study 1: Framing the Problem</th>
<th>Study 1:a</th>
<th>Study 1:b</th>
<th>Study 1:c</th>
</tr>
</thead>
<tbody>
<tr>
<td>Approach</td>
<td>Instrumental case study</td>
<td>Instrumental case study</td>
<td>Explanatory</td>
</tr>
<tr>
<td>Purpose</td>
<td>Exploratory/Assessment</td>
<td>Exploratory/assessment of user created video game reviews</td>
<td>Conclusion of findings in Papers I and II.</td>
</tr>
<tr>
<td>Unit of Analysis</td>
<td>Self-produced heuristic tool, popular and unpopular video games.</td>
<td>User created game reviews on selected games from Study 1:a</td>
<td>Study 1:a and Study 1:b</td>
</tr>
<tr>
<td>Method</td>
<td>A summative evaluation, using a heuristic inspection method, of games in different quality strata, and a formative evaluation of the self-produced evaluation tool through practical use.</td>
<td>Word frequency analysis, content analysis</td>
<td></td>
</tr>
<tr>
<td>Results/Contribution</td>
<td>New heuristics, further development of the evaluation tool, most commonly found issues were found in both high and low rated games.</td>
<td>Findings show that users are inclined to comment negatively; when the gameplay actively breaks their sense of immersion; when they dislike the game controls; when the game story feels uninspired. The inspection methods in Study 1:a only capture issues with controls, implying that issues with story or immersion are not possible to find using inspection tools.</td>
<td></td>
</tr>
</tbody>
</table>
Contribution to dissertation: Findings leading to future research (Study 1:b): are players experiencing the same issues that were found in Study 1:a?

User feedback can be accessed from online reviews to garner user experience and attitudes. This is a first step in answering the first sub-question of the RQ.

<table>
<thead>
<tr>
<th>Study 2: Addressing the Problem</th>
<th>Study 2:a</th>
<th>Study 2:b</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Approach</strong></td>
<td>Intrinsic case study</td>
<td>Instrumental case study</td>
</tr>
<tr>
<td><strong>Purpose</strong></td>
<td>Exploratory</td>
<td>Exploratory / descriptive</td>
</tr>
<tr>
<td><strong>Unit of Analysis</strong></td>
<td>Participants’ attitude and experience</td>
<td>Aspects in user created game reviews</td>
</tr>
<tr>
<td><strong>Research Questions</strong></td>
<td>Finding root-causes for negative disengagement, longitudinal examination of changes in user attitude</td>
<td>Testing the null hypothesis that there is no relationship between the game rating and the sentiment value of the examined aspects. Testing the method as such</td>
</tr>
<tr>
<td><strong>Method</strong></td>
<td>Longitudinal user study, user diaries, repeated interviews, content analysis</td>
<td>Word frequency analysis, content analysis, aspect based sentiment analysis, statistical hypothesis testing (chi-square)</td>
</tr>
<tr>
<td><strong>Results/Contribution</strong></td>
<td>A set of “slogans” that describe game design elements that cause user disengagement and negative attitude/frustration.</td>
<td>The sentiment of an aspect reflected the rating of the review, indicating that user created content can be used to quickly assess user attitudes towards game aspects.</td>
</tr>
<tr>
<td><strong>Contribution to dissertation</strong></td>
<td>Longitudinal feedback from diaries showed that the initial user experience and attitude changed after a relatively short period of time. This indicates that cross sectional studies may give insufficient or false information about user attitudes. Answers RQ 1.2, in part answers 1.3</td>
<td>The results from this study are used in the solution of RQ 1.1, and in part 1.3. Furthermore, the Playsona-model introduced in the discussion is based on the results from Study 2:b.</td>
</tr>
</tbody>
</table>
Case Study Approach

Although my methods have varied with each paper, the overall approach has been a case study. The case study approach is used when the research is focused on providing in-depth understanding of the studied object. Done “by the book,” they focus on one singular instance of the item of investigation. To be able to make general assumptions from case studies, the researcher creates more studies and combines them into one concluding work, which is what I do in this dissertation.

The aim of a case study is to “illuminate the general by looking at the particular” (Denscombe, 2014, p.53). The approach of case studies varies with the intention of the researcher. Creswell (2012) described various approaches: the intrinsic case study, for example, is performed with the intent to examine a unique case that is inherently and intrinsically interesting. The instrumental case study is performed in cases when the researcher wants to understand a phenomenon, often consisting of several cases. The exploratory case study explores the phenomenon. The explanatory case study attempts to explain the studied phenomenon.

My case studies are a combination of at least three varieties - Intrinsic, because the games have unique attributes that I wanted to understand; instrumental because the games, attributes, and game UX taken together form a phenomenon; exploratory, as each case has presented a new and unexplored field within the context of game UX. For example, Study 2:a is presented as an intrinsic case study as it considers the specific case of playing a newly released action game. It is also instrumental as the purpose of the study was to examine the general phenomenon of player retention and longitudinal changes in attitude.

Case studies are appropriate when the researcher wants to study the “how” and “why” of a given object of research (Yin, 2002). It is also an appropriate approach when the researcher cannot directly interact with the participants (for example when the data is gathered from written documents such as newspapers or journals) or when the context of the matter is relevant.

Creswell described the “case” as a bounded system, process, and/or activity (Creswell, 2012), and Baxter and Jack (2008) describe the case as the unit of analysis that the researcher wants to analyze. In my studies, the case study approach fits well. Study 1:a used games (bounded systems/activities) and a
list of heuristics for evaluation (process) in a context sensitive situation. Study 1:b had user created content (game reviews) as the object of analysis; we, as researchers could not get involved with the users, and the relevant question of the study was “why” and “how” regarding the users’ attitudes. Study 1:c is odd in that it is a theoretical discussion regarding Study 1:a and Study 1:b, but it still concerns a bounded system (the two papers). Study 2:a clearly fits the description, as it is used for participants engaged in a system, where the purpose of the study is to understand how and why players change their attitude towards the game they are playing. Study 2:b replicates parts of the methods used in Study 1:b.

My data collections have been recordings, observations, interviews, participant diaries (play diaries for a longitudinal data collection in Study 2:a), and online user created content (game reviews).

The longitudinal study in Study 2:a can be considered a variant of phenomenology. The phenomenological approach is used with the purpose of describing or mapping the experience of several individuals who share the same experience (Creswell, 2012), which was the case with Study 2:a.

**Data Collections and Units of Analysis**

Study 1 used the following data collections.

- In-game observations
- Field notes
- Written documents, in the form of user created reviews
- Literature reviews

The in-game observation data and field notes data used in study 1:a were collected by playing the games involved in the study. The collection methods suited the purpose of the study, which was to test a set of heuristics and to gather information on new heuristics.

Study 1:b collected data from an internet web forum. This data could have been gathered from interviews with video game users, but a number of inconvenient issues arise: the respondents may not be as eloquent when they form their thoughts in a spoken reply; the number of respondents would be much lower (there were several hundred posts gathered from the web forum in study 1b); the recruitment process would require us to find a large number of users for each particular game; the respondents may not have a strong opinion of
the game at all; and finally, the researchers would not be able to formulate questions at a high enough fidelity to reach the same result.

Study 1:c is a theoretical discussion regarding Study 1:a and Study 1:b, which means that the data collection was already performed.

Study 2 used the following data collections
- Interview data
- Field notes
- Written documents, in the form of user diary entries (2.a) and user created reviews (2.b)
- In-game observations
- Word frequency analysis
- Semi structured interviews

Interview data was collected from the participants in Study 2:a, prior to, and after they had tried the game. The other data source for Study 2:a was written documents, from the Focused Player Diaries that the participants wrote during the study.

Study 2:b used written document data from Metacritic.com online. All user-created reviews of each of the games involved were collected from the site. Furthermore, I played each game involved in Study 2, to become familiar with the environment and to build a basis for participant interviews.

Data Analysis Methods

Content Analysis
The user created reviews from Metacritic used in Study 1:b were analyzed through content analysis. The purpose of Study 1:b was to explore the users’ attitudes, which makes the content analysis a mix between Directed and Summative (Hsieh & Shannon, 2005). The summative content analysis aims to discover underlying meanings of the content, after identifying and quantifying this content; we wanted to understand the underlying meaning of the user reviews. The directed content analysis identifies key concepts of the material, using an existing theory or prior research, in our case a list of heuristics. A variant of content analysis was used in Study 2:a, where the participants were instructed to write diaries, reply to surveys, and participate in semi-structured interviews.
Aspect Based Sentiment Analysis
Study 2.b used an aspect based sentiment analysis as the major analysis method. An aspect based sentiment analysis (ABSA) is performed when user sentiments of certain aspects of a multi-aspect entity are measured, in a dataset gathered from user comments, such as online forums or user created reviews (Pontiki et al., 2016). Video games have plenty of aspects (sometimes referred to as attributes) that the user considers when playing, e.g., playability, graphics, and storyline.

Aspects are words or phrases that exist either explicitly or implicitly in the dataset. Explicit aspects are the actual word in context, and implicit aspects are inferred from the context. For example, if the aspect is gameplay, an explicit occurrence could be “I really enjoyed the gameplay,” and an intrinsic could be “I really enjoyed the challenges and the features of X.”

The aspects are determined through a word frequency analysis. After the dataset is collected, product or domain relevant words that occur on a frequency above a pre-determined threshold are retained for the following sentiment analysis step. The sentiment analysis is performed either through a scripted natural language processing algorithm, or through a manual read through. The result will show the sentiment for each aspect, for example in terms of positive, neutral, or negative sentiment.

Word frequency and selection
As mentioned in the metacritic description in the background section, metacritic authors rate their own comment to reflect their experience of the game in question. This is a rating from 0 to 10, which is used to categorize the comment as: low, medium, or high rated.

In the cases where user reviews have been used as a data source, a word frequency analysis has been performed to find which word or concept is most frequently used in the reviews. The word frequency is the determinant for the continued research process, e.g., for selection of reviews to be used in the ABSA.

Ethical Considerations
All the studies involved in this dissertation has followed the ethical standards of the Swedish Research Council (Vetenskapsrådet 2002). All participants involved in the studies in this thesis are anonymous. At the time of their participation, the participants were instructed that they could discontinue any part of the survey, interview, or data collection at their own discretion and without
any questions asked. Participation has been completely voluntary in all cases. Full transparency of the research purpose has been conveyed to the participants. To the utmost extent, the researchers have strived to give all participants all information needed to achieve informed consent. No study can be said to break any laws at the time of the conduct. All video games and other artifacts in the studies involved in this thesis have been selected based on certain attributes, including but not limited to year of release, genre, and popularity (as evaluated by a third party).

The data collected from the web forum comments (metacritic.com) has been anonymized: in the articles with users, all the user names have been removed and all user comments have been omitted or only partially quoted. It has been my ambition to have broad, inclusive and representative samples of video game players. I have, to the best of my knowledge, never willingly or intentionally excluded or in any other way treated differently any participant on behalf of their gender, creed, sexuality, abilities, or ethnicity.

Method Description per Paper

Study 1:a

For Study 1:a, we selected the five highest and the five lowest rated PC video games on metacritic.com. The game genres were first-person shooters, action games, and role-playing games. These genres were chosen because they often provide a large, immersive and interactive “world” for the player to act in. Table 3 lists the games with title, developer, and Metacritic rating and user score.

Games used in Study 1:a

Table 3: Alphabetical list of games used in Study 1:a

<table>
<thead>
<tr>
<th>Title</th>
<th>Developer / Publisher</th>
<th>Metascore, user score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assassins Creed IV: Black Flag</td>
<td>Ubisoft Montreal/Ubisoft</td>
<td>86/100, 7.8/10</td>
</tr>
<tr>
<td>BioShock Infinite</td>
<td>Irrational Games/2K Games</td>
<td>94/100, 8.5/10</td>
</tr>
<tr>
<td>Borderlands 2</td>
<td>Gearbox Software/2K Games</td>
<td>89/100, 8.1/10</td>
</tr>
<tr>
<td>Brothers: A tale of two sons</td>
<td>Starbreeze Studios AB/505 Games</td>
<td>90/100, 8.6/10</td>
</tr>
</tbody>
</table>
DARK
Realmforge Studios/Kalypso Media
Digital
41/100, 4.2/10

Legends of Dawn
Dreamatrix/Dreamatrix
29/100, 4.8/10

Ride to Hell: Retribution
Eutechnyx/Deep Silver
19/100, 1.3/10

Realms of Arkania: Blade of Destiny
Crafty Studios/United Independent Entertainment GmbH
18/100, 2.3/10

Takedown: Red Sabre
Serellan LLC/505 Games
32/100, 2.5/10

Tomb Raider
Crystal Dynamics/SQUARE ENIX, Eidos Interactive
84/100, 6.9/10

Data collection
To gather data for this paper, we performed heuristic evaluations on the games, using our own set of heuristics. The heuristic evaluations followed the process according to Nielsen (1995), where each evaluator individually played and evaluated each game, and made notes of all the issues they encountered on a report card. One of the evaluators also had the role of evaluation manager. Each game was played until no new interaction styles could be detected. Data saturation normally occurred after two hours of playing. Each evaluator performed their evaluations individually and with no communication with any of the other evaluators. This is standard procedure for heuristic evaluation, with the purpose of maintaining inter-observer reliability within the evaluation group.

Data analysis
Following the evaluation phase, the findings were analyzed. Any ambiguities in the issue descriptions were clarified; issues that were deemed to not concern game world interaction were removed. All remaining issues were rated according to a severity rating ranging from 0 to 4, where 0 is “cosmetic problem” and 4 is “unplayable.” The findings were then aggregated into a list by the evaluation manager. As a final step, the aggregated list of issues was matched with the Net Heuristic List by the evaluation manager.

Each issue was matched to all applicable heuristics from the Net Heuristic List. If an issue could not be matched to a heuristic, it was referred to as “Heuristic Missing.” Two categories emerged in the following analysis of the “Heuristic Missing” issues.

Study 1:b
For Study 1:b, we chose three games with similar interaction styles and genres. Just as with Study 1:a, we selected games based on their metacritic user
rating. A high, a medium and a low ranked video game, released in 2013 were selected. From these three games, we collected all user reviews. Please see the background chapter for details on metacritic and user metascore.

**Games used in Study 1:b**

These games were chosen due to their similarities in gameplay, genre and interaction styles. The game rating was also a criterion: we wanted different strata of game qualities, which are represented in the Metascore (see Table 4 for details).

*Table 4: Games used in Study 1:b*

<table>
<thead>
<tr>
<th>Title</th>
<th>Developer / Publisher</th>
<th>Metascore, user score</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assassins Creed IV: Black Flag (ACIV)</td>
<td>Ubisoft Montreal/Ubisoft</td>
<td>86/100, 7.8</td>
<td>High</td>
</tr>
<tr>
<td>DARK (DARK)</td>
<td>Realmforge Studios/Kalypso Media Digital</td>
<td>41/100 , 4.2</td>
<td>Medium</td>
</tr>
<tr>
<td>Ride to Hell: Retribution (RTHR)</td>
<td>Eutechnyx/Deep Silver</td>
<td>19/100, 1.3</td>
<td>Low</td>
</tr>
</tbody>
</table>

**Data Collection**

We gathered all player reviews from metacritic.com into three sets, one for each game. For each set of player reviews, we conducted a word frequency analysis. From the resulting word lists, we picked the first five words according to the following categories:

- Words that were most frequently used to describe the players' feelings towards the game (Adjectives/Adverbs).
- Words that were most frequently used to describe the game or aspects of the game (Game descriptive).
- Words that describe specific actions and activities within the game (Gameplay).

For example, for ACIV, we found:

Adjective/Adverb: good, great, best, bad, repetitive

Game descriptive: story, graphics, storyline, control (controls), mechanics

Gameplay specifics/activity: combat, sailing, missions, battles, gameplay

**Data Analysis**

These words were inserted into Key Word in Context (KWIC) concordance lists, which shows the selected word in its context. That way it is possible to
analyze how the author of the review intended for the word to be interpreted. We then compared the KWIC with the same heuristic list used in Study 1:a, to see if we could detect any heuristic matches in the players' reviews. Finally, each review was classified as either positive or negative.

Study 1:c
Study 1:c was a consolidation of the two previous studies, 1:a and 1:b. It should be viewed as a theoretical discussion regarding the findings from the two papers. No further data collection was performed for Study 1:c.

Study 2:a
The purpose of the study was to measure how players’ attitude towards a game changes over time, and in which ways the game meets their expectations. For this purpose, we needed a game that participants knew of, but had not yet played. It was therefore important that the game was not yet released at the start of the study. We deemed it important that the game should be a high budget production; it needed a strong marketing budget to ensure that it was well known in video game news media. We also wanted participants who knew about the game to some degree, and who were interested in playing it. This requirement existed to make sure that the participants’ ability to form a pre-understanding of the game would be as equal as possible.

Game used in Study 2:a
The game that was selected for this study was Tom Clancy’s: The Division (TD), by Ubisoft and Ubisoft Massive (2016). It was released on March 8, 2016. It was highly anticipated (the promotional campaigns had been going on since 2013), and the participants of the study were aware of it, but had never played it. These attributes were all in favor for picking TD as the object of study.

The game is a third person shooter, with role playing game elements. The game can be played in single player mode, cooperatively online with other players, and in a player versus player competitive mode.

Participants
During the recruiting phase, a number of individuals were asked if they knew about the game, and whether they were interested in playing it. If they matched these criteria, they were invited to join the test. Our budget allowed for eight games to be purchased and handed out to the participants. Four male and four female participants, in the age range from early twenties to mid-thirties, were
picked. One female participant could not participate in the study due to technical issues. These technical issues occurred after the recruiting period was over, so we were not able to find a substitute. The final participant count was seven individuals: three female, four male.

According to the startup interview, all participants enjoy role playing games, strategic games and cooperative games. They all identify as video game players, in the sense that they play video games as a hobby. They play commercial triple A titles, and they keep themselves informed of new game releases and related news. Their video game experience ranges from 15 to 30 years, and they play video games between 4 to 20 hours a week. Five of the seven participants preferred game world, setting and narrative over game play elements and two of the seven participants preferred game play over narrative and storyline.

Data Collection

The research method was designed based on Hassenzahl’s (2003) model of UX (please see Background chapter, section UX theories and Models, Figure 1). The model of UX illustrates how a product is perceived, from the user's perspective, based on its apparent product character (APC). The attributes of the APC in combination with the user’s situation (the user's experience or attitude towards a similar product) ultimately leads to a consequence, in the form of the user’s attitude towards the product (the model also shows the designers perspective, but I did not utilize this in Study 2:a, since only the players’ change of attitude was of interest).

As the user gets more accustomed to the product, for example through continuous engagement, the user’s situation changes. In Study 2:a, this change was possible to measure by applying the ideas behind the theoretical model. The APC consists of two groups of attributes: pragmatic and hedonic, where the hedonic attributes stimulation, identification, and evocation were used to form the questions and surveys that were part of the player diaries in Study 2:a. Through the diaries, we could collect data that specifically concerned the participants’ attitude to the game prior to engagement (when the game only has an apparent product character), during engagement (when the participants are familiar with the game product character), and eventually after a prolonged time of engagement (when the players have formed their final attitude towards the game). The final diary was followed by individual interviews to capture the participants’ attitude towards the game. As we had three data entry points for the diaries, I performed three separate coding sessions, for the first, fifth and twentieth hour respectively. The data was divided into four groups, derived from the attributes in the model of UX: Manipulation, Stimulation, Identification, and Evocation.
Startup Interview

I interviewed each participant individually in order to get to know their video game preferences and their previous experiences of the genres involved in TD. This was done before the release of the game, and before any of the participants had played it. The interviewer and the participant watched an official gameplay video, released by the producer, which showed various key elements and features of the core game: combat, loot, player versus player combat, abilities and items. During the screening, I asked the participants what they thought about the game: did they think that it looked interesting or exciting, did they want to play etc.

By showing the players the video before they could play the game, they were introduced to the producers “intended product character” (described in more detail in the chapter Theoretical Framework, section Model of UX) of the game. This was done as I wanted the participants to start to form a “character” of their own.

Focused Player Diaries

I decided to use player diaries as my main data collection method as I wanted to capture the participants’ feelings and thoughts as close to the experience as possible. By repetition and standardization of questions (1, 5 and 20 hours, asking the same questions), I could also gain data that is difficult to obtain through other methods, such as observations.

The intervals of the diaries were 1, 5 and 20 hours respectively. The 1-hour mark captured the absolute first impression of the participants. At the 5-hour mark, the participants would have familiarized themselves with the core mechanics and story of the game and at the 20-hour mark there was a good chance that they had advanced or even finished the basic storyline of the game.

At the beginning of the study, I handed out the first player diary, with instructions to play for exactly one hour, and then fill out the diary and email it to us. Once I had the first hour diaries, I sent out the second diary with instructions to fill it out and email it back after they had played a total of five hours. This process was repeated for the last diary, which was to be filled out when the participant had played a total of twenty hours.

In the diaries, the participants were instructed to write about their gaming experience so far, to write a short summary of a typical game session, and to answer a survey about their play style. The diaries contained questions about their experience with certain game elements such as missions, non-player character enemies etc. and in each diary the participants were asked to write their three favorite things about the game, as well as their three least favorite things. The questions were inspired by the attributes manipulation, stimulation, evocation, and identification.
In each of the diaries (1, 5 and 20 hours respectively) the participants were prompted to answer a set of six statements on a Likert scale from 1 to 10 (where 1 is “not at all” and 10 is “very much”). These statements regard interest in playing more, sense of excitement, fright or uneasiness, frustration, boredom, and disappointment. If they rated below or above a certain grade (depending on the question), they were prompted to write a motivation for their grading. In the last diary (after 20 hours of playtime), I asked the participants to score their opinion on specific game mechanics: the avatar’s abilities, if the game was repetitive, if the missions, enemies, and loot was varied interesting, and finally if the game content was making them feel uneasy or scared.

The purpose of the diaries was to gather a wide range of data to see if and how the participants changed their attitude as they progressed in the game. I call this method “Focused Player Diaries” since the design of the diaries was focused around the topics of the game that I was interested in measuring.

Follow-up Interviews
The participants’ comments were also used to build the core of the semi-structured follow-up interviews which were conducted after the participants had handed in their last game diary. The questions and statements in the diaries and follow-up interviews were all based on attributes presented in the Model of UX (Hassenzahl, 2003).

To ensure that we had a good understanding of the game, the researcher who performed the interviews played the game extensively during the study. This prolonged engagement with the game made it easier to discuss game-related topics during the post-study interviews and to relate to the content of the diaries.

To ensure credibility of the interview data, the researcher’s notes were read back to the participant, who would then be able to confirm or correct their answers.

Study 2:b
We used a qualitatively driven mixed methods approach, where quantitative methods supplement and improve the study’s results. The qualitative analysis is done through a manual aspect based sentiment analysis. The quantitative analysis was done through hypothesis testing using a Chi-square test.

Games used in Study 2:b

- Dragon Age: Origins (DA1) (Dragon Age:Origins, 2009)
- Dragon Age II (DA2) (Dragon Age II, 2011)
- Dragon Age: Inquisitions (DA3). (Dragon Age: Inquisition, 2014)

**Word frequency and selection**

The data collection for our ABSA was performed in the following steps. First, we collected all user reviews on the PC-version of the three games from the Dragon Age franchise: DA1, DA2, and DA3, from Metacritic. As mentioned in the Metacritic description in the background section, Metacritic authors rate their own reviews to reflect their experience of the game in question. This is a rating from 0 to 10, but in effect it will categorize the comment as one of three tiers: low, medium, or high rated. We decided to only work with the reviews of the PC-version (the games exist for multiple platforms) as it was the versions that we were familiar with.

For each game, we did a word frequency analysis, using AntConc 1, to find which aspect was most frequently used in the reviews. As we had no previous practice with this method in this context, the threshold was set after we saw the results – we decided to pursue the three most frequent explicit aspects that were shared by all three games. These explicit aspects were: *Story*, *Combat*, and *Character*. All reviews that did not contain any of the aspects were omitted from the dataset. As the reviews were rated by the authors, we already had the rating categories.

Since the review rating and the sentiment of the aspect may differ – for example, a high rating review may use an aspect in a negative way – it was important to collect all reviews of all ratings. Figures 2, 3 and 4 show the frequency of the aspects in relation to review rating. As can be seen, the aspects tend to be more frequent in low rated reviews than high and mid rated reviews.

As a result of the data collection, we had a dataset of reviews for each game, regarding the three aspects (story, combat, character). Each review was categorized into its original rating level.

So, in conclusion for this section:

- Aspects were determined through word frequency analysis of all the user reviews
- The three most frequent aspects were combat, story, character
- Each game had a number of reviews
- A review contains at least one of the aspects
- A review is rated as either low, medium, or high

---

1 AntConc, by Anthony (2012), is a freeware concordance and text analysis tool by Dr. Laurence Anthony at the Faculty of Science and Engineering at Waseda University, Japan (http://www.antlab.sci.waseda.ac.jp/index.html).
The dataset contains all reviews, sorted by game, rating, and aspect.

Sentiment Analysis
The sentiment analysis was performed through an online crowdsourcing service. The rating and name of the game were omitted for the evaluators to limit the risk of bias. The evaluators were asked to read a review, or excerpt of a review, which contained one of the aspects, and to determine if the author of the review had used the aspect in a positive, neutral, or negative way. The following quote is an example of an excerpt that the evaluators judged:

*The menus, crafting and combat are so totally and completely cumbersome. Everything is very statically organized and takes so much time. I spent an ungodly amount of hours collecting resources, crafting things, comparing items to what I already owned and it is just so, so, so cumbersome and tiresome, it really damages the game.*

The aspect of combat occurs in the quote, and the overall use of the aspect is considered negative. In total, 9358 review excerpts from all three games were analyzed this way, and each aspect was judged by at least three evaluators.

If an excerpt contained more than one aspect, it would be run again, through a second (or third) sentiment analysis, where that aspect would be in focus for the evaluator. When the sentiment analysis was done, the dataset was reconstructed with rating and game name.

The relevance of each of the three aspects was tested. The following null hypothesis was constructed: *there is no relationship between the values of aspect X (character, combat or story) and the overall review rating.*

---

2 [www.crowdflower.com](http://www.crowdflower.com): a data mining and crowdsourcing service where researchers can upload their data e.g., for manual sentiment analysis by anonymous evaluators.
Results

In this chapter, I will present the results from the separate studies and how these results are related to the research question, and sub-questions, investigated in this thesis. The individual papers were written with their own purpose, meaning that they are self-contained and can be read separately and independently. Therefore, it is not necessarily the original outcome for each paper that is listed as a contribution here. For example, the research design in Study 2:a is of greater interest in response to the RQ of the dissertation, than the results presented in the paper itself. Table 5 shows which papers that, in whole or in part, address the different research questions. In the sections following Table 5, each Study is described and discussed in relation to the relevant RQ.

Table 5: Overview of the papers included in the dissertation. Cells marked with Ω indicate that the paper has a relevant contribution to the research question. Cells marked with α indicate that the Study in some sense contributes to the solution.

<table>
<thead>
<tr>
<th>Study 1</th>
<th>RQ 1.1</th>
<th>RQ 1.2</th>
<th>RQ 1.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study 1:a</td>
<td>α</td>
<td>α</td>
<td></td>
</tr>
<tr>
<td>Study 1:b</td>
<td></td>
<td></td>
<td>Ω</td>
</tr>
<tr>
<td>Study 1:c</td>
<td></td>
<td></td>
<td>Ω</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Study 2</th>
<th>RQ 1.1</th>
<th>RQ 1.2</th>
<th>RQ 1.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study 2:a</td>
<td>Ω</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study 2:b</td>
<td>Ω</td>
<td></td>
<td>Ω</td>
</tr>
</tbody>
</table>

These are the research questions that were introduced in the introduction chapter. They are addressed in the following sections.

1. RQ 1: How can novel methods be used within the context of game user research?
   1.1. In what ways can UX theories be used within the context of game user research?
   1.2. In what ways can indirect user feedback be integrated into game user research?
   1.3. In what ways can longitudinal user feedback be integrated into game user research?
In what ways can existing UX theories be used within the context of game user research?

This question is supported by Study 2:a and Study 2:b, as well as Study 1:a to an extent. In the background chapter, I describe how UX has developed in non-game industries, and how industry and academy both contribute to create descriptive models for UX. I further discuss how game research established GUX as the need for a video game domain specific approach to UX was emerging. In this context, it is valuable to assume pre-existing UX theoretical models and apply them to game research situations to determine the degree to which they can be utilized.

Study 2:a

As mentioned in the Methods chapter, the study had seven participants; four male (M1...M4) and three female (F1...F3). Study 2:a used the model of UX (Figure 1) (Hassenzahl, 2003). By using the attributes from the model, I could extract the participant’s views and attitudes from a game specific perspective. Each attribute was used as a basis for interview questions and survey questions in the focused player diaries. The attributes are manipulation, stimulation, identification, and evocation (please see chapter Theoretical Framework, section Model of UX, for detailed descriptions of the attributes).

**Manipulation**
The participants felt that the user interface contained too much information and that “there is so much going on so I just focus on this one thing” (M2). The game offers many ways to optimize the play, such as improved health and armor, or better special abilities. However, the participants found these opportunities to be presented in a “cluttered” way: “I can’t be bothered to try and understand how this 2% improved this and 3% improved that works, I just want to get a better weapon” (M3) and “the amounts of skills are overwhelming, so it becomes hard to see the purpose and benefits – this makes it uninteresting to find optimal combinations” (F3). In short, by explicitly asking the participants about manipulation of the game, I could extract information that is essential to a game designer who wants to evaluate a prototype.

**Stimulation**
The survey showed that all but one of the participants felt less interested in playing, less excited about playing and more bored with the game at the 20-hour mark than at the 1- and 5-hour marks. One participant (F2) lowered her want-to-play-more score drastically – from 9/10 to 1/10. In the post-game interview, this participant said that the repetitiveness of the game made it an
uninteresting experience. As an example, the participant mentioned a situation where she had finished one mission and went to do a second one, which turned out to be exactly the same type of mission. As this was a recurring event, it eventually made the participant feel uninterested in the game: “This is tedious and cheerless; the rewards are rarely worth it. It is hard to keep motivated through this grind [repetitive gameplay]” (F2).

Most of the participants commented on the reward system. Gear and weapons that are rewarded after defeating a challenging enemy are randomized in both type and abilities. This makes the game too unpredictable according to the participants. They agree with the comment about overwhelming information on gear and weapons, which makes it even more difficult to plan ahead. This reward system forces the player to repeat easy missions in attempts to get “lucky rolls” on the rewards. As the participants already find some missions repetitive, this is an unsatisfying reward mechanism.

According to Hassenzahl (2003), stimulation can be achieved by challenging our senses, creating opportunities and preconditions for personal development and successful use of the artifact. Judging from the user scores and comments, the game is challenging, but sometimes in a way that is not meeting the user’s anticipation or satisfaction.

Identification
The participants did not identify as TD players, and they did not get a sense of community with other players, even though they played online with anonymous players. One reason for this was that there are too few options when creating the avatar.

Evocation
The game successfully evoked feelings of uneasiness, compassion and anxiety. One participant (F3) felt so strongly about some in-game scenes, that she left the screen for a while. Another participant (M1) explained that not only were the graphical scenes disturbing, but also the voice memos that the players found: “I listened to a recording of a conversation between a dying woman who anxiously said she did not want to die alone, and a 911 operator on the other side who was incapable to assist her or even offer any consolation. It affected me way more than I have ever experienced in any game before.” Overall, these powerful situations enforced the game narrative in a positive way, according to most of the participants.

However, the game’s overall approach of realism was overshadowed by unrealistic opponents and repetitive missions. In the interviews, all participants said that the game is repetitive, but not all found this to be a problem. One participant (F2) found the repetitiveness so problematic that she could not imagine ever playing the game again after the study had been completed.
On the topic of enemies, all participants believed they were unreasonably hard to defeat, not only from a game perspective, but also from a true-to-life perspective.

One participant (F1) mentioned a dissonance between the player actions and the character role: “The things you do in the game do not match the game narrative. I would rather have a shorter game but a tighter story. A super realistic world cannot contain this much grind [repetitive gameplay]. Also, the cleaners [an enemy type] are completely unrealistic. They are all insane people armed with flame throwers? This doesn’t fit the realistic context of the game!”

This sense of dissonance was paraphrased by other participants. One participant (M3) maintained that “[the game] is not believable. It’s like it tries to be a realistic story but the actions that the players do and the lack of realism of the combat (the fact that enemies can survive dozens of bullets for example) break this sense of realism. Suspension of disbelief is broken.”

**Conclusion Study 2:a**

From the analysis of the data, I created a list of generalizable slogans. The slogan attributes (Table 6) are from the Model of UX (Figure 1, *chapter UX theories and models*). The slogans could be attributed to stimulation and/or evocation attributes of the game. *Stimulation* relates to the game challenges and the ability to inspire exploration and task solving, to show the player what lies ahead, and encourage continuous play to reach the player’s goals. The game in Study 2:a failed to do this on several points; for example many of the participants brought up the reward system. Early in the game, they found the unpredictable reward system intriguing and interesting, but eventually it caused frustration rather than joy.

*Evocation*, in a game setting, is defined as the ability to evoke emotions regarding the game activity, such as joy, fear and satisfaction. The repetitive nature of the game activities and the overcomplicated game serve as an example of why the players felt less inclined to continue playing.

Table 6 is a result of using the UX theory in game user research, and as such contributes to RQ 1.1.
Table 6: "Slogans" from Study 2:a as a result of using the UX theory in game user research.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Retention breakdown description</th>
<th>Retention Slogan</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stimulation</td>
<td>Complicated game elements (talents, skills, gear bonus and penalties) make the player feel less committed – indicating poor stimulation and evocation.</td>
<td>Game - not Gear</td>
</tr>
<tr>
<td>Evocation</td>
<td>Reward systems should be accessible and/or predictable. If the player feels that the reward is both inaccessible and unpredictable, they feel less inclined to continue playing and exploring the game’s opportunities for stimulation.</td>
<td>Reward through Strategy – not Randomness</td>
</tr>
<tr>
<td>Stimulation</td>
<td>Avoid repetitive missions. Missions lacking in the attributes stimulation and evocation lead to lowered sense of appeal.</td>
<td>Challenge – not Chagrin</td>
</tr>
<tr>
<td>Evocation</td>
<td>Mismatch between actions performed in game play and setting. Player actions and computer controlled entities’ behavior should match game’s narrative setting.</td>
<td>Relevant Realism – not Snake-oil-salesmen</td>
</tr>
</tbody>
</table>

The retention breakdown description explains how game elements can disrupt the users’ GUX, and cause a negative experience. The connection between the UX attributes and GUX is clear. The possibility of both collecting written data, preparing interviews and finally analyzing makes the framework an important tool in this study. By using the model as an analytical framework, and by using the perspective of the theoretical UX model, I was able to discuss the game with the participants from the perspectives of Manipulation, Evocation, Identification, and Stimulation. The attributes helped both me and the participants to verbalize and concretize their thoughts about the game. This framework became especially useful as the play time progressed and the participants’ attitude towards the game changed.

Study 2:b

In Study 2:b, I used Janlert and Stolterman’s theory on “Character of things” (1997) as an explanatory model for users’ dissatisfaction with the three games involved. I discovered that the users focused on three major game aspects and how the user review rating matched the sentiment of the game aspect: i.e. a
low rated review would present an aspect in a negative sentiment and vice versa. Furthermore, I found that the same aspects in all three games were rated differently in each game, for example in the first game in the Dragon Age series, most of the reviews containing the “combat” aspect were positive, and used the aspect in a positive sentiment. In the second installment of the game, the opposite was found: most of the reviews and sentiments regarding the “combat” aspect were negative.

According to Janlert and Stolterman’s (1997) concept of RoC, too many alterations in a known or expected artifact will go against the user repertoire, and put the artifact’s popularity at risk. Users can, to some degree, predict or anticipate the behavior of the artifact, based on the perceived character and the context where the artifact is meant to be used. When the designer introduces a new feature, s/he must be careful not to take too big a leap from the familiar. Doing this may be perilous as the users feel that the artifact is too disparate from their repertoire of expected characteristics of the artifact.

The frequency of the aspects implies that they are important to the users, and thus a part of their RoC – this implies that the low rated review authors are disappointed with the aspects in the games. The theory of Character of Things proved very useful as an analysis framework, and thus I can claim that it provides a contribution to RQ 1.1.

In what ways can indirect user feedback be integrated into game user research?

By analyzing indirect user feedback in the form of user created reviews, I found critical issues with the games involved in the studies, without having any users present. Both Study 1:b and Study 2:b used user created reviews as a data source.

Study 1:b & 2:b

In Study 1:b, I used user created game reviews to find what the users commented on, and analyzed these comments with a set of video game specific heuristics. The users experienced issues that the heuristics were not designed to measure. In Study 2:b, I continued to research user created reviews by gathering all reviews from all three games of a game series and through a word frequency analysis, I produced three main aspects that were discussed by the authors of the reviews.

The analysis from Study 2:b showed that the frequency of the aspects increased in the low rated reviews. Study 2:b showed that users tend to comment on aspects they have strong feelings about, and the sentiment of these feelings
will comprise the overall rating of the game. This was statistically significant using a chi-2 test (please see table 7 and table 8).

Study 2: examined the null hypothesis: there is no relationship between the values of aspect X (character, combat or story) and the overall review rating.

Aspects used in a negative sentiment occurred in low rated reviews, implying that a user who is discontented with an aspect often is discontented with the entire game, which in turn implies that the aspects are relevant measures for the game. The method used to show this should be considered as an evaluation technique that can be used as a complement for a game developer during an early research phase in a production project.

Table 7: Aspects distributed on review ratings, for all three games in the franchise. The numbers are from the evaluators’ sentiment analysis, and show the frequency of each entry.

<table>
<thead>
<tr>
<th>Aspect</th>
<th>Low</th>
<th>Mid</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Char.</td>
<td>bad</td>
<td>633</td>
<td>257</td>
</tr>
<tr>
<td></td>
<td>neutral</td>
<td>1038</td>
<td>72</td>
</tr>
<tr>
<td></td>
<td>good</td>
<td>68</td>
<td>148</td>
</tr>
<tr>
<td>Comb.</td>
<td>bad</td>
<td>520</td>
<td>211</td>
</tr>
<tr>
<td></td>
<td>neutral</td>
<td>358</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>good</td>
<td>43</td>
<td>83</td>
</tr>
<tr>
<td>Story</td>
<td>bad</td>
<td>993</td>
<td>278</td>
</tr>
<tr>
<td></td>
<td>neutral</td>
<td>1056</td>
<td>119</td>
</tr>
<tr>
<td></td>
<td>good</td>
<td>72</td>
<td>142</td>
</tr>
<tr>
<td></td>
<td>4781</td>
<td>1360</td>
<td>2127</td>
</tr>
</tbody>
</table>

Using the chi-2 test, the following values were obtained (Table 8):
Table 8 Chi-square values for each aspect, of all three games in the franchise

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Aspect</td>
<td>Chi square</td>
<td></td>
</tr>
<tr>
<td>DA1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Character</td>
<td>120.2</td>
<td></td>
</tr>
<tr>
<td>Combat</td>
<td>100.4</td>
<td></td>
</tr>
<tr>
<td>Story</td>
<td>196.6</td>
<td></td>
</tr>
<tr>
<td>DA2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Character</td>
<td>304.9</td>
<td></td>
</tr>
<tr>
<td>Combat</td>
<td>299.6</td>
<td></td>
</tr>
<tr>
<td>Story</td>
<td>426.6</td>
<td></td>
</tr>
<tr>
<td>DA3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Character</td>
<td>1072.5</td>
<td></td>
</tr>
<tr>
<td>Combat</td>
<td>374.4</td>
<td></td>
</tr>
<tr>
<td>Story</td>
<td>1250.2</td>
<td></td>
</tr>
<tr>
<td>DA series</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Character</td>
<td>1541.3</td>
<td></td>
</tr>
<tr>
<td>Combat</td>
<td>813.8</td>
<td></td>
</tr>
<tr>
<td>Story</td>
<td>1963.4</td>
<td></td>
</tr>
</tbody>
</table>

All values exceeded the threshold at p= 0.001 and 4 degrees of freedom thus in all cases the null hypothesis can be rejected.

In what ways can longitudinal user feedback be integrated into game user research

The purpose of Study 2:a was to see if and how users’ attitude changes over time, and to find the cause for this change. The longitudinal approach proved fruitful, as the participants showed a positive attitude towards the game before they played it, and after the first hour. At 5- and 20-hour time points, attitudes had changed across the entire group. Through the use of a longitudinal study and focused player diaries, I found issues with the game that a single session may have missed.

Pre-play Impressions

The participants’ reactions after viewing the official game play video were unanimously positive. I interpreted their comments “It will be interesting to see how they integrate story driven gameplay with mmo[ massive multiplayer online game] emergence, new narrative mechanics etc.,” “The graphics are visually stunning,” “…the concept feels new and cool,” and “The cooperative game play looks like great fun!” as positive, and that all participants were interested in playing the game.
Graphical Representation of Attitudes

The following graphs show how the participants’ attitude towards the game changed over time. Figure 2 shows their willingness to play more, Figure 3 shows the sense of excitement, and Figure 4 shows the sense of boredom the participants felt. The graphs were the basis for the follow-up interviews, presented in the section Playing Experiences.
Figures 2, 3, and 4: The graphs show willingness to play more (top graph), changes in sense of excitement (second graph), and boredom (bottom graph) the participants feel when playing the game after 1, 5 and 20 hours.
Through the longitudinal study, using the focused player diaries and pre- and post-interviews, I can visualize the changes in users’ attitudes. Game user research performed only on a singular session without consideration of change over time will most likely miss the retention breakdowns as described in the slogan table (Table 6), and it will become more difficult to understand the root cause of user disengagement.

Concluding the Results Chapter

The user experience models that I have used have mainly contributed to the study design. The models show which questions I need to ask the participants, to capture their perspectives and attitudes. The longitudinal approach also opens for an increased understanding of how these attitudes change over time, and how the design of a game should be prepared for this change. This result is illustrated in the slogans (table 6), which are supposed to serve as a template or ruleset for the game developer. Furthermore, the value of the focused player diaries is the visualization of the development of the participant’s Repertoire of Character (Janlert & Stolterman, 1994).

The indirect feedback that can be read from user created reviews shows that this type of data, and source, is of great value to a developer who critically wants to research the genres they want to develop within. Although the method as such does not guarantee a successful game, it may help the developer to avoid the same mistakes as previous versions. The method is fairly light-weight, can be performed by just one or two developers, and the results are – as shown – valuable.
Discussion and Conclusion

In the background chapter, I established that GUX is considered a fundamental approach to video game design, and that knowing the target user group is fundamental to successful GUX design. Recall Hassenzahl’s model (Figure 1) describing the users’ situation. Situation is, according to the model, established by the users’ RoC (Janlert & Stolterman, 1997). Thus, to optimize the GUX design, a designer can familiarize themselves with the repertoire, or prepare the target group by shaping their repertoire through commercial channels and design accordingly.

Study 2:b illustrated the effects of a mismatch of user repertoire and developers’ GUX design. While it may be too late to repair the game-franchise presented in Study 2:b, important lessons can be drawn from the less-than-positive user reviews. The repertoire can be recovered from the analysis of the reviews (as we did in Study 2:b), and the results can be used to gain a better insight into the repertoire.

If the repertoire is shaped by commercial channels, such as in game footage, pre-release videos, concept art and the like, the designer must still be attentive to the outcome. Does the resulting game match the signals they send out? In Study 2:a, we found a poor match, as the participants felt frustrated and bored with the game, leading to termination of game engagement.

A successful video game release is worth more than sales numbers. If the product can meet the expectations of the users, it is likely to improve the developer’s reputation, leading to other benefits, such as improved potential for cooperation with other developers, increased staff motivation, easier to recruit special competences and much more. A reputation is a fragile thing: it takes time and effort to build, but can very easily be ruined by a single misstep. A good reputation requires constant vigilance, which in turn can be costly, so any method that can lower the cost of development should be explored. The chance for a successful release improves if the developer keeps the users’ situation and Repertoire of Character in mind throughout the whole development process, but this requires, in turn, close and regular contact with the end users.
The Birth of the Playsona

A common method for representing the end user during the design process of an artifact is to create Personas (Pruitt & Adlin, 2010) – invented, imaginary individuals with abilities and features that are intended to represent the real-life end-users. Personas help design teams to personify the user group rather than view the users as a generic mass of people. This personification helps the designers to think of their design from the perspective of each of their Personas, which in turn leads to a more user focused design.

Personas can be created in many different ways, but most methods share the same basic steps. Data is collected from the target group/groups, and categorized in a way similar to content analysis to find patterns of behavior and personality. The result is used to form archetypes, which are then fleshed out to more “real” personalities. The process of developing Personas is rather extensive; data is gathered on essential user groups, product stakeholder demands are taken into consideration, collected data is analyzed, subcategories are identified and evaluated, etc. A process of such proportions can have a deterrent effect on design teams, due to cost and time investments. It is easy to imagine that short cuts are made, and that all Personas that the industry is using are not fully developed according to due process. Nonetheless, Personas are powerful design tools that help the design team to keep the users’ needs in mind throughout the development process. Bearing this in mind, my work suggests a Persona that can be created more easily and on-the-fly, still using data directly obtained from the users: a Playsona.

The Playsona (a portmanteau from player and Persona) encompasses the aggregated opinions about aspects that are collected through an ABSA method, as presented in Study 2:b. A Playsona can be created for any game genre, provided that the researcher has access to a dataset of user created reviews or opinions. The utility value of the Playsona should be obvious – to know which aspects of the game genre that the users deem important, and to know the sentiment that the users feel about the implementation of these aspects in previous productions. This knowledge is crucial for creating new and engaging content and to avoid mistakes from previous games.

Creating the Playsona

The developers have an idea of what kind of game they want to produce, and an idea of their users’ expectations. We can consider this stage as an instance of the Intended Product Character (IPC), from the Model of User Experience (Hassenzahl, 2003) (please see chapter Theoretical Framework, section Model of User Experience, figure 1). The developer’s vision is encompassed in the IPC, and by extension, so is the developer’s idea of the intended user.
We can now create a model similar to the Model of User Experience but at a lower level of complexity, where the IPC is replaced with the intended user’s expectations (IUE). The developer must ensure that the IUE is in conformity with the actual user’s expectations (AUE). This can be achieved by performing the Playsona process (figure 5a & 5b).

By collecting reviews and performing a word frequency analysis, much of the work is done. A first glance at the results from a data collection and word frequency analysis, as performed in Study 2:b, will give important clues. By selecting aspects based on the frequency of appearance in the data, and in what estimated sentiment they are used, the developer can form their Playsona.

For example, from a data set, we might find that the aspect of **Combat** occurs with a high frequency in mostly low rated reviews, and that the aspect of **Character** occurs with a high frequency in high rated reviews. This tells us that it is probable that the user (the AU) enjoys aspects of the character in the game, but dislikes aspects of combat. If the developers familiarize themselves with the game (for example by playing it), and read samples of reviews, it is possible to create a Playsona that reflects the data. If more aspects are collected and matched, the Playsona will be more fine-tuned. The Playsona should then be used in the same manner as the original Persona method, to keep the end user in mind during the design process.

An advantage with the Playsona method is that it is relatively quick to implement. One person can extract multiple aspects and create several Playsona in less than a work day. For the developer who wishes to have a quick overview, it is a convenient and practical option. The disadvantage is that the method is not guaranteed to be valid. To be sure that the data is correctly analyzed, or academically sound, the sentiment analysis, the ABSA, step should be performed, followed by an appropriate statistical evaluation method (in study 2:b a chi square test was used for this).

Considering the above, I claim that the Playsona is a step in the right direction towards the main RQ of this thesis. It is my firm belief that the game industry needs quicker, light weight, easy-access tools to improve their productions, both from a producer point of view, and from the view of the end user.
Figure 5a: The developer must ensure that the intended user’s expectation is in conformity with the actual user’s expectations. The Playsona process (Figure 5b) may be used to achieve this.

Playsona Process

Figure 5b: The Playsona is created from aspects derived from user reviews of similar games. The developer combines suitable aspects into a Playsona. The purpose is to help the developer to focus on the right things.
Future Work

In conclusion, my contributions presented in this dissertation are the following:

- By applying the analytical framework of a UX model, I was able to isolate a GUX-relevant set of retention mechanics.

- Longitudinal data collected through focused player diaries and interviews provides information on why and when users “drop off.”

- Indirect user feedback is a valuable tool to be added to the GUX toolkit. The semi-automated way of using online user reviews is a lightweight tool for GUX input in the design process. It will indicate, for instance, what game characteristics are essential for customer satisfaction.

- Indirect user feedback may fill the gap between expert driven evaluations and user opinions.

I claim that some tools available for developers are lacking in relation to GUX and user needs. This is not to say that these tools are without merit. As shown in Study 1:c, game interface was not mentioned unless they were considered poor. Failure to adhere to video game specific heuristics, that cover such issues, will most certainly cause more problems and even more disgruntled users. In fact, many users will not even reach a state of discontent with the actual game, as they will terminate their engagement due to poorly designed user interfaces. But once the basic needs are met, the “real” user experience starts.

My suggested use of UX theories and data collection- and analysis techniques can produce lightweight tools for GUX research and game designers. These qualitative oriented techniques can be used to gain an in-depth look at drop-off rates and retention mechanisms. The results of the research also point to a potential use of UX tools such as personas which can be adapted to GUX based on lightweight tools. The Playsona model must be tested under real circumstances. It may, on a theoretical level, seem like a convenient method, but its value cannot be truly known until it has been tested.
in a real design case. Such a baptism of fire is necessary to enhance and improve the model, and a future study should apply the model to various stages of the design process; the model may be useful in an early stage, but less useful in the later stages etc. The many different genres of video games may also present challenges to the model. The suggested future studies for the Playsona will thus be: Testing in a real environment, testing on various stages in the design process, testing on different genres. Furthermore, the method of extracting user created reviews – and other types of comments – must be further developed. Many forums exist for the users to express themselves, and this data should be extracted and put to use!

A potential drawback of the sentiment analysis method is that it requires a previously released and reviewed game. This requirement, however, is unlikely to be a hindrance, as the market is ever expanding, and while genres do mix, not many new genres see the light of day.

Focused Player Diaries was a very convenient method for me as a researcher, and it may be a viable method for a game developer. A developer can use user diary entries on a smaller scale as an inspirational starting point for creating interview questions for focus groups and other test situations. The data collected through the diaries captures the participants’ thoughts very close in time to the actual experience, which is why they have an advantage over interviews, where the respondents are forced to re-think about their experiences. The diary notes give rise to new questions in subsequent interviews, and it is easy for the researcher to follow the respondent's development over time. A semi-structured interview based on the diaries can be expanded over multiple diaries from all the participants, giving both width and depth to the interview data collection. Study 2:a could also have benefitted from at least one more interview session. We conducted interviews prior to the first diary entries, and after the last 20 hours of game time. A mid-time interview, at 5 or 10 hours, would have been interesting, since the attitudes started to shift at the 5-hour mark for most of the participants.

Aside from conducting additional interviews, the diary itself should be explored; can it be streamlined so that it works as an “off-the-shelf” solution for game research? Study 2:a should be viewed as a prototype model, from which we can keep developing the Focused Player Diaries into a complete model, applicable for all kinds of games.

In closing, two major models of GUX research are presented in this thesis. They are not complete, but by working with them I have realized that they both have great potential to become valuable tools for game research communities and game developers.
Papers

Studies included in the dissertation
In this chapter, I present the studies comprising this dissertation, with a more in-depth description of each paper, its main research question, and the relevant contributions.

Study 1

Study 1:a

Study 1:b
Stråät, B., & Verhagen, H. (2014). Vox Populi–A Case Study of User Comments on Contemporary Video Games in Relation to Video Game Heuristics. In *Game on, 9-11 September 2014, University of Lincoln, Lincoln, United Kingdom* (pp. 5-9). University of Lincoln, United Kingdom.

Study 1:c

Study 2

Study 2:a
Study 2:b

Study 1


This paper was written by me, Johansson, and Rutz (Strååt et al., 2015). We wanted to see if our previously crafted Net Heuristic List was appropriate for the task of evaluating games, and we wanted to see if there was a connection between violation of game heuristics and game rating. A total of 10 games were evaluated according to the heuristic evaluation method as described by Nielsen (1995). The video games were selected based on their popularity rating on metacritic.com: low rated games had a score of 40/100 or less and high rated had a score of 80/100 or more. We found that the low rated games violated more heuristics than the high rated games, and that some of the issues we found were not represented as heuristics at all. As a result, we could tentatively suggest two new heuristics to be added to the Net Heuristic List.

**Research Question:** First - evaluate the Net Heuristic List. We sought to know whether our selection of heuristics could be used, productively, to measure game play, according to our definition, and to establish whether new heuristics can be developed.

Second - examine whether there is a connection between the assessed quality of a video game and the heuristics violated in the game, and, accordingly, whether equally rated games violate the same heuristics.

**Contributions:** Two new heuristics were suggested: *False Affordance* and *Lack of Meaningful Play*. False Affordance occurs when the Gameworld affords interaction in a certain way (for example, the graphics of a wall makes the player think they can climb it) but this interaction is not possible. Lack of Meaningful play occurs either when choices the user makes have no impact
on the outcome, or when the user is forced to complete game stages that contain no value.

The study also showed that the low rated games violated game usability heuristics much more than high rated games. The results provide valuable information towards answering the first question of this thesis. It showed common and shared problems within the different games. It also showed that low budget games can easily alleviate many video game issues by adhering to the Net Heuristic List, which in turn indicates that the Net Heuristic List is a valid tool.

Study 1:b: ‘VOX POPULI - A Case Study of User Comments on Contemporary Video Games in Relation to Video Game Heuristics’

This study, written by me and Verhagen (Strååt & Verhagen, 2014), focuses on the users’ ideas on game design issues. We gathered all user reviews at metacritic.com, from three of the games that were used in Study 1:a (Strååt et al., 2015). One high, one medium, and one low ranked game were selected. We designed word categories to describe the games, and performed a word frequency analysis on the user reviews. The categories were ‘Adjective or Adverb,’ ‘Genre Description,’ and ‘Game Specific.’ The words were then viewed in the context sentences, which in turn were analyzed with the Net Heuristic List, where we examined if it was possible to interpret a sentence to match any of the heuristics.

The results showed that the highly-ranked game received critiques on the storyline and poor optimization of the video software. The middle-ranked game received critiques on the game play and storyline, while the lowest ranking game got most of its critiques on game play, interaction elements, and boring tasks. These results might indicate that game play and interaction design is important, which in turn correlates with the findings of Study 1:a.

Research Question: How do players express themselves when they review a game, and can their opinion be tied to the heuristics of the Net Heuristic List?

Contributions: The results showed that the players focused on issues with storyline and immersion. These aspects are covered in the Net Heuristic List, but it was interesting to see that almost no other aspects of the included games were discussed. This indicates that the view of the experts (the top-down ap-
proach) does not coincide well with issues that the users find relevant. Moreover, users rarely mentioned usability aspects. Comments about controls were predominantly negative, which indicates that players mention game controls when they are poorly implemented. If the controls are good, the players do not seem to mention them at all.

Study 1:c: ‘Top Versus Bottom: Game Evaluation from an Expert or Player Perspective’

This is a theoretical discussion that connects Study 1:a and Study 1:b and works as a segue into Study 2. It was written by me and Verhagen (Strååt & Verhagen, 2015).

**Research Question:** This paper examined the extent to which heuristics developed by experts can address the issues that the end users experience, what kinds of design issues are most frequently mentioned, and whether these issues are reflected in online user comments.

**Contribution:** The results showed that users focused their criticism on narrative and aesthetics rather than the issues that the heuristics were not designed to address. This raised the question of how to determine the users’ most frequent issues, leading to Study 2.

Study 2

Study 2:a: ‘Exploring Video Game Retention Mechanics – A Longitudinal Case Study’

This paper was written by me and Verhagen (Strååt & Verhagen, 2017). It explores how users’ attitudes changed over time when playing a newly released video game. The participants of the study were interviewed during the recruitment phase to see if they were interested in the game involved, prior to its release. The game can be played cooperatively, competitively, or solo. The game genre is best described as action first person shooter, with role playing elements. In the game, players solve missions and fight enemies. The overarching narrative is that a virus has caused chaos in New York, and the players are part of a special team that must find a cure. On the release day, the participants were given the game with instructions to play it just as if they had bought it themselves. They were also instructed to keep a game diary where they should write down their thoughts and answer a survey. The diary was to be completed during certain intervals: after 1 hour, after 5 hours and after 20
hours of game time. After the 20-hour diary, the participants were interviewed about their experiences with the game. The interview questions were based on all the diary entries from all the participants. After the data was collected, it was analyzed using content analysis methods.

**Research Question:** How does player attitude change over time, and what are the major causes of attitude change?

**Contributions:** The results showed that most participants started to lose interest in the game after the 5-hour mark. This was mostly due to boredom (e.g., the missions were repetitive), frustration (e.g., the user interface was cluttered and confusing), and unpredictability (e.g., the reward system was non-transparent and too random). The final analysis result could be generalized to a set of four “slogans” that describe the issues that the participants regarded as the strongest deterrent factors.

**Study 2:** ‘Probing User Opinions in an Indirect Way: An Aspect Based Sentiment Analysis of Game Reviews’

This paper was written by me, Verhagen, and Warpefelt (Strååt, Verhagen, & Warpefelt, 2017). The goal of this study was to see if users’ sentiments differ between games that are released in a series. To this end, we decided to examine the user reviews of the game series “Dragon Age.” At the time of the study, Dragon Age had three installments: Dragon Age: Origin (DA1) (2009), Dragon Age 2 (DA2) (2011), and Dragon Age: Inquisition (DA3) (2014).

Data was collected by gathering all user reviews from Metacritic.com. Data analysis was done through word frequency analysis, aspect based sentiment analysis (ABSA), and statistical Chi-square analysis. From the word frequency analysis, the three most frequent words related to the game or gameplay were selected. The aspects were “combat,” “story,” and “character.” These aspects were submitted to an ABSA, through an online service. In total, 8268 user reviews or excerpts of reviews containing the aspects were evaluated for sentiment.

**Research Question:** If we select the most frequent aspects that the users seem to find important as discussion topics, which of these are relevant for post judgment of a video game?

- **Null hypothesis:** There is no relation between sentiment of aspects and ratings of a game

**Contributions:** The result showed that the users used the aspects in a sentiment – negative, neutral, positive – that matched their rating of the game – low, medium, high. This implies that the aspects reflect the users’ sentiment towards the game, thus, the method can be used to find problem areas within a game series.
Other publications by Björn Strååt

**Licentiate Thesis**
Strååt, B. (2015). TOP-DOWN DESIGNERS vs BOTTOM-UP USERS
Limitations of Heuristic Evaluations on Games
SU Report Series No. 15-007, ISSN 1101-8526
Department of Computer and Systems Sciences,
Stockholm University, April 1, 2015, Kista

**Conference papers, reviewed**


Artman, H., Karlgren, K., Ramberg, R., & Strååt, B. (2012). Designing Interaction in Interaction Design: Using interactionaries in order to understand student use of interaction design concepts. In *DESIGNS FOR LEARNING*
2012, 3rd International Conference Exploring Learning Environments; Copenhagen, Denmark, 25-27 April 2012 (pp. 14-16). Aalborg University.
References


Dragon Age II. (2011). BioWare.


64


Strååt, B., & Verhagen, H. (2014). Vox Populi–A Case Study of User Comments on Contemporary Video Games in Relation to Video Game Heuristics. In Game on, 9-11 September 2014, University of Lincoln, Lincoln, United Kingdom (pp. 5-9). University of Lincoln, United Kingdom.


66


