UNIVERSITY OF CALIFORNIA,
IRVINE

Challenge and Retention in Games

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Informatics

by

Thomas Debeauvais

Dissertation Committee:
Professor Cristina V. Lopes, Chair
Professor Gary Olson
Assistant Professor Joshua Tanenbaum

2016

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to thank the National Science Foundation for partial support during my doctoral program. Even though not part of this thesis, I enjoyed working with my advisor and my colleagues on other projects funded by the NSF.

CURRICULUM VITAE

Thomas Debeauvais

EDUCATION

Doctor of Philosophy in Information and Computer Science **2016**
University name *City, State*

Master of Science in Information and Computer Science **2012**
Another university name *City, State*

ABSTRACT OF THE DISSERTATION

Challenge and Retention in Games

By

Thomas Debeauvais

Doctor of Philosophy in Informatics

University of California, Irvine, 2016

Professor Cristina V. Lopes, Chair

Game designers and researchers agree that the main motivation for starting playing a game is challenge. It is only a small step to say that when the game becomes too difficult, players can become frustrated and quit. While extensive work shows that challenge is central in player enjoyment, its influence on player retention has received little attention. Establishing this influence is difficult for several reasons: (i) Definitions of challenge tend to be game- or genre-specific. (ii) Measuring retention accurately requires real-life data, which is often hard to access. (iii) Challenge can be intertwined with other factors, such as social interactions in multi-player games. Addressing these three limitations requires a mixed-method, holistic, and cross-game approach paying close attention to the game mechanics. Therefore, we use a diverse mix of data sources, ranging from interviews to online questionnaires and telemetry logs, to explore player behavior in five games: Ragnarok Online, World of Warcraft, Forza Motorsport 4, Jelly Splash, and an undisclosed commercial mobile game.

This work provides empirical evidence that challenge does influence retention, and that many contextual factors, namely player segments, social interactions, time, and money, nuance and mediate this influence. These findings may help game developers make better games, expand our knowledge of challenge, retention, and their relationship, and improve the bottom-line of game companies by making games more engaging and players stay longer.

Chapter 1

Introduction

The past decade has seen the rise of three concurrent practices in the game industry. First, games user research, the branch of user experience research applied to games, has matured and is now standard among large game developers and publishers. Games user researchers study what players enjoy or dislike in a game. Their tools range from qualitative focus groups and interviews to quantitative surveys and telemetry data analysis [67]. A second phenomenon that has seen its way into the game industry is big data. More and more data is collected about how people play and what they do inside games. The data collected is also becoming increasingly complex and diverse: from eye tracking to diary studies, A-B tests, and physiological data, all triangulating player enjoyment [135]. A third phenomenon is the recent major shift in the typical game's genre, audience, lifecycle, platform, and business model. The 1990's and 2000's were the (second) golden era for video game consoles [47]. At the time, most games were released once and for all, in a fire-and-forget manner[1]. Game prices were somewhat standard, and no game was given to players for free. By 2009, Facebook had reached a quarter billion users, offering a new platform for game developers to tap into a new and more casual audience [73]. By the early 2010s, social games like FarmVille were

[1]Exceptions include long-lasting franchises such as Madden Football or Final Fantasy.

played by tens of millions of players daily [85]. Successful games are now expected to start new or continue existing franchises [115] and release new content through regular patches.

1.1 Motivation

In this context, there has been an increased focus on player retention, i.e. how long players engage with the game, and when and why they drop out (or churn, as it is also called). Studying retention and its counterpart, churn, is interesting and valuable because they are a proxy for player enjoyment: when players like the game, they keep playing; when they stop liking it, they stop playing and churn [110].

Game designers are interested in how they can build more engaging games with better retention. One knob in their control is the game's challenge level [3, 111]. Best practices recommend making games that are easy to understand, progressively more difficult, and ultimately hard to master [98]. The assumption is that if the game is too difficult, players may become frustrated and churn. If the game is too easy, they may become bored and also churn [80].

This assumption is reasonable, but does it always hold empirically? After all, some players enjoy games that are incredibly difficult [74]. Others have fun playing games that are sometimes trivially simple just to pass time or socialize [73]. And although designers know better than players what is best in a game, they sometimes make mistakes [29]. They also have certain tastes, and tend to only design and play games that fit their tastes [80]. So designers may not always understand the appeal of certain game mechanics on certain segments of players [17]. That is where academic games user research can help study niches of players with particular tastes and expectations [37, 92, 101, 121]. While these academic studies focus on the motivations of players, they ignore or at least stay vague on game design particulars.

In the end, the relationship between challenge and retention does not seem straightforward. Improving our understanding of this relationship may help game designers build better games, expand our knowledge of challenge, retention, and the influence of one on the other, and improve the bottom-line of game companies by making players stay longer.

1.2 Thesis and Research Questions

In this thesis, I claim that *challenge influences retention, but this influence is highly contextual.* The goal of this thesis is to empirically explore the influence of challenge on retention by highlighting contexts and situations in which this influence is manifest or surprisingly missing. In the process of conducting this work, I focus on the following research questions:

(1) Is the influence of challenge on retention observed empirically? As will be shown in the rest of the dissertation, the first-order approximation is yes. However, defining challenge and churn is not trivial, and heavily game-specific. Once adequate metrics of challenge and churn are found, correlating them is more straightforward.

(2) What factors mediate the influence of challenge on retention? To answer this question both broadly and in depth, I gather the results of different methodological approaches applied to several games of various genres. Consistent with a wide range of literature covered in Chapter 2, I find that social interactions, time, money, and player segments often mediate, nuance, and sometimes even eclipse, the influence of challenge on retention.

1.3 Approach

My work comprises seven studies using a total of three different techniques on five different games. By applying different techniques to study these games, I am able to find more ways in which challenge and retention relate, and I am able to identify more factors and contexts mediating this relationship. In each study, I place a strong emphasis on comparing the empirical behavior of players to the intended purpose of a game's mechanics. In the remaining of this section, I list the five games investigated in my work. Each of the games listed below is covered in its own chapter later in this thesis.

The first study, detailed in Chapter 3, concerns private servers of a massively multiplayer online game (MMO) called Ragnarok Online (RO). Although these private servers are illegal, thousands of players would rather play on them than on the official servers. Do these players prefer private servers because they are easier? Through qualitative interviews and participant observation, I find that difficulty plays a role on player enjoyment and retention, but so do social interactions and the money that players are ready to spend into the game. I also provide game mechanics introduced exclusively on private servers to alleviate, if not completely address, the problems most commonly raised by players on official servers.

The second, third, and fourth studies are combined in Chapter 4 and relate to World of Warcraft (WOW), another MMO. As of 2016, WOW remains one of the most popular MMOs, and a lot of academic research covers it [55, 91, 92, 96, 100, 133]. With 10 million players in 2012 [1], WOW gathered players from around the world, across genders, and of all ages [109, 131]. Do different demographic segments play and churn differently? Challenge in WOW revolves around grinding for game gold, but players sometimes (illegally) buy game gold with real money. Which segments of players buy gold most? In MMOs, the lack of new content is also said to be the main cause of churn [89]. How fast do players consume new content once it is released? I answer these questions by analyzing the data from two online questionnaires answered by several thousand WOW players (to be able to segment players demographically), as well as in-game player behavior data collected over seven months (to measure actual churn, and not just the breaks that players report in the questionnaire).

The fifth study looks at progression and how players adapt to the game's challenge over time in a racing game called Forza Motorsport 4 (F4). This study, presented in Chapter 5, is particularly relevant to my thesis because in the game, players can manually configure the amount of challenge they are ready to face. As for the approach, I analyze millions of logs from telemetry data collected over two years. Large-scale data mining is a very appropriate technique to study retention because it measures churn objectively, reliably, and unobtrusively [53].

The sixth study, covered in Chapter 6, focuses on the influence of challenge on churn in the mobile tile-matching game Jelly Splash (JS). Its main question is the core of this thesis: does difficulty make players churn? The approach consists, again, in mining millions of logs from telemetry data for the reasons mentioned above. This study also introduces a process and metrics for quantifying challenge and players' perception of challenge from in-game telemetry data in level-based games.

The seventh and last study, detailed in Chapter 7, centers on the tutorial of an undisclosed commercial mobile game (U) installed by millions of players. Since the literature reports that tutorials can influence long-term retention [5, 28], the goal of this study is to observe player churn before and during the tutorial. By mining data from a dozen versions released throughout the game's soft launch, I am able to post-hoc A-B test different aspects of the game and see how they influence retention. I find that time, and not a particular game mechanic, can explain nearly 90% of the churn happening during the tutorial.

These seven studies result in a long and heterogeneous list of empirical findings, graphs, behaviors, and player segments related to challenge, retention, or both. To aggregate and summarize items from this list, I build an affinity diagram. The goal is to surface practical lessons learned that may be applicable to a wide range of games. These lessons learned are presented in Chapter 8.

Game	N	Methods
Ragnarok Online (RO)	9	interviews
	30+	participant observation
World of Warcraft (WOW)	2,865	online questionnaire
	1,350	online questionnaire paired with game data
Forza Motorsport 4 (F4)	220,000+	telemetry data
Jelly Splash (JS)	274,000+	telemetry data
Undisclosed game (U)	78,000+	telemetry data

Table 1.1: Summary of the dissertation data sources.

Game	Challenge	Challenge Metric	Retention Metric
RO	Acquiring equipment	Player reports	Player reports
	Boss monsters		
WOW	Boss monsters	Bosses killed	Weekly play time, breaks
	Amassing gold	Buying gold	Weekly play time, breaks
F4	Driving	Disabling assists	NA
JS	Passing a level	Level difficulty	Churn
U	Winning a level	Level difficulty	Churn

Table 1.2: Listing of challenge and retention metrics used in the dissertation.

1.4 Contributions

My thesis makes the following contributions:

(i) It is the first systematic, holistic, design-centric, large-scale study of the influence of challenge on retention.

(ii) It shows that time can sometimes be a stronger predictor of churn than challenge.

(iii) It provides methods and protocols to define, analyze, and understand challenge and retention in a range of games.

(iv) It catalogs dozens of retention patterns across games of different genres.

(v) The literature agrees that many factors contribute to making a game engaging. Yet when looking at players, few works go beyond a binary engaged/not engaged label. I show that engagement can be broken down in two distinct dimensions, intensity and consistency, and detail four behavioral archetypes illustrating these dimensions.

1.5 Organization of the Dissertation

This dissertation is organized as follows. In Chapter 2, I break down the questions asked in my work into three dimensions, namely player enjoyment, retention, and in-game behavior, and provide an overview of the answers found in the literature to these questions. Chapters 3, 4, 5, 6, and 7 detail the findings obtained in the studies of RO, WOW, F4, JS, and U. Chapter 8 aggregates the findings obtained from the previous chapters and provides practical lessons learned. Chapter 9 concludes with a summary of the dissertation and suggests directions for future work.

Chapter 2

Related Work

Enjoyment and retention are two sides of the same coin. Studying enjoyment is studying why people play; retention, why they keep playing. Retention is enjoyment over time. There is a sizable community of researchers studying fun and game enjoyment. They want to know what makes games appealing, and what is fun. There is also a significant research community studying retention. They focus on explaining or predicting when people stop playing. Despite their common interests, these two communities do not overlap often, and talk very little with each other. Meanwhile, more and more experts, designers, and user researchers from both industry and academia analyze player behavior. This field is called games user research. Their goal is to describe or encourage certain player behaviors in games, such as engaging with content in a certain way, playing longer, or spending money. Although their findings take into account the game mechanics, their conclusions are often game-specific. This dissertation aims at reducing the gap between these three communities by showing how certain game mechanics and player behaviors relate to each other, to enjoyment, and to retention.

2.1 Enjoyment

The core questions asked by the studies focusing on player enjoyment are: Why do people play? What are their motivations to play? What makes a game enjoyable? What is fun?

2.1.1 Motivations

Players have three overarching motivations to play games: challenge, social interactions, and immersion.

Challenge is probably the most commonly found motivation to play games in the literature. One of the oldest studies, conducted by Malone in 1980 by interviewing 65 children in elementary school, suggests that motivation for playing games consists of challenge, fantasy, and curiosity [87]. This study was seminal, in that most of those following it have identified similar motivations. Although a later study based on questionnaire data from 2,226 college students expands the list of motivations to competition, cooperation, recognition, fantasy and curiosity, challenge, and control, the study finds that challenge remains the top motivation to play games [60]. Another quantitative study, conducted on 5,751 players of first-person shooter games, reports that the more committed players, who are members of (semi-)professional clans, are more motivated by challenge and competition than players in amateur clans or without a clan [71].

Social interactions with other players provide another motivation to play games. Not all games allow players to interact with each other, and the extent to which they can interact varies, so this motivation does not always apply. Nonetheless, these interactions fall in three categories: competition, cooperation, and spectating [27]. Although one could argue that competition is nothing else than challenge introduced by other players, cooperation and spectating are clearly different from challenge. In social games like FarmVille, social interactions play such a strong part that among the four motivations players mention in forum posts, three are social: friends and family, competition, emotion (partly social, e.g. when receiving gifts), and time investment [48]. Another ethnographic study focused on social learning finds that players of World of Warcraft particularly enjoy learning about the game and teaching each other [6]. A statistical study shows that the motivations to play obtained from a questionnaire given to 1,993 players include 1) pleasant social interactions through appropriate communication places and tools, and 2) effective personal interaction with the system by providing appropriate goals and feedback – a qualification of how players should be able to tackle challenge in games [31]. Moreover, challenge and social interactions are motivations found across age categories: a survey administered to 124 people aged 45 to 85 years old finds that while the main motivation to play is challenge, the largest predictor of time spent playing is social interactions [37].

Immersion is a motivation that is somewhat less well-defined in the literature than challenge and social interactions. It varies between games, and different players attach different meanings to it [22]. But a previous statistical work gathering thousands of survey data from online game players shows that immersion, as players define it themselves, is as strong a motivation for playing games as challenge and social interactions [131]. Immersion is also mentioned as a motivation to play games by 69% of 271 Baby Boomers surveyed, while 80% mention challenge, and only 12% like to play with others [101]. Broadly defined, immersion consists of two parts: effectance and escapism. Effectance is being able and wanting to interact with and have an impact on the game world. Escapism relates to our imagination and feeling somewhere else than in the real world. In Malone's terms, effectance is curiosity, and escapism fantasy [87]. Another work attributes enjoyment in games to the immediacy of feedback (effectance), and to two aspects of escapism, namely being part of an alternate reality and the story [78]. Game stories can immerse players to some extent by providing familiar narrative schemas such as the hero's journey, but they provide a more potent form of immersion when they "overturn or conjoin conflicting schemas", such as in

Hitchcock's Psycho, in which the protagonist dies in the middle of the movie [49]. Switching from escapism to effectance, a lab study with 500 participants playing different versions of the same game finds that reducing effectance, by making the controls less reliable, severely decreases enjoyment, but making the game more difficult does not [79]. This suggests that while players can accept fluctuations in challenge, they are not as forgiving with immersion. Another author suggests that immersion is necessarily sensory, involves the imagination, and requires some challenge [58]. In fact, many studies on immersion cite flow, the feeling of being in the zone, when focusing intensely on a task [35]. Yet flow has emotional ramifications, and requires the task at hand to be neither too easy nor too hard, which is an assumption about challenge. While flow is a recurring concept in the literature, it actually encompasses multiple motivations to play.

2.1.2 Player Types

The idea around player typologies is to find segments of players more motivated by certain aspects than others. Where motivations focus on aspects of games central to enjoyment, typologies focus on how players are motivated by these aspects. Most of player typologies rely on questionnaires administered to players. Some are based on actual in-game player behavior. A few are purely theoretical. This subsection provides an overview of typologies for a variety of games and game genres. For a more comprehensive review of player typologies, see [124].

In the industry and among gamers, the most well-known typology is probably the casual-hardcore dichotomy. Casual players are frequently described as playing for short periods of time, attracted by simpler and easier games, and with escapism and social interactions as their strongest motivations. Hardcore players, on the other hand, are characterized as more dedicated, playing for long stretches of time, attracted by complex games, and motivated by competition. Even though this simplistic dichotomy is limited [20, 33, 73], it is often the basis for other typologies. For example, a questionnaire of 754 players identifies four profiles: Hardcore, Casual, Well-rounded, and No gamers [88]. Another recent study obtains four player types from the auction house transactions of 20,000 players in the online game Glitch: casual, moderate, hardcore, and forum players [52]. Going beyond the casual-hardcore labels, Bartle suggests two binary dimensions to classify players of online games: acting vs interacting, and with the world vs with other players. Bartle's model results in four categories: the killers like to confront other players, the socializers like to socialize, the achievers like challenge, and the explorers discovering the world [10]. A similar typology based on 60 contextual interviews proposes four ways in which players have fun in games, and suggests a typology based on these kinds of fun. Achievers experience fiero (also called hard fun) when overcoming obstacles. Socializers have "people fun" when interacting with others. Serious fun is when players feel like their actions have an impact, and the rewards are meaningful. Easy fun is more about curiosity, wonder, and surprise [82]. A more recent study pairs questionnaire and telemetry data of nearly 40,000 players of two online games. It suggests six types, some of them already mentioned: socializer, completionist, competitor, escapist, story-driven, and smarty-pants [75]. Another well-known player typology, the DGD1, is based on the Myers-Briggs personality test and playing habits of 408 participants. Some of its four player types overlap with some from the other typologies: conquerors like challenge, managers strategy, wanderers toyplay and variety, and participants social interactions [11]. The latest iteration of the DGD1 mode is BrainHex. Based on survey data from more than 50,000 players, it identifies seven archetypes: seekers, survivors, daredevils, masterminds, conquerors, socializers, and achievers [90]. Despite the large sample size, the questionnaire and types can be improved [24].

Several other works have clustered players based only on their behavior inside a particular game. The results are less typologies than they are player profiles and ways to play a specific game. As such, they are overall more precise, but do not provide player types or play motivations directly applicable to other games or genres. One work looks at the number of deaths, causes of deaths, completion time, and times asked for hints in puzzles for 1,365 players of Tomb Raider: Underworld, an action-adventure console game. It suggests four player profiles: veterans complete the game fast and die little, solvers never ask for help in puzzles and progress cautiously, pacifists die mostly from AI opponents, and runners complete the game the fastest but rely on many hints [50]. A similar study on 1.2 million players of Forza Motorsport 5, a console racing game, identifies 11 profiles based on the amount and diversity of content that players engage with [64]. Another work clusters 6 million World of Warcraft characters based on the kinds of achievements they have unlocked, i.e. related to social interactions, immersion, or challenge. They also find four profiles of players [14].

2.2 Retention

When studying retention, the main questions asked are: What keeps people playing? Can we predict when they will stop? Do players change over time?

2.2.1 Engagement

Engagement is a concept **lacking a clear definition** in the literature. It is often used as a synonym of retention, i.e. as a desire to continue playing, or to go back to playing after a break [23]. Based on 41 interviews, a study echoes this definition by showing that engagement is a process that makes players coming back to play more by linking objectives, activities, accomplishments, and affect within a game [110]. A systematic review lists several ways to measure engagement (from player reports to physiological metrics) and links these measurements to play motivations, time spent playing, and impact of playing on life satisfaction [21]. Another review of 66 purely quantitative papers also lists several variables measuring engagement: the self-reported intention of replay, actually observed replay, loyalty, and likelihood of recommendation. However, only 17 of the 114 independent variables reported are about actual in-game behavior, and the 97 others about self-reported intentions [63]. There are also approaches in the game industry to list and discuss game mechanics increasing retention in online games [7] and social games [61, 62].

Play motivations and engagement seem related, but studies are conflicting in terms of the direction of this relationship. One study performs regression analysis on questionnaire data and finds that social interaction is the strongest predictor of time spent playing [71]. Another study based on seven semi-structured interviews indicates that players need to first invest time, effort, and attention into a game to begin feeling immersed. Only then are they engaged and want to keep playing [22]. Nonetheless, judging from these studies, it is clear that a link between motivations and engagement exist.

Another branch of research on engagement focuses on the impact of **tutorials** on long-term player engagement with the game. One study compares three games of varying complexity. Measuring session duration, activity, and return rate from telemetry data for 45,000 players, it finds that a tutorial does increase the return rate for the more complex game, but not for the two simpler ones [5]. In another study, the authors conduct a qualitative text analysis of hundreds of game reviews. They find that tutorials are generally very enjoyable, fun, and rewarding, but recommend developers to add elements of intrigue and more informational mechanics to help players determine that they should continue playing the game [28].

2.2.2 Churn

Churn is a synonym for player drop-out. A large community of researchers focus on explaining or predicting churn. Their large-scale approaches make extensive use of analytics, data mining, and machine learning techniques [18, 65, 113]. Some studies are descriptive in nature. For example, one looks at data from 250,000 players from 5 action-adventure and FPS console games and finds that a Weibull distribution is a good fit summarizing the distribution of players' lifetime [13]. Most studies, though, are predictive in nature. Their goal is to build models that predict accurately when players will churn. Interpretability of these models by humans is generally of no concern, although some models can be visualized.

Defining churn: Churn is defined as the moment when players drop out of the game. However, players may have stopped being engaged for a while. This prompts some studies to define churn not as the time when players are completely inactive, but rather less active. For example, a study looking at 38,000 players of the online game Aion over 6 months finds that it is possible to predict with 95% precision and recall when players drop below 6 days of activity in the past 30 days [95]. This definition is adapted in another study predicting churn in Diamond Dash, a free-to-play mobile game: players are labeled as churners when they do not play for two weeks. Following this definition, the study finds that it is possible to redirect high spenders to other games of the same developer by giving players an incentive in the form of in-game currency before they churn [108]. Another study predicts churn based on the state and activity of players' peers in their social network [76]. Yet another study uses information gain on in-game behavioral features such as number of sessions, total play time, and inter-session length to identify differences between soon-to-be churners and engaged players [19].

Adaptive difficulty: One community among game AI research uses churn prediction to dynamically adapt games to players' tastes or skill level [136]. The side effect, and sometimes even the end goal, is a reduction in churn. One work from this field predicts when Scrabble players are going to churn based on their performance, and adapts the tiles drawn accordingly to reduce their likelihood to churn [66]. Another work procedurally generates race tracks fitting the player's driving style [123]. Two others add handicaps to balance games and make them more fun for both experts and less-skilled players alike [12, 81].

2.2.3 Longitudinal Studies

Some studies aim at describing changes in player behavior over time. The findings can be at the scale of the whole player base or for each particular player. One such study, mentioned earlier, looks at game data from 20,000 players to cluster them in four clusters and show that players change cluster over time [52]. A similar clustering of 60,000 players based on the type of actions performed in an action-adventure game (e.g. jumping or fighting) also shows that players change cluster over time [116]. Another study combines game data from 690 World of Warcraft players gathered over 8 months with their in-game behavior data. It finds that the players most motivated by teamwork and competition are tend to level up and progress through the game the fastest [16]. Finally, a qualitative analysis of player posts on the World of Warcraft forum shows that players expect to progress and acquire better and better rewards as they spend more time in the game [100].

2.3 In-Game Behavior

The core questions asked in games user research are: How do players behave? Which mechanics encourage or discourage certain in-game behaviors? How to make better games? How to increase revenues?

2.3.1 Social Sciences

One branch of games studies analyzes how people play together using an anthropological or sociological lens. In **anthropological** game studies, a frequent goal is to observe how online players build a culture and interact within a community existing both in-game and in real life [103, 122]. Noteworthy anthropological studies include an analysis of World of Warcraft (WOW) chat logs to see how players learn strategies, discuss item usage, and negotiate game ethos [94]; ethnography and participant observation in the online multi-player puzzle game URU and the events following its closing down [102]; an ethnographic analysis of the social structures and practices among dedicated EverQuest "power players" [121]; and a textual analysis of forum posts and game reviews focusing on the negotiation of social capital among the most devoted players of a Facebook game [33].

Sociological game studies focus on the structure and nuances of player interactions. Most of them are based on World of Warcraft. A recent study using mixed methods tracks the 25-man dungeon progression over four months for two mid-core guilds in WOW. Progression is "slow and painful": for every three attempts at defeating a boss, two fail and one succeeds. This pushes guild leaders to alternate between farming easy bosses to boost morale or try harder bosses as a challenge [9]. Another quantitative study tracks the characters present in one WOW server every 15 minutes using a bot. It finds that only 20-40% of player's time is spent in a group, with an upward trend as characters level up, and that guilds with 16-60 members have a tight core of only 6-9 players [55]. Another study of WOW guilds, qualitative this time, finds that guilds with more than 35 members have a formal leadership structure and rules, whereas 75% of the guilds with 10 or less members consist of informal gatherings of real-life friends or family. Interestingly, that study also reports that women are more active than men in recruiting new members to their guild [128]. Switching from guilds to players themselves, a survey of 180 players of three online games (Second Life, Mapple Story, and WOW) finds that a quarter of players, mostly men, play a character of the opposite gender [54]. Other gender differences include women playing more than men, being more likely than men to play with their partner, and underestimating their weekly play time more [127]. In-game behavior does not just differ between men and women, but also between male and female characters. For example, women playing a male character spend more time in PVP than women playing a female character, and men playing a female character heal more than men playing a male character [134].

2.3.2 Improving Gameplay

Designers in the game industry suggest guidelines and provide lessons learned, generally through talks they give at game developer conferences, sometimes in books or blogs, but rarely in peer-reviewed articles. For example, Pardo, a famous game designer, lists tweaks brought to Starcraft 2 and WOW mechanics so that players would enjoy them more [99]. Another Blizzard designer explains why they removed caps on daily quests in WOW: players had a "pressure to maximize everything", leading them to burnout [29]. Designers also publish dos and don'ts, such as eight ways to design tutorials badly [2]. Raph Koster, another prominent designer, suggests that players quit playing when they have mastered all the patterns the game offers [80]. This echoes academic research in learning and games for learning [59]. In terms of difficulty and challenge, designers generally recommend increasing difficulty smoothly over time [3, 111].

Games user research is another branch of the game industry measuring the success of game mechanics on players, and observing how players interact with a game. Games user researchers gather consumer feedback through focus groups, retrospective surveys, beta tests, traditional usability testing, and first hour trials [36]. Playtests, in which the game studio invites players to their lab to observe them and track their behavior while they play an unreleased game, can focus on parts of the game other than the first hour, and provide a somewhat controlled experiment setting which can provide both qualitative and quantitative data about players' in-game behavior [4]. Although qualitative methods are widespread [67, 86], game companies also use quantitative methods such as eye-tracking, psycho-physiological data and game telemetry [135]. There are even more complex instrumentation and data collection platforms for games, such as Microsoft's TRUE [77].

In **academia**, although the methods to study gameplay tend to be quantitative, the study goals are somewhat more diverse. For example, some studies aim at describing challenge and player skill over time. Tracking 5-man pick-up groups over time in WOW, and defining encounter difficulty as character health lost and distance between group members, one notices that challenge decreases as characters gain gear [8]. Another tracks the evolution of skill in the first-person shooter Halo: Reach, with a particular focus on how fast players regain skill after breaks [69]. In Counter-Strike, another first-person shooter, expert players are not concerned by particular actions such as jumping, but rather by specific chains of actions such as those required to take control of an area [105]. In yet another shooter called Kane & Lynch 2, players can fail several times in a row at completing an objective. To understand this frustration, a visualization of player movement is overlaid on the game map and combined with quotes from 22 player interviews [25]. Frustration can also stem from latency, also called lag, in online multi-player games. One lab study shows that players do worse when latency spikes irregularly than when it is constant – players get used to it and manage their character accordingly [119].

2.3.3 In-Game Purchases

In some games, players can buy virtual items for real money. The earliest studies of in-game purchases focus on the in-game economies of virtual worlds [26] and the value of playbor, i.e. players making real money from selling virtual items to other players [45]. But more and more companies realize that they can satisfy that demand themselves following a free-to-play business model where the developer releases the game for free, but sells virtual items to players for real money through micro-transactions. More and more academic research is devoted to understanding the factors driving these in-game sales: the items sold can be functional (e.g. power-ups or bonuses), social (e.g. gifts), or aesthetic (e.g. garments or accessories) [84]. Small spenders tend to purchase functional consumables, whereas big spenders buy decorations, character customization, and gifts for other players [129]. Note that this categorization matches the list of three play motivations mentioned above: challenge, social interactions, and immersion. The practice of gifting also brings to mind the concept of social capital frequently mentioned in anthropological game studies.

In practice, only a small portion of the player base purchases virtual items, and there are noticeable cultural differences. For example, Asian players spend more than Americans or Europeans [83]. There are also differences between game genres: on the Kongregate portal where thousands of games can be played, the average paying user spends $10 in single-player games, but $300 in multi-player games. Players in guilds also spend 20 times more [30]. Players' position in the social network also determines how much they are ready to spend, and how much they can make their peers spend [126]. There are also works predicting which mobile game players will purchase virtual currency, and how much [117].

2.4 Summary

Challenge is the main motivation for play. It is the factor most often mentioned in the literature, being involved in most player types, theoretical explanations of engagement, empirical predictions of churn, and several other in-game behaviors such as purchasing. This makes challenge the prime factor to consider when studying retention. Moreover, challenge is a concept that is both very broad and game-specific. For example, it includes progression and daily quests in World of Warcraft, aiming skill and reflexes in Halo: Reach, and letter difficulty in Scrabble. Besides challenge, social interactions and immersion are play motivations frequently mentioned in the literature and encountered in players' in-game behavior.

Retention is related to engagement, a concept on which the more theoretical literature does not completely agree. However, more empirical and longitudinal studies predicting when players stop playing quantify retention more clearly in terms of the number of players who stop playing for a certain duration. Again, this duration seems game-specific.

Chapter 3

Ragnarok Online

Ragnarok Online (RO) is a Korean Massively Multi-player Online Role Playing Game (MMORPG) developed and published by Gravity since 2002. It is still running as of 2016. In 2009, when we collected data for this work, Gravity provided servers in 15 locations worldwide and claimed that the game counted more than 50 million players[1]. But alongside official servers, there is a large community of illegal private servers. This chapter is organized as follows. First, we briefly introduce the game, and describe the methodology used to collect data on the private and official servers. Then we discuss how private server administrators modify the game, and how these modifications answer the needs of their players in terms of difficulty and quality of life.

3.1 Gameplay

In RO, players control a character in a virtual world shared with other players. In towns, they can trade items and socialize with other players. In the wilderness and in dungeons,

[1] According to Gravity, RO "has more than 50 million registered users worldwide with over 3 million in the US". See `http://www.warpportal.com/news/press.aspx?page=12`

they can engage in Player-versus-Environment (PvE) activities such as slaying monsters and completing quests. In guild castles, they can battle other players in bi-weekly player-versus-player (PvP) events.

All these activities grant experience points (XP) and levels. When reaching the maximum level of 99, players are granted an aura shining around their character. The aura is a cosmetic way for players to brag about their level. XP is obtained mostly only by killing monsters. Players can "grind" in solo by killing thousands of easy monsters, or team up with other players and try tackling challenging bosses. These appear in pre-determined locations, and disappear for several hours after they have been killed. Unless players "camp" a boss' reappearance, they have no chance of even seeing it. Thus RO's end-game PvE provides little challenge, prompting some players to call the game a "dead end" [44]. More modern MMOs like World of Warcraft provide more challenging end-game PvE by instantiating bosses in separate dungeon replicas for each group of players attacking them.

Another common feature of MMOs found in RO is the item dropping system. When monsters die, they sometimes drop certain items at a certain rate. The drop rate ranges from 70% for the most frequent items to 0.01% for the rarest. Items can be sold for money or traded for other items with other players. Merchant players can sell items to other players through their roadside shop, as shown in Figure 3.1.

RO also has extensive character customization. Players can choose their character's cloth color, hair color, hair style, hat, glasses, and even mouth-guard. Since RO is based on the eponymous Korean manga, the in-game art is very "kawaii": characters have big heads and eyes, and they are drawn in flat 2D sprites with color palettes, whereas the world is rendered in 3D polygons with textures. The game background is loosely based on Norse mythology with gods such as Freya and Loki, places like Valhalla and Niflheim, and equipment such as the Sleipnir shoes. Some game updates brought content based on other ancient cultures including ancient Japan, China, or Turkey.

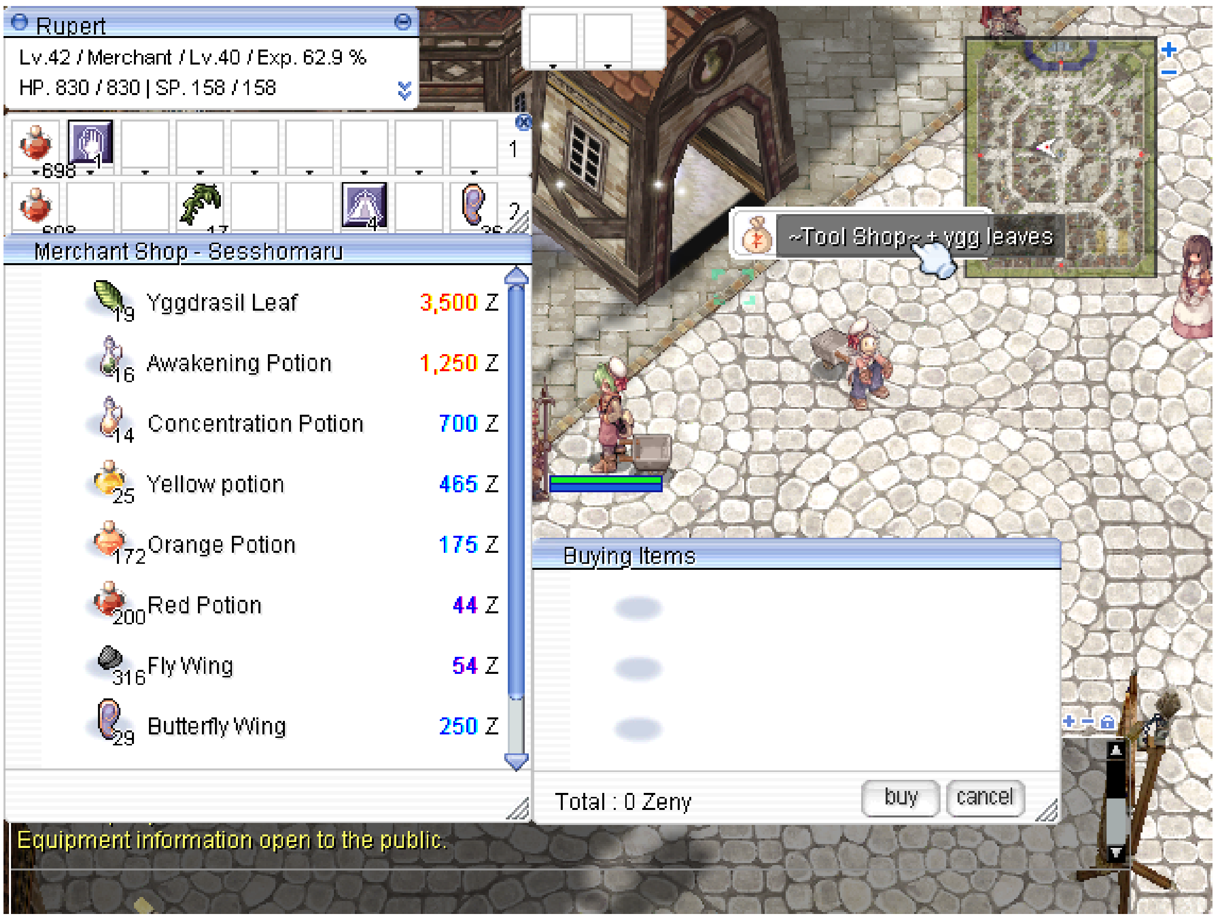


Figure 3.1: Screenshot of a player shop in Ragnarok Online

In Ragnarok, the virtual world is made of 16 x 16 self-contained square maps – monsters can roam within a map, but cannot leave it for another. Maps are linked together via portals on their North, East, South, and West borders. Traveling from one end of a map to another takes a few minutes, and traveling from one end of the world to another more than an hour. This is too long and tiresome. Therefore each city has a Non-Player Character (NPC) offering portals to the nearest 3-5 cities for a fee. Some player characters such as Acolytes can memorize 3 locations anywhere in the world, not just cities, and can teleport other players there, sometimes charging a fee.

3.2 Private Servers

Gravity runs 20 or so official servers of RO worldwide. Some servers are subscription-based, and some are free-to-play. There are also thousands of individuals worldwide running their own RO game server[2]. Running a private server is illegal in countries where Gravity runs the game. Technically, private server administrators use a software called a server emulator. Emulators accept connections from game clients like official game servers do. Emulators can be built from scratch, or leaked from the developer. RO's server source code, Aegis, was leaked and used in the jAthena and eAthena emulators. Emulators are commonplace for popular MMOs: the Mangos project reverse-engineered the code for World of Warcraft servers, and Aion-emu did the same for Aion.

Information concerning emulators can be hard to find because developers are well aware of the illegality of their servers. Therefore, setting up an illegal server is not easy and requires technical knowledge in server administration. Some private server administrators add new content that does not exist on the official game server in a way similar to game mods. So private server administrators tinker with both the server- and client-side of the game.

[2] The international support board for RO private server administrators, eathena.ws, counted more than 100,000 registered users worldwide in 2009.

Websites such as ratemyserver.net allow RO private server administrators to advertise their server, and players to vote for their favorite.

When this study was conducted in late 2009, two broad categories of online entertainment were well-researched: virtual worlds like Second Life [15] and popular Western MMOs like Everquest II [26, 127] and World of Warcraft [55, 72, 93]. A third category seems to have been ignored: Asian MMOs counting tens of millions of users worldwide including MapleStory [120] and Ragnarok Online. There is also no literature on private servers at all. To shed more light on Asian MMOs and private servers, this chapter tackles the following questions: Why do private server administrators step into illegality and host servers for players they do not even know? Why are private server players not going to official servers when they are free-to-play? What are the most common modifications brought to the game by private servers? What lessons can official servers learn from them?

3.3 Methods and Limitations

In France, the official game server, called french Ragnarok Online (FRO), launched in June 2007 with a subscription-based model. In December 2008, FRO went free-to-play. Despite this change, a French community of illegal private server administrators and players still exists as of late 2009, and FRO only has a thousand players online at peak time. Thus the French official server is a great candidate to compare player experience between private and official servers.

To pick a private server, we go to ratemyserver.net, the main website where players can search for a private server among the more than 500 available[3]. For each private server, ratemyserver.net provides a link to their website and their difficulty "rate", a number ranging

[3]In a period of three months, at least 350 of the 500 private servers present on this website were active. As a comparison, there were about 30 active official game servers worldwide in 2009.

from 1 to a million. Low-rate private servers have rates under 30, medium-rate between 30 and 150, high-rate between 150 and 10,000 and very high rate over 10,000. We choose a rate of 50 because it is somewhat in the middle of the range of possible rates. A rate of 50 means that monsters on the private server give 50 times more XP and are 50 times more likely to drop items than on an official server. We pick two private servers: PS1, a French private server with 200 concurrent users at peak time, and PS2, an international private server with 2,000 concurrent users at peak time. PS1 hosts players from France, Belgium, Quebec, and other French-speaking locations. PS2 hosts players from across the world, with large communities from Germany, Brazil, and the US.

This chapter combines several qualitative sources, all conducted or accessed in late 2009: 1) Informal in-game chat with dozens of players from all three servers. 2) Eight semi-structured interviews, with four players from PS1, two players from PS2, and two players from FRO. Three of the interviewees were female players. In-game interview logs were saved using the /savechat game command, and analyzed offline later. 3) A semi-structured interview with the administrator of PS1. 4) Participant observation while playing the game with dozens of players. 5) Public results from player surveys conducted by iRO, the official international English-speaking server[4]. And finally, 6) forum posts published on eathena.ws and the forum boards of the three servers.

Table 3.1 lists three in-game issues and their solutions on private servers. First, players on small private servers have trouble finding groups of other players to join. Various social incentives address this issue. Second, the in-game economy on small private servers can be weak. This is solved by making items easier to find from both monsters and players. Last, private servers improve the general quality of life and provide extra features for their players compared to official servers.

This study has several limitations. The first limitation is the sampling of players. Eight

[4]See Kafra Survey at `http://iro.ragnarokonline.com/community/polllist.asp`

Issue	Solutions
Group play	Tweaking group parameters, Warpra and Healer, @who command
Economy	Rates, @autotrade and @whosell commands
Quality of life	Avatar customization, antibot measures, control panel, in-game events

Table 3.1: In-game issues and their solutions found on private servers.

interviewees do not represent all players of the three servers. We can only conduct participant observation with players who like to play with others. Thus the solitary players, however rare they may be, are under-represented in this study. Nonetheless, we gather a range of different sources to triangulate and confirm interview statements.

A second limitation is the sampling of servers. Players on the French official server FRO may differ greatly from those on the most populated international server iRO or the most up-to-date Korean server kRO. After all, the game's Kawaii art reminds us that the game was first designed for an Asian audience. As for private servers, PS1 and PS2 are middle-rate servers. Players from very high rate or low rate servers may have drastically different experiences and opinions about private servers. PS1 peaks at 200 concurrent users, and PS2 at 2,000. They probably do not reflect what happens on servers with 20 peak concurrent users.

3.4 Supporting Group Play

A key feature of MMOs is teaming up with other players to defeat challenging monsters and complete quests. Larger groups can reach greater achievements. When servers do not have enough players, groups tend to be smaller and more scarce, and the game more difficult and less rewarding.

3.4.1 Tweaking Group Parameters

In RO, players in the same group can share their experience points (XP) and drops. On an official server such as FRO, XP is equally divided between all members on the same map, whether or not they hit the same monster. The maximum group size is 12, and all members must be within 10 levels of each other. On the private server PS1, the smaller population makes it harder for players to find a group. Three techniques facilitate and encourage group play. First, the maximum group size is increased to 15. When players within a certain level range are scarce, groups cannot be picky about their members. Their composition is often suboptimal to tackle challenging objectives. Adding 3 players can make the difference between utter failure and epic success. Second, the level range within a group is increased to 15. This increases the likelihood of finding suitable group members by 50%. It also consolidates existing groups: level-75 players can join groups of level-60 players or level-90 players rather than make their own. Third, an XP bonus is given for each player hitting the same monster. This encourages co-located group play and social interactions. When players are spread around the map, they collaborate with each other only remotely.

3.4.2 The who Command

Private server administrators can give their players certain commands that official servers only grant to their game masters. Commands are typed in the regular chat box and start with the @ (at) symbol. For example, @who displays the names of all players currently online, whereas the basic /who player command only displays how many are online. An official server player typing @who will receive an error message such as “no such command”. To play in groups, players need to know the name of other simultaneously connected players. On official servers, often very populated, players can just look around in their current map. On young or small private servers, such as PS1, building and strengthening the community

Figure 3.2: Warpra and Healer.

is crucial to ensure the server's survival. So private server administrators give their players a tool to find other players.

3.4.3 Warpra and Healer

Downtime occurs mostly when regenerating life and traveling the world. Acolyte players can instantly heal and teleport other players. They are very valuable in a group and greatly improve other players' game experience. On PS1, Acolyte players are scarce. Two non-player characters fill their role in every city: the Warpra teleports players to any city or dungeon entrance, and the Healer regenerates players' life instantly. The Warpra and Healer are displayed in Figure 3.2. On the more populated PS2, Acolytes are common. Yet the Warpra and Healer are still implemented and widely used by players. As a result, players play alone more often because if they are killed by strong monsters, instead of trying to find a group to defeat them, these two NPCs combined instantly restore their health and send them back to the dungeon where the wounded monsters are. On ratemyserver.net, nearly all middle-, high- and very high-rate private servers advertise their Warpra and Healer. Some low-rate servers have them too, but no official server does.

The Warpra and Healer also have a major drawback: they make boss fights much easier. Monk characters can quickly localize bosses on their map by teleporting around, and inflict huge damage with their ultimate skill Asura Strike. After having used this skill, they let the boss kill them, reappear in town, are healed instantly by the Healer, and teleport back to the boss through the Warpra. Since bosses are free-for-all, boss health does not regenerate. Following this process, monks can defeat bosses in a few suicide runs. While bosses are supposed to be a challenging end-game PvE activity for groups, they are now a very easy task to solo.

It is surprising that neither PS1 nor PS2 tried to solve this issue. Maybe their administrators fear losing players if the original game is changed too radically. However, at least three solutions exist. The easiest solution consists of limiting the damage of Asura Strike on boss monsters. This involves changing a couple lines of code. The drawback is that monks would not be used as damage dealers in end-game PvE anymore. A more sustainable solution consists of regenerating bosses to full life when they have not been attacked for a certain time. This is similar to the concept of leashing in World of Warcraft[5]. This solution still does not address the scarcity of challenging end-game PvE. An even more comprehensive solution consists of implementing an instance system for all bosses, similar to World of Warcraft. This would allow multiple groups of people to kill the same boss at the same time. This instancing system is already implemented for several hourly, daily, and weekly dungeons[6], so it would not require too much development effort. This solution provides the extra possibility of tailoring the boss difficulty and drops to the composition of each group.

Another drawback of the Healer and Warpra is that they encourage solo play through multi-boxing. When a player multiboxes, she runs multiple game clients and alt-tabs to control multiple characters simultaneously. In free-to-play games, making another account costs nothing. And although multi-boxing is against FRO's terms of service, the technical solutions

[5]See `http://wowwiki.wikia.com/wiki/Leash`
[6]See `http://irowiki.org/wiki/Instance`

in place (if any) do not seem effective. Players report multi-boxing to level up one of their weaker character with their strongest one. The administrator of PS1 reports that during PvP events, players leave their secondary priest character near the Warpra and Healer in order to buff their main character when they die and need to teleport back to fight. Allowing multi-boxing encourages solo play because it makes for a "faster avatar leveling". All interviewed players mention multi-boxing, but only one deplores it: "people play as if they were alone".

3.5 Maintaining an Economy

RO was designed with a certain number of players per server in mind. The drop rate and price of each item are based on thousands of players looting and selling millions of items every day. For supply and demand to stabilize item prices, the player base needs to be large enough. This is an issue on private servers. How do private servers maintain their in-game economy?

3.5.1 Rates

MMOs are games with many numbers. Private server administrators can tweak most of these numbers, but they pay a lot of attention to one in particular: the rate at which monsters drop items and grant XP to players, i.e. the difficulty rate. Different difficulty rates set different player expectations and lead to radically different experiences. For instance, a rate-3 server is very similar to the official game. Players expect incentive structures, prices, and activities, similar to official servers. On a rate-100,000 server, leveling up is so easy that players expect more activities than simply grinding XP. Most of the time, monsters die in less than a second without having time to hit the player. Administrators of very high rate servers often increase the bosses' skills and health points so that bosses stay challenging. PS1 and PS2 are both medium-rate private servers. They stand somewhere in-between. Interestingly, player demographics vary depending on the rates. The private server administrator we interviewed remarks that “The easier the game is, the younger the players are, and the sooner they come back from school and start to play, moving the peak time earlier”.

A previous work on Maple Story, Second Life, and World of Warcraft found that “older users broadly prefer creating an avatar that looks like an idealized version of themselves” [54]. According to the PS1 administrator, the population in high-rate and very high-rate private servers tends to be younger than in lower rate servers, probably because the game is easier. No player interviewed confirms or rejects this claim, but there could be a correlation between player age and tolerance to difficulty.

3.5.2 The autotrade Command

Players can buy equipment and potions from non-player characters or from other players. When players trade with each other, both players need to be online, unless the transaction occurs through a merchant player’s roadside shop while they are away from the game, but still online. This is shown in Figure 3.1.

Some private servers enable the @autotrade command. It allows merchant players to let the game server handle their shop for them while they are offline. With autotrade, they do not need to be logged into the game for their character to sell goods. To other players, the auto-trading merchant player behaves exactly like a non-player character. Players can play another character while their auto-trading character is selling their stock. Consequently, the number of players logged-in at the same time increases, which is desirable for private server administrators looking to boost their stats. At peak time on PS1, up to 25% (45/170) of the logged-in avatars can be auto-trading merchants. During the least busy hours, the proportion of auto-trading merchants can rise to 90% (35/39). This leads players connected in the middle of the night to compare the server to a ghost ship.

Autotrade merchants regularly connect to update their stock, check the competition's prices, and eventually update theirs. This creates a reverse auction system, also called undercutting. We observed this first-hand: In PS2, player 1 starts selling a chain mail for 400k in the autotrade area. A few hours later, player 2 logs in, and starts selling a chain mail for 380k. Several hours later, player 1 comes back, and decreases the price to 370k. This may seem paradoxical: since the rates are higher than on official servers, players should earn money more easily, and item prices should raise as well. But because the rates are higher, monsters drop items more frequently. Having more items dropped and merchants selling them permanently increases supply, and prices remain comparable between low/middle-rate private servers and the official servers despite the rate differences.

3.5.3 The whosell Command

Private server administrators can enable the @whosell command for their players. A private server player typing "@whosell Chain Mail" in the chat box receives a list of auto-trading merchant players currently selling a Chain Mail, and for what price. The @whosell command is not enabled on any official server. Some FRO players complain that prices can vary a lot. FRO actually hosts a Price Watch feature on its web site so players can search which merchant player is currently selling which item. However, the Price Watch web pages are not easily accessible from the web site, and require players to log out or at least alt-tab out of the game. This can explain why no FRO player mentions it. PS1 and PS2 players are generally satisfied with item prices and their server's economy.

3.6 Improving the Quality of Life

Private servers add extra features to the game to please their players. These features include: more ways for their players to customize their character, a control panel allowing players to see their stats outside the game, dedicated community management, and an effective detection and removal of bots.

3.6.1 Character Customization

In hacking the original executable file used to launch the game, private server administrators can make more character customization available for their players. Hats, hair styles, hair colors and clothes colors can be extensively modified. As Ducheneaut et al. suggest in [54], an attractive character customization system enables players to customize “those parts of their virtual bodies that will be most immediately visible and recognizable by others, and which are easily adaptable and commonly modified in real life”. Hair style, hair color, hats, and clothes are perfect candidates. While the official game provides only a few customization possibilities, some private servers boast dozens of hair styles and level-99 auras surrounding the player, and hundreds of hats, hair colors, and clothing colors. Table 3.2 compares the amount of customization available on official servers vs private servers. With many different possible styles and colors, private server players have more unique characters compared to official server players. However, more customization comes at a cost: players report that some hair color palettes in the 400 range can be buggy, flashy, and very ugly. As a result, some private server players can spend a lot of time selecting a satisfactory appearance.

Feature	FRO	PS1 and PS2
Hair styles	23	50
Hair colors	8	500
Clothes colors	4	1500
Hats	around 350	500
Level 99 auras	1	70

Table 3.2: Comparison of character customization on official vs private servers.

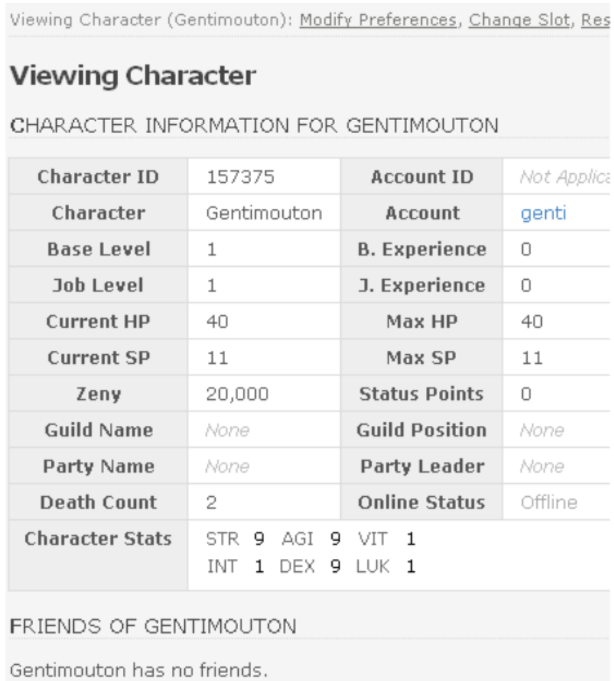
3.6.2 Control Panel

The control panel is a part of a private server's website linked with the game database. It shows the players currently online, allows players to transfer money between their characters, and ranks players on leaderboards based on their money or level. Nearly all the private server players reported using the "Who is Online" functionality to check for logged-in friends. Players on official servers would like to have a control panel. A survey posted on iRO's website shows that 71% of official server players "would like to see rankings [...], guild castle holds [...], and class rankings" ($N = 5230$). While official servers like iRO do not provide such functionality as of 2009, a private servers control panel offers players a way to interact with the game without being logged-in. Other MMOs such as World of Warcraft publicly provide players' information through a website called the Armory, a popular addition to the game[7]. Figure 3.3 is a screenshot of the Control Panel provided on private server PS2.

3.6.3 Bot Hunting

Private server administrators can also use the control panel to detect bots posing as humans. A bot is a program that players launch to play automatically for them. Bots react faster and perform repetitive tasks more effectively than humans. In analyzing packet traffic, the Control Panel allows private server administrators to detect that a particularly high-traffic

[7]See `http://icrontic.com/article/world-of-warcraft-us-armorys-got-a-brand-new-bag-of-tricks`



Viewing Character (Gentimouton): Modify Preferences, Change Slot, Res

Viewing Character

CHARACTER INFORMATION FOR GENTIMOUTON

Character ID	157375	Account ID	Not Applica
Character	Gentimouton	Account	genti
Base Level	1	B. Experience	0
Job Level	1	J. Experience	0
Current HP	40	Max HP	40
Current SP	11	Max SP	11
Zeny	20,000	Status Points	0
Guild Name	None	Guild Position	None
Party Name	None	Party Leader	None
Death Count	2	Online Status	Offline
Character Stats	STR 9 AGI 9 VIT 1 INT 1 DEX 9 LUK 1		

FRIENDS OF GENTIMOUTON

Gentimouton has no friends.

Figure 3.3: Control Panel of a Ragnarok Online private server.

player following certain repetitive patterns in the game is actually a bot. Administrators also sometimes log into the game to chat with suspicious players and assess whether they are robots. As a result, administrators can boast having a bot-free server and a fairer game which players may greatly appreciate.

3.6.4 Community Management

Game masters (GM) organize in-game events such as treasure hunts or monster invasions, support players facing technical difficulties inside the game, and moderate the forums. Two surveys on the iRO website show that official server iRO players are unsatisfied with their Game Masters. The first survey asks "How would you rate the service of iRO compared to the service of other MMORPGs that you have played?". 55% of players say they are not satisfied, 27% find the iRO service good, and 18% have no opinion ($N = 1,334$). A second iRO survey reveals that the quality of GM work was appreciated by 22% of the players, but 62% judge their actions insufficient or bad ($N = 1,329$). FRO players are unsatisfied too: one player complains that events on the official server are too rare and of poor quality. Another FRO player wishes for more available and responsive GMs. On the other hand, a player from PS1 mentions that GMs on their private server are responsive probably because they have few players to deal with.

3.6.5 Free-to-Play vs Pay-to-Win

Community management is also very important with respect to the free-to-play business model. Although the official server FRO is free-to-play, players can pay a monthly subscription to receive in-game privileges such as powerful items and faster XP. As a consequence, guilds have become "elitist in the way they recruit" since the game became free-to-play: "if you do not have any subscription we do not take you, if you do not have any cranial [piece of equipment] we do not take you". Private server PS2 has no subscription, but makes certain useful commands and powerful equipment available in-game only through a Paypal donation system. After all, the server infrastructure needs to be paid. But a PS2 player finds that players with Paypal-purchased equipment have an unfair advantage over those without. On PS1, there is neither subscription nor donation: the administrator spends her own money to run the small server. Players do not complain about the game being unfair or "pay-to-win".

3.7 Summary

Even though official servers are free-to-play, some players expect a quality of service they do not seem to find. These players turn to illegal private servers because they are easier, provide more character customization, encourage grouping while allowing solo play, and have responsive game masters. Even though the competition between private servers is

Category	Finding
Challenge	Warpra and Healer compensate for the lack of players.
Challenge/Social	Warpra and Healer are popular, but incentivize solo play.
Challenge	Players prefer higher rates because it reduces grinding.
Challenge	@whosell and @autotrade increase trade activity.
Segments	Higher rates may attract younger players.
Social	Players like active and personal community management.
Social	Groups encouraged through @who, group size, and level range.
Social	Customization and control panel allow players to show off.
Money	Elitist guilds on free-to-play servers require members to pay.
Infrastructure	Players acknowledge and appreciate bot detection tools.

Table 3.3: Summary of findings for RO.

keen, creative administrators find ways to satisfy their players in getting them know each other better, preventing bots, or even stabilizing a fragile economy. Therefore, official game servers could benefit from the following design recommendations: 1) Reduce the amount of grinding by increasing the experience and drop rates. 2) Provide an interface for players to better navigate and participate in the in-game economy. The private server commands @whosell and @autotrade are rudimentary and text-only, but a centralized auction house may be an appropriate solution. 3) Provide a greater amount of character customization, such as more hair styles and clothes colors, and expose player character data through a web API or at least a web front-end.

Small private server administrators have managed to turn what could have been a weakness, namely the lack of players in an MMO, into a strength. Gravity has been hunting down large private servers singe the game launched[8]. But as the game is aging and competition on the MMO market remains fierce, these private server communities could be a cheap solution to increase the game's longevity.

[8]See `http://pacificepoch.com/china-investment-research/articles/shanda-to-crack-down-on-ro-pirates/`

Chapter 4

World of Warcraft

World of Warcraft (WoW) is a fantasy role-playing MMO developed and published by Blizzard since 2005. It is available worldwide, and is still the MMO with most subscribers as of 2016. The number of subscribers peaked at 12 million subscribers in 2010, and has kept decreasing since then to reach 5 million in late 2015[1]. This chapter relies on three studies we have already published. The first relies on survey data from 2010 to explore various churn patterns and player segments [40]. The second study uses the same survey data from 2010, but focuses on players who purchase in-game currency for real money to gain an advantage in the game [41]. The third study pairs survey data from 2011 with longitudinal gameplay data from 2012 to explore how several player segments progress and churn after new game content is released [39]. The timeline of WoW content patches, the two surveys, and the gameplay data collection is listed in Table 4.1. After briefly introducing WoW's gameplay, we go in more details about the methods and data of the three studies. We present several segments displaying large differences in terms of player commitment, willingness to purchase virtual currency, and pace of progression.

[1]See `http://www.statista.com/statistics/276601/number-of-world-of-warcraft-subscribers-by-quarter/`

Date	Version	Event
2004/11	1.0	Launch in the US
2005/02	1.0	Launch in Korea and Europe
2005/06	1.0	Launch in China
2007/01	2.0	Burning Crusade expansion
2008/11	3.0	Wrath of the Lich King expansion
2009/12	3.3	Icecrown Citadel dungeon
2010/03	–	**First Online Survey**
2010/12	4.0	Cataclysm expansion
2011/06	4.2	Firelands dungeon
2011/10	–	**Second Online Survey**
2011/11	–	**Start of gameplay data collection**
2011/11	4.3	Dragon Soul dungeon
2012/06	–	**End of gameplay data collection**
2012/09	5.0	Mists of Pandaria expansion

Table 4.1: Partial World of Warcraft timeline.

4.1 Gameplay

In WoW, players wander in a large virtual world made of several continents each cut in dozens of zones. Zones fall in three categories: cities, player versus environment (PvE), and player versus player (PvP). In cities, players can socialize, trade items through the auction house, or find other players to team up with. Cities also start many quest lines that players can complete to gain experience points (XP) and gold, the in-game currency. XP and leveling in WoW is much faster than in RO. Making a character reach the maximum level takes only a couple month. Once players reach maximum level, the end-game starts. The goal is to continuously improve character equipment by participating in difficult PvE and PvP activities.

PvE consists of battling monsters in the open world and dungeons. Open world zones host monsters and non-player characters. Dungeons host particularly challenging monsters called bosses. New dungeons are added to the game every 4-8 months. The rewards for defeating monsters are XP, gold, items, and sometimes completing a quest line. Whereas players

Figure 4.1: World of Warcraft screenshot.

can roam the open world in solo, they have to team up in groups of 5, 10, or 25 to go in dungeons. These groups can originate from guilds, in which players socialize and make friends, or through ad-hoc pick-up groups assembled automatically by the game. The PvE difficulty follows the game design motto “easy to learn but hard to master” [98]: monsters in the beginner zones are very easy to solo, but end-game bosses require groups of players to coordinate their actions perfectly. These bosses drop the best equipment in the game. Some guilds require their members to dedicate several nights per week to raid these bosses.

PvP zones are self-contained areas where players fight other players. PvP activities involve duels, arenas (2v2, 3v3, or 5v5), battlegrounds (10v10 or 40v40) and temporary open-world zone-wide fighting against players of the opposite faction (up to hundreds of players on each side). Difficulty in PvP comes from other players having better hand-eye coordination.

4.2 Methods and Limitations

This chapter explores patterns of churn, gold purchase, and progression. The first and second rely on the same survey from 2010. The third pairs survey data from 2011 with longitudinal gameplay data from 2011 and 2012. In this section, we describe these two datasets in more details.

4.2.1 Dataset 1: First Survey

The first survey was conducted by Nicolas Ducheneaut and Nick Yee between March and May 2010. They let us add to their questionnaire several questions of our own about retention and gold buying. Links to this survey were sent to mailing lists from their previous studies, broadcasted on social media like Facebook and Twitter, and posted on gaming websites such as wow.com and wow insider, and on forums from Taiwan, Hong-Kong, and the US. The survey was composed in English and translated in Traditional Chinese for Hong-Kong and Taiwan. Nicolas and Nick then shared the survey data with us.

Population: Although the survey was targeted at the US, Taiwan, and Hong-Kong, it was taken by 2,865 respondents from 48 countries speaking 47 languages. Therefore we classify respondents who took the Chinese survey as "Asians", since they comprise a diverse population from East Asia, and respondents of the English survey "Westerners". The sample contains 76% Westerners and 24% Asians. Around 31% of respondents are women, a rate higher than in previous MMO studies (24% in [96], 20% in [127], 14% in [131]), but probably justifiable by a constant increase in the proportion of women playing games over the years. As of 2016, 44% of gamers are women [57]. The average respondent is 28 years old ($\sigma = 9$), a number similar to those found in previous studies of MMO players [127, 131].

Survey content: The survey contains three types of questions: 1) general demographics such as age, gender, country of residence, marital and job categories, languages spoken, years of formal education, and a standard list of questions for the big 5 psychological traits; 2) WoW-specific questions such as whether the respondent's main character is in a guild, the number of years respondents have been playing WoW, and whether they ever took a break from the game or froze their subscription; 3) patterns of play include playing with real-life acquaintances, having ever made a real-life friend or partner from the game, number of years playing MMOs, defining oneself as casual or hardcore, and a sub-questionnaire about MMO play motivations. This sub-questionnaire was constructed and validated by Nick Yee [131]. In this sub-questionnaire, players rate on a 5-point Likert scale the importance of 5 game elements related to achievement, such as "Becoming powerful", 5 related to socializing, such as "Chatting with other players", and 5 for immersion, such as "Feeling immersed in the world". For each of the 3 motivation factors, their score is the average of the scores of the 5 game elements relating to this factor. We normalize each of these 3 motivation scores (ie mean of 0 and standard deviation of 1).

4.2.2 Commitment Metrics and Their Limitations

From the questions asked in the questionnaire, we pick three metrics characterizing three dimensions of player commitment. These metrics are not exhaustive, but rather a first step towards a more comprehensive model of player commitment and retention in MMOs.

The first metric is the **weekly play time**. It measures the intensity of play at the time of the survey, and is measured in hours per week, or H/w. It was obtained by asking respondents how much they play. The very few respondents with more than 120 hours per week are discarded. The distribution of the weekly play time in WoW, plotted in Figure 4.2. The average weekly play time is 23 hours per week ($\sigma = 15$), a number consistent with other

MMOs such as EverQuest and Dark Age of Camelot [112].

The second commitment metric is the **stop rate**, ie the ratio of respondents who stopped playing the game and eventually came back to play again. It was obtained by asking players if they ever took a break of 1 month, 6 months, or 1 year from the game, and if they ever froze their subscription. The stop rate is a reverse measure of players' tenacity, ie how much they (do not) stick to the game. Although the stop rate is not churn (the monthly ratio of customers who stop using a product), it is the metric closest to it. The stop rate of the whole sample is 77%: out of 100 players, 77 took a break at some point.

As shown in Figure 4.2, 40% of respondents have stopped for six months or a year before returning to the game. Among these 40%, 57% (i.e. 23% of the entire sample of the first survey) never canceled their recurring subscription. In other words, 23% of players were paying but not playing for six months or even a year! Player data remains in Blizzard's databases indefinitely even if they do not pay. Hence it is very surprising that a quarter of respondents keep their subscription active.

The third commitment metric is the number of years playing the game, or **WoW years** for short. It shows player loyalty, and whether players are early adopters or followers. When the survey was conducted in Spring 2010, WoW had been up and running for 6 years. We discard players who report playing WoW for more than 6 years. 96% of respondents have been playing WoW for more than a year, and 70% for more than three years. These numbers are consistent with another 2010 study of WoW players, which reports 94% and 65% respectively [96]. Figure 4.2 shows a histogram of the years that respondents starting playing WoW and MMOs. The spike for MMOs in 2005 suggests that WoW is the first MMO played by many respondents.

These metrics have several limitations. First, they are **obtrusive and rely on the memory** of respondents. Respondents are asked to pick a bin among 1 month, 6 months, and a year,

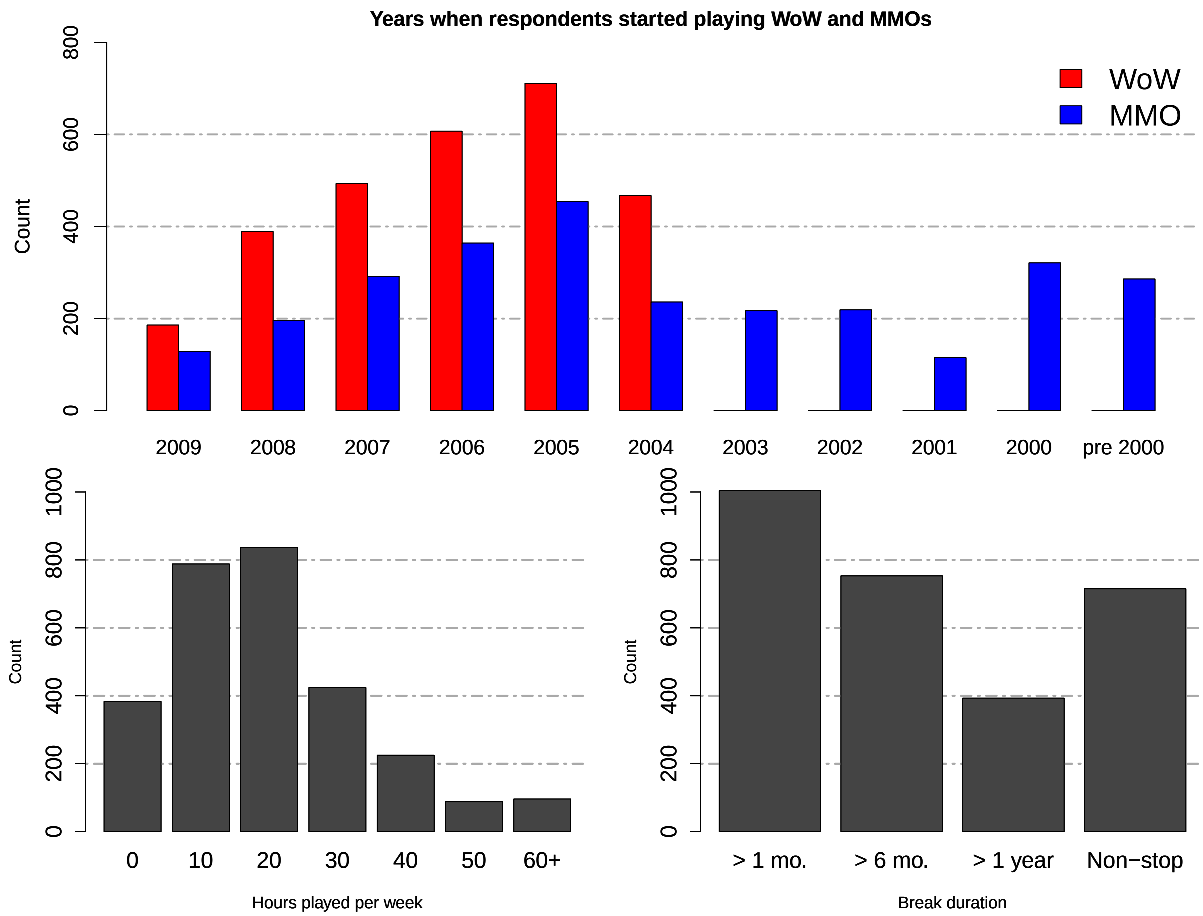


Figure 4.2: Distribution of Commitment Metrics in WoW

for the duration of their break(s). Knowing exactly and unobtrusively the length of breaks may provide additional insights about retention. Moreover, knowing why, when, or how often people stop playing is hard, and knowing why people come back to the game might be even harder. The release of new expansions can be a first phenomenon to look at; people may not play so much before expansions, as they might have explored most content or finished most dungeons. However, they tend to return to the game after expansions are released [104]. The first survey was conducted in Spring 2010, 4 months after patch 3.3 and 8 months before expansion 4.0, which seems far enough from both to suggest that respondents may belong to the more dedicated part of the player base.

A second limitation is that they only **approximate of churn**. To be able to accurately measure churn, one would have to look at a longitudinal measure of the play time, in which case churners could be those players spending less and less hours per week in the game. Churn could also be defined as the number and duration of breaks taken. Sadly, we could not ask for this in a questionnaire: people are not good at remembering how many times they did X or when they last did Y in the last 5 years.

A third limitation is the possible difficulty of **generalizing** these three metrics to other MMOs. Some metrics such as the weekly play time should be taken carefully when measuring player retention. In the case of children playing an MMO such as Club Penguin for example, parents may have more control over the player's accounts than the player herself. In this case, the weekly play time is not so much a measure of the player's will to play but rather how long the parents allow their children to play. Moreover, some metrics may be more useful than others throughout an MMO's lifecycle. For instance, counting the number of years playing an MMO that just came out does not make sense.

4.2.3 Gold Buying Metric and its Limitations

To compare players who buy gold to those who do not, respondents are asked "Have you ever bought WoW gold using real money?". In the following, we refer to the rate of gold buyers for a particular category of players as the number of respondents in that category who report having bought gold, divided by the total number of respondents in that category.

Typical gold-buying rate: Slightly more than 14% of respondents report having ever bought gold (0.5% chose to not answer that question). In his 2005 online questionnaire, Yee found that 22% of MMO players report having ever bought gold [130]. Two reasons may explain this difference in rate. First, Yee's study was conducted on MMO players at large, not on WoW players specifically. WoW players might be less likely to buy gold than the average MMO player. Second and most likely, the average MMO player may have changed since 2005: in 7 years, the number of MMO subscriptions has more than tripled [125]. Moreover, compared to the 14% of WoW players who ever bought gold as of 2010, only 10% of the Puzzle Pirates MMO players bought virtual items in 2009 [70]. Social games report between 1 to 20% of their players engaging in real-money transactions in 2010 [97, 106]. So it seems that the ratio of WoW players willing to buy virtual items is somewhat typical.

One limitation of the gold buying metric is that it **lacks breadth and depth**: In the questionnaire, respondents report if they ever bought gold or not. A more comprehensive study of gold buyers should include questions about the quantity of gold bought, how frequently, and when. The motivations behind buying gold should also be investigated. Some players may buy their gold from a friend or a relative, while others may just turn to eBay or middleman organizations who also provide power-levelling services. Player emotional responses to and attitudes toward RMT should also be the subject of future qualitative work. This exploratory quantitative work misses these subtle distinctions.

4.2.4 Dataset 2: Second Survey and Gameplay Data

Nicolas Ducheneaut and Nick Yee hosted a second online questionnaire written in English. They posted links to it on popular WoW and online gaming websites from China and the US in October and November 2011, right before patch 4.3 launched. The links posted on Chinese-speaking gaming websites pointed to a translation of the survey in simplified Chinese. Respondents who report their country to be Hong-Kong, China, or Macau are labeled as the CN respondents. As for English respondents, we only keep those who report their country to be the US.

Survey content: The questionnaire asks respondents demographic questions such as their age and gender, as well as WoW-specific questions such as the number of years they have been playing the game for, and whether they ever stopped playing the game for at least a month. Yee's 15 gameplay motivation questions give us the achievement, immersion, and social scores for each participant. The questionnaire also asks respondents the names and servers of their active characters. This allows us to match their survey data with data generated by their game characters.

Clean-up: Since the gameplay data collection starts on December 1, we only keep participants who took survey on or before November 30, 2011. We also select players who answer all questions about their age, gender, number of years playing WoW, and motivations. To be able to pair gameplay data with player data, we select players who logged in with at least one of the characters they listed between December 1 and June 30. Finally, we ignore character "resets", ie characters who were deleted during data collection, re-created with the same name, and leveled-up again.

Overall statistics: After cleanup, the sample consists of 1,350 players, among which 29% are women, and 41% from the CN region. The average CN respondent is 23 years old, while the average US respondent is 35 years old. Both the average CN and US respondent have been playing WoW for close to 5 years as of November 2011. We note that the players who frequent the websites on which the survey was advertised may be more expert and dedicated than the average player. However, our sample matches other samples found in previous works in terms of age and gender statistics [127, 131].

Overall, 4,389 characters logged in at least once during the 7 months. Distinguishing by character level, 65% of all characters have already reached the maximum level of 85 as of December 1st, 2011, 7% reached it sometime between December and June, and 28% did not reach it by the end of June. These numbers show that two thirds of the active characters are in the end-game phase of the game, and able to raid FL and DS.

Longitudinal gameplay data: Character data is exposed by the Armory, a Blizzard web service connected to the WoW database[2]. Character data was pulled every day from December 1, 2011 to June 30, 2012. While the Armory API provides hundreds of measurements at the character level, we focus on three monthly aggregates at the player level. First, we measure the number of characters used by each player during the month. In the rest of the paper, we refer to the *active player base* of a particular month as the sample of players who used one or more characters during that month. We refer to the *total player base* as the 1,350 players. Second, we define the number of raids that each player participates in during a month as the sum of all the 5-, 10-, and 25-player raids entered by all her characters that month. The Armory API does not provide the number of "Looking for Group" raids entered, so the number of raids does not take this number into account. And finally, we define the number of PvP occurrences that the player participated in during the month as the sum of all the duels, arenas, and battlegrounds that the player's characters participated in. Zone-wide PvP battles such as Wintergrasp are not returned by the Armory API. Table 4.2 lists the 11 variables collected in this second dataset.

[2]The documentation of the Armory API lives at `http://blizzard.github.io/api-wow-docs/`

Domain	Features
Demographic	Age, gender, region
WoW-specific	Number of years playing WoW, whether they stopped before
MMO play motivations	Achievement score, immersion score, social score
Monthly game data	Number of characters played, PvE raids, and PvP occurrences

Table 4.2: Variables collected in the second WoW dataset

4.3 Commitment Segmentation

Players of an old MMO like WoW have had opportunities to leave the game, come back, and leave again as expansions are added to the game. This section relies on dataset 1, the first online survey conducted in Spring 2010, to explore various patterns of churn between different player segments. It is important to note that players can pay for their subscription manually every month, per semester, or yearly. Players can also choose to be billed automatically every month, and they can stop being charged automatically at any time. It is financially interesting for Blizzard to keep player subscriptions going as long as possible, even if they are not playing anymore.

As shown earlier in Figure 4.2, 35% of respondents stopped playing for at least a month, 26% for at least six months and 14% for at least a year. A one-sided t-test shows that the 25% who never stopped playing started playing WoW more recently (2.9 versus 3.7 years, $p < .01$).

In the rest of this section, we break down each row of Table 4.3.

4.3.1 Region and Gender

The first dataset shows large differences between Asian and Western players in terms of churn behavior. Asians play four more hours per week on average, yet they are more likely to stop. Asian and Western players started playing WoW around the same time. Table 4.3

	N	H/w	Stop rate	Years playing
Western players	2183	22	75%	3.6
Asian players	673	26	85%	3.4
Women	875	23	68%	3.4
Men	1981	23	81%	3.6
Casual	819	17	83%	3.4
In-between	1792	24	76%	3.6
Hardcore	243	32	67%	3.7
Not in a guild	277	19	88%	3.4
Guild member	1569	22	79%	3.5
Guild officer/leader	1006	24	71%	3.7
Made real-life friends in game	1530	24	78%	3.9
Did not make RL friends	1319	21	76%	3.2
Met RL partner IG	378	28	84%	4.0
Did not met RL partner	2485	22	76%	3.5
Single/divorced	1441	24	81%	3.7
Play with partner	826	23	71%	3.6
Play without partner	581	19	75%	3.8
Has no children	1800	23	100%	3.5
Play with children	127	21	56%	3.7
Play without children	435	19	74%	3.9

Table 4.3: Average retention metrics for different segments of WoW players.

compares Asian and Western players in terms of the three retention metrics. Looking at gender, two one-sided t-test indicate that women are not playing significantly more per week than males, yet they are less likely to stop (23 hours per week for both, $p < .01$). Men tend to be earlier adopters than women.

4.3.2 Hardcore-Casual Dichotomy

Another way to segment players is the hardcore-casual dichotomy. Casual players tend to prefer easier games, while hardcore players generally look for efficient strategies to beat the game [11, 121]. In the first survey, 29% of respondents consider themselves "casual", 8% "hardcore", and 63% "in-between". Hardcore players are less likely to stop than in-between players, who are themselves less likely to stop than casual players ($c^2(2, 2854) = 6.73$, $p < .01$). Moreover, players who label themselves as hardcore tend to spend more time in the game per week ($r(2828) = .19$, $p < .01$).

4.3.3 Play Motivations

The hardcore/casual dichotomy is a very simple player typology. Players are either one or the other. Yee's motivation model is more nuanced [132]. It considers the extent to which players are motivated by three factors: achievement, socializing, and immersion. For each factor, we bin respondents with a (non-normalized) score between 1 and 2 in the [1,2] bucket, between 2 and 3 in the (2,3] bucket, and so on. The weekly play time is then averaged for each bin of each factor. Figure 4.3 illustrates the relationship between these three motivation factors and the weekly play time. For both immersion and achievement, players in the highest bin play more per week than players in the lowest [1,2] bin.

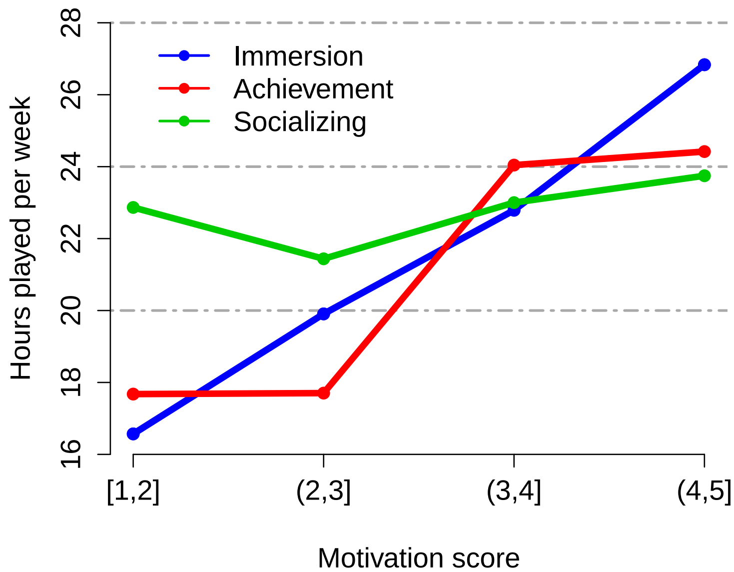


Figure 4.3: Average WoW weekly play time for the three motivation factors.

4.3.4 Guild Position

Players socialize mostly in guilds [56, 93]. Previous studies show that players in guilds spend more time per week in the game [55, 102]. Our first survey goes beyond a simple in-guild/not-in-guild dichotomy and asks players the position of their main character in a guild. Only 10% of respondents do not belong to a guild, 55% are regular guild members, and 35% are guild officers or guild leaders. A one-way ANOVA using guild position as the independent variable and weekly play time as the dependent variable indicates that a higher guild rank leads to spending more time per week in the game ($F(2, 2826) = 12.90$, $p < .01$). Another ANOVA shows that players with more guild responsibilities have been playing WoW for slightly longer than basic guild members, who themselves have been playing for slightly longer than players whose main character is not in a guild ($F(2, 2840) = 10.88$), $p < .01$). Yet the stop rate decreases as prominence in a guild increases ($c^2(2, 2852) = 9.27$, $p < .01$).

4.3.5 Playing With One's Partner

Approximately 13% of respondents report having met someone who became a boyfriend, girlfriend, or spouse while playing WoW. Those respondents play longer per week, and are also earlier adopters. Surprisingly and similarly to those respondents who made real-life (RL) friends from the game, RL partners met in the game do not seem to decrease the stop rate. Women are twice more likely to meet a RL partner in the game than men (20% of women versus 10% of men). Although some players manage to have met someone who became a real-life partner, some actually do play with their partner. We denote as "partnered" the respondents who report being engaged or married. An ANOVA using the weekly play time as dependent variable shows that players who play with their partner play slightly less per week than single players, but four more hours than players not playing with their partner ($F(2, 2831) = 29.00$, $p < .01$). They are also less likely to take a break ($c^2(2, 2848) = 6.94$, $p < .01$).

4.4 Gold Buying Segmentation

Since it only takes a couple months to reach maximum level, the bulk of WoW does not consists of amassing XP, but rather equipment and gold. This is called the end-game. The best end-game equipment comes from defeating difficult boss monsters. This requires a large time commitment to a raiding guild, from 10 to 30 hours per week. Raid equipment is usually bound on pick-up: once players loot it from a monster, they cannot trade it with other players. Gold comes mostly from selling or re-selling on the auction house the equipment, crafting materials, and other items that are not bound on pick-up. Based on our experience of the game, we hypothesize that players with little time to raid buy second-best equipment from the auction house for gold. But since they have little time to play, they do not have much gold. So they buy game gold with real money from third-party websites.

Effectively, they trade game time for real money.

The WoW terms of use stipulate that players "may not sell in-game items or currency for real money, or exchange those items or currency for value outside of the game"[3]. Buying gold with real money is against the terms of use, and can result in players being banned. Yet, the demand for gold is high. When Blizzard banned 50,000 gold farmer accounts in 2006, the cost for 100 gold went from \$6 to \$35 [46]. And to cut down on gold sellers, Blizzard launched in April 2015 the WoW token, which allows players to purchase game gold with real money[4]. In 2010, when this study was conducted, the token did not exist yet.

In this section, we compare the ratio of gold buyers for several player segments. The weekly play time is included to check whether players do buy gold because they have too little time to play. We also suspect age to be a confounding factor, since older players may have more disposable income to spend. We finish this section with a simple model combining all segments together to identify the variables that best explain buying gold.

4.4.1 Demographics

Table 4.4 breaks down age, weekly play time, and proportion of gold buyers for all demographic segments under consideration. Overall, men report having bought gold twice more often than women ($c^2(1, 2841) = 31.11$, $p < .001$). Looking at regional differences, Asians seem as likely to buy gold as Westerners ($c^2(1, 2850) = .13$, $p = .93$).

Focusing on job status, 17% of full-time employees buy gold, whereas at most 12% of part-time employees, students, home-makers and unemployed respondents ($c^2(5, 2499) = 18.12$, $p < .001$). This confirms that respondents with more disposable income, such as full-time employees or retired people, are more likely to buy virtual gold. A one-way ANOVA us-

[3]See `http://us.blizzard.com/en-us/company/legal/wow_tou.html`
[4]See `https://us.battle.net/shop/en/product/world-of-warcraft-token`

	N	Age	H/w	Gold buyers
All	2865	28	23	14%
Asian women	136	25	26	5%
Western women	739	32	23	9%
Asian men	537	23	26	16%
Western men	1444	29	21	17%
Home-maker	91	34	26	7%
Student	733	22	23	12%
Part-time employee	406	27	25	12%
Unemployed	226	27	30	13%
Full-time employee	1360	32	20	17%
Retired	22	53	35	32%
Now in school/college	625	21	24	11%
College graduate/dropout	1865	32	22	15%
High school grad/dropout	350	24	28	21%
Play with real-life ties only	888	28	21	12%
Neither RL nor IG	404	29	21	14%
Both RL and IG	1251	28	24	15%
Play with in-game ties only	312	28	23	21%

Table 4.4: Average features and ratio of gold buyers for different WoW segments

ing the weekly play time as the dependent variable and the job category as the independent variable shows that full-time employees are generally older than full-time students ($F(5, 2835) = 258.93$, $p < .001$). Moreover, they play fewer hours per week than other categories ($F(5, 2823) = 22.83$, $p < .001$), which again fits the picture of players buying gold when they do not have enough time to play.

The retired segment has the highest rate of gold buyers, at 32%. The average respondent in this segment is 53 years old, and more likely than other segments to play with their children ($c^2(5, 2853) = 105.66$, $p < .001$). Thus it could be that retired respondents buy gold for their children or grand-children. However, the very small size of this segment ($N = 22$) cautions against reading too much into these results.

4.4.2 Play Motivations

Immersion: Buying gold is very weakly inversely correlated with immersion ($r = -.08, p < .001$). Looking at differences across gender, men and women are equally motivated by immersion, with motivation scores of 3.4 and 3.5 respectively ($t(874) = 2.55$, $p = .005$). However, the ratio of women buying gold decreases sharply with the immersion motivation score, from 18% for scores below 2 ($N = 34$), to 6% for scores above 4 ($N = 319$). As seen in Figure 4.4, the ratio of men buying gold only dwindles from 19 to 14% for those same categories ($N = 135$ and $N = 643$ respectively). Men with immersion scores above 4 are twice more likely to buy gold than their women counterpart ($c^2(1, 962) = 11.5$, $p < .002$).

Achievement: Buying gold is weakly correlated with achievement ($r = .10$, $p < .001$). Men are more motivated by achievement than women: their average score is 3.6 against 3.2 for women ($t(874) = 13.6$, $p < .001$). Looking at respondents with an achievement score above 4, 21% of men buy gold ($N = 743$), but only 10% of women do ($N = 165$) ($c^2(1, 908) = 8.10$, $p < .007$). Note that among respondents with achievement scores below 2 (both $N = 90$), only one man and one woman report buying gold.

Social: Buying gold is not correlated with socializing ($r = .05$, $p < .001$). Socializing in WoW happens largely in guilds [56, 92]. The ratio of gold buyers does not differ significantly between guild-less players, guild members, and guild officers or leaders ($c^2(2, 2839) = 2.69$, $p = .21$). However, the last row of Table 4.4 shows that the ratio of gold buyers varies with the nature of the social ties. Around 75% of respondents play with real-life ties such as a partner, relative, friend, or colleague. Moreover, 55% of respondents have made in-game ties, ie other players met in the game whom they now consider real-life friends or partners. Only 12% of players who play only with real-life ties buy gold, compared to 21% for those who play only with in-game ties. In-between, players with both real and virtual ties are nearly as likely to buy gold as players without ties at all ($c^2(3, 2842) = 14.16$, $p < .001$).

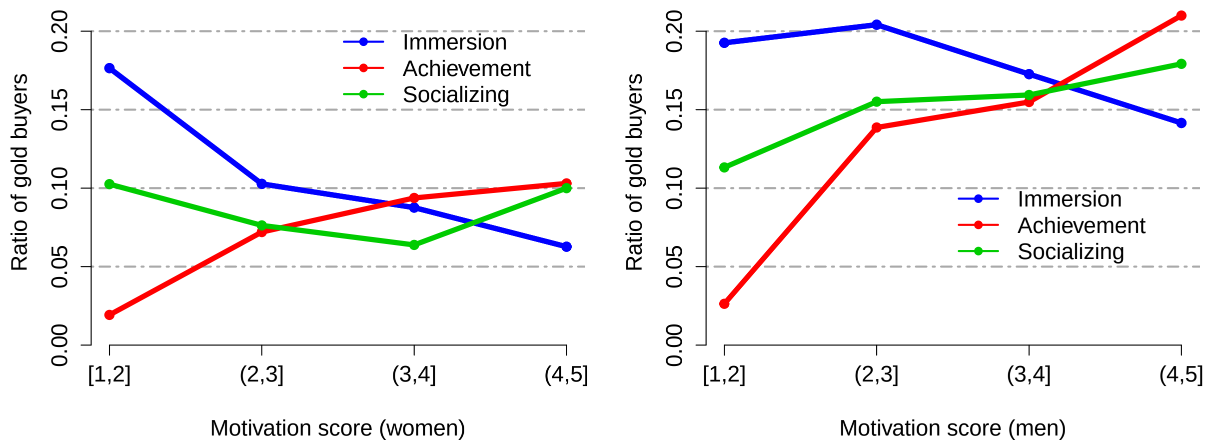


Figure 4.4: Ratio of gold buyers against the three motivation scores and across genders.

4.4.3 Links With Commitment

As detailed in the previous section, the first survey provides three metrics measuring commitment: weekly play time, number of years playing WoW, and having taken a break of at least a month, 6 months, or a year. Buying gold is not quite correlated with the weekly play time ($r(2825) = .05$, $p < .001$). Earlier WoW adopters are slightly more likely to buy gold than more recent players ($r(2839) = .11$, $p < .001$). This difference may be due to the fact that earlier adopters are also more likely to take a break from the game ($r(2852) = .24$, $p < .001$).

Players who have taken breaks are more likely to buy gold than players who never stopped (16% vs 10%, $c^2(1, 2850) = 14.31$, $p < .001$). Figure 4.5 shows that Western players who take longer breaks are more likely to buy gold ($c^2(3, 2177) = 22.35$, $p < .001$). This does not seem true for Asians ($c^2(3, 671) = 2.73$, $p = .37$). Possibly, when Western players who have taken a break come back, they may feel a little behind compared to other players and may be tempted by a gold bonus.

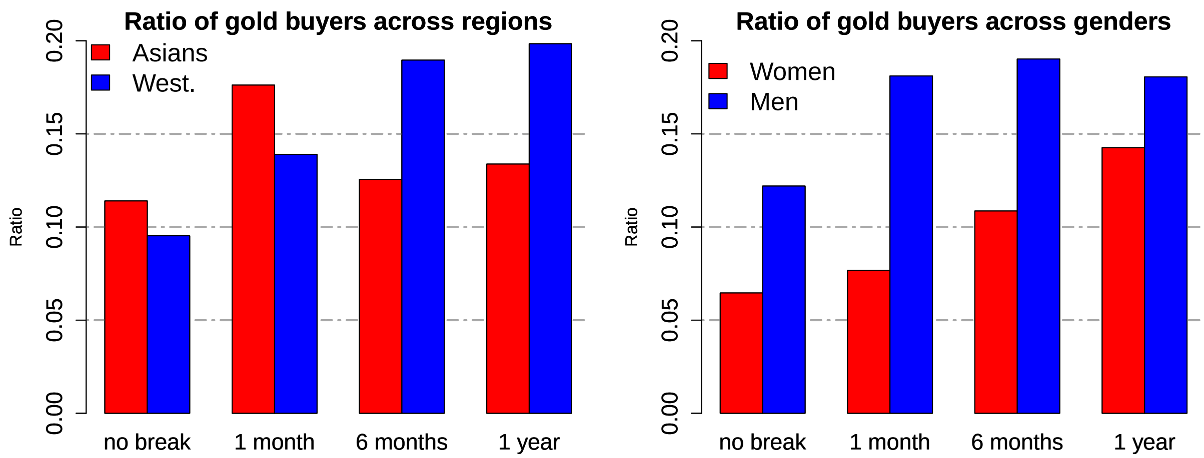


Figure 4.5: Ratio of gold buyers against duration of breaks, across regions and genders.

4.4.4 Generalized Linear Model

Figure 4.6 provides a summary of the significant and sizeable correlations found so far. Each node represents a variable. The value on the edge between two variables indicates the Pearson correlation coefficient r between them. For nodes other than buying gold, only r values above .10 in absolute value are displayed. Edge thickness reflects correlation strength. Demographic variables are purple, play motivations blue, and MMO-related in green.

Most variables are correlated with at least four other variables. This makes it difficult to take away any crisp finding with just t-tests and correlations. Moreover, series of comparisons can reveal statistically significant differences that could be attributed to chance only. Several methods exist to reduce the emergence of significant results due to chance only. Sequential Bonferroni corrections, for instance, adjust the required significance level with the number of comparisons performed [107]. One of the drawbacks of Bonferroni corrections, though, is their zealous dismissal of results when the number of comparisons to execute gets large, which is our case.

Another method involves using more advanced statistical tools such as a generalized linear model (GLM). GLM is more rigorous than a series of simple tests because it considers all the

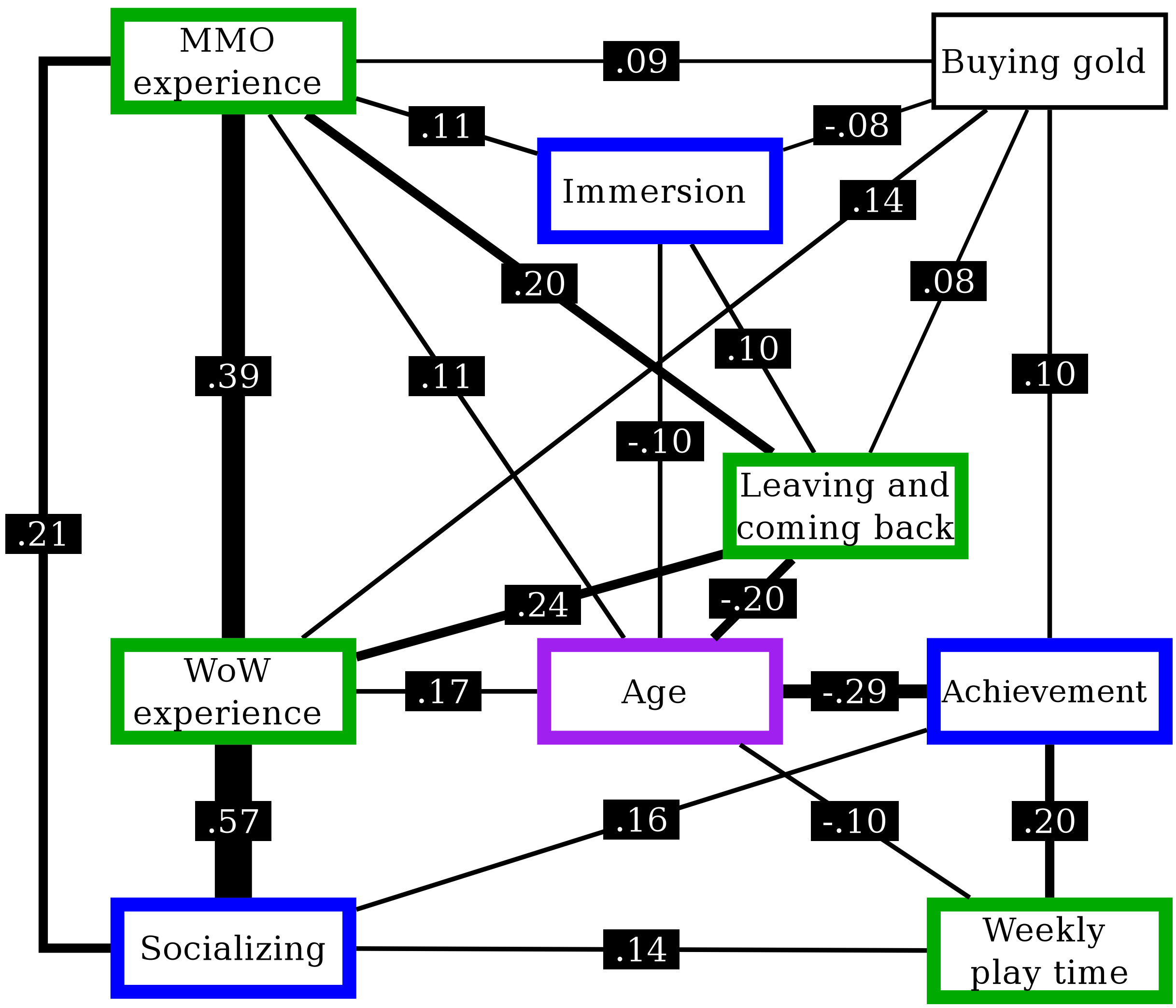


Figure 4.6: Correlation network between variables from the first WoW survey.

variables at once. GLM returns the variables most likely to predict an outcome of interest, such as the likelihood of buying gold. We use multivariate logistic regression as packaged in R with the command `glm`. Since buying gold is a binomial outcome (yes or no), we add a binomial family parameter to the `glm` command.

Preparing the data: We remove from the analysis respondents who chose not to answer one or more questions, since the model would try to predict the respondents with missing values as respondents in a fictitious "missing-value" category. Some variables contain too few respondents in some of their categories. For example, there are only 84 respondents still in high-school. Therefore, we collapse similar categories to make more precise claims. For instance, we group respondents still in K12 and high-school graduates together, and college students and college graduates together. Since logistic regression only accepts numerical or binomial variables, categorical variables like education (originally with 6 different categories: K12, currently in high-school, currently in college, and so on) have to be transformed into binomial ones (high school or less = 0, college or more = 1). Some variables with too small categories can not be collapsed, since the result would not mean anything. For instance, only 22 respondents are retired, and 91 home-makers, but each category is very different from the other. The variable "country of origin" also contains too many answers with too few respondents (2 respondents from Macau, 3 from Malaysia, and so on). Therefore we have to leave those variables out of the model.

Limitation: collapse of categories: When preparing the data for the generalized linear model, we collapsed categories such as education levels of K12 and high-school graduates together because of the too few respondents in them. However, we could have collapsed categories differently: K12, high-school, and college students together in one group, and high-school and college graduates together in another group. Our collapsing choices may have introduced noise in the analysis.

Screening variables: Given our previous analyses, gender seems to strongly influence the likelihood of buying gold, and is unlikely to cause most of the other variables. Therefore, we take gender as our main predictor. To prevent interactions between gender and other variables, we test whether each variable is correlated with the outcome (buying gold) or the main predictor (gender). For instance, a one-tailed t-test shows that the respondents who buy gold are significantly older than those who do not ($t(411) = 1.93$, $p = .027$), and another t-test shows that women are significantly older than men ($t(874) = 9.15$, $p < .001$). Therefore, age is a **confounder**: it is related to both buying gold and gender, and may confound the relation between the outcome and the predictor, so we include it in the analysis. The weekly play time, however, is related to buying gold ($t(411) = 2.53$, $p = .006$) but not to gender ($t(874) = .92$, $p = .178$). Hence the weekly play time adds **precision** to the analysis, and we also include it. As another example, the marital status is related to gender ($c^2(1, 2836) = 42.44$, $p < .001$) but not to buying gold ($c^2(1, 2830) = .02$, $p = .83$). Thus it would add **noise** to the analysis, and has to be left out. Variables unrelated to both gender and buying gold are also left out.

Building the model: We then run the `glm` command in R with the selected variables. The model shows that gender, our main predictor, still has a significant effect ($p < .003$) on buying gold when controlling for all other variables. We then proceed to remove all the variables with insignificant effect. For instance, taking a break of 1 month ($p = .53$), 6 months ($p = .29$), or a year ($p = .55$) are insignificant and taken out. However, having ever frozen one's subscription is significant ($p = .013$) and kept in. In the end, the 17 variables that significantly influence buying gold are summarized in Table 4.5.

Interpreting the results: The `glm` command returns the odds for each of the 17 variables. For categorical variables such as gender, an odds ratio of 1.696 means that controlling for all other variables, the odds of buying gold are estimated to be 69.6% higher for men compared to women. For numerical variables such as age, the odds can be interpreted as follows: between two groups that differ 1 year in age but that are equal on all other variables, the

Variable	p-value	Odds
Having played on a private server	.000	1.830
Gender (Being a man)	.000	1.696
Having frozen subscription	.001	1.477
Made in-game friends	.123	1.202
Playing with spouse	.039	0.690
College education	.014	0.684
Playing with sibling	.019	0.614
Playing with cousin	.103	0.512
Achievement score	.000	1.383
Years playing WoW	.000	1.283
Extraversion score	.045	1.142
Age	.000	1.040
Weekly play time	.100	1.006
Socializing score	.067	0.853
Agreeableness score	.048	0.850
Immersion score	.001	0.817
Conscientiousness score	.003	0.800

Table 4.5: Logistic regression model explaining gold buying.

odds of buying gold are estimated to be 4.0% higher for the older group. In other words, controlling for all other variables, the odds of buying gold increase 4.0% per year of age. Similarly, the odds of buying gold decrease by 19.3% for each point of immersion score.

Summary: Using a more advanced statistical model, gender and play motivations remain strong predictors of buying gold. Job category and break duration are replaced by college education and having frozen one's subscription. Playing with real-life ties such as cousin, sibling, or spouse, remain strong predictors for not buying gold.

4.5 Progression and Churn

Blizzard adds new content to WoW in the form of expansions and patches released every 4-8 months. Expansion 4.0 called "Cataclysm" was released in December 2010. It raised the maximum level of a character to 85. Blizzard launched a series of three content patches for Cataclysm. Each patch starts a new PvP season, which resets the PvP ratings of all characters to zero. Each patch also introduces new end-game PvE dungeons for maximum-level characters to raid. Small dungeons only accept groups of 5 players. Larger dungeons accept groups of 10 or 25, and have a normal and heroic difficulty. In heroic difficulty, monsters have more health points, inflict more damage, and drop more powerful equipment. To try a dungeon in heroic difficulty, players must first complete it in normal difficulty. Each boss can only be defeated once per week for each difficulty and each group size.

Patch 4.2 launched in June 2011. It started PvP season 9, and added the Firelands (FL) dungeon. FL contains 7 bosses: Alysrazor and three others are outdoors and can be approached in any order, then Baleroc, Staghelm, and Ragnaros must be defeated in that order. FL bosses dropped the best character equipment in the game until the release of patch 4.3.

Patch 4.3 launched on November 29, 2011. It started PvP season 10, and introduced the Dragon Soul dungeon (DS). DS bosses drop even better equipment than FL bosses. DS features 8 sequential boss encounters, with Morchok the first and easiest, Ultraxion the fifth and of medium difficulty, and Madness of Deathwing the last and most difficult. Patch 4.3 also added on-demand raiding with random players through a "Looking for Group" (LFG) difficulty. DS remained the hardest dungeon in WoW until expansion 5.0 in September 2012. Starting at the end of January 2012, Blizzard reduced the damage of all DS bosses in normal and heroic difficulty by 5% per month until its cap at 30% in July 2012[5].

This section uses dataset 2, ie the survey data collected just before patch 4.3, and the gameplay data collected for 7 months during patch 4.3, to answer the following questions: Is there a relationship between difficulty and churn? Which player segments are more adverse to difficulty? Can we predict presence from one month to the next? How much do players replay older content such as FL? How does the player base change over several months?

[5] Although players could opt out of this damage reduction, the Armory API did not provide this information. Since there is no incentive to opt out, we suspect that most players benefited from it.

	CN		US	
	women	**men**	**women**	**men**
N	84	469	313	484
characters played	1.5	1.3	3.0	2.5
raids 5	18.3	16.5	16.2	13.1
raids 10	4.1	3.6	3.0	2.4
raids 25	6.1	6.2	4.5	4.7
Total PvE raids	28.6	26.3	23.6	20.2
duels	2.1	4.2	0.4	0.5
arenas	8.9	8.1	2.2	2.6
battlegrounds 10	0.5	0.5	0.2	0.1
battlegrounds 40	1.4	1.9	1.2	1.8
Total PvP occurrences	12.9	14.6	4.0	4.9

Table 4.6: Stats of the average player over 7 months, split by region and gender.

4.5.1 Segments

Table 4.6 provides the average number of characters played over the 7 months across genders and regions. It also breaks down the total number of raids and the total number of PvP occurrences participated in over the 7 months. The average CN player uses fewer characters, PvEs as much, and PvPs three times more often, than the average US player. CN players seem more competitive and more focused on a single character than US players. Women seem more inclined to PvE and less inclined to PvP, but these gender differences are smaller than the regional differences.

For a given month, we define the ratio of active players as the size of the active player base divided by the size of the total player base. In December 2011, the active player base represents 93% of the total player base, ie the ratio of active players is 93%. This percentage decreases gradually to reach 64% in June 2012. In other words, a third of the player base stops playing after seven months. Figure 4.7 plots the ratio of active players over time, across regions and genders. CN men seem to churn twice as fast as US women, with CN women and US men somewhere in-between. We distinguish three overall phases: first a loss of 15% from December to February (5% churn per month), then somewhat flat until April

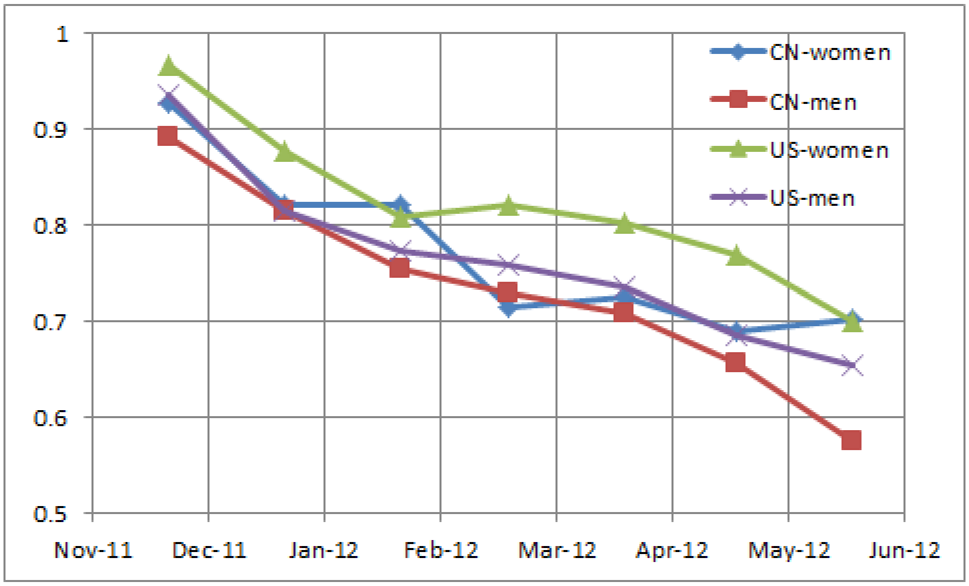


Figure 4.7: Ratio of active players, split by region and gender.

(2% per month), and a final drop of 10% until June (again 5% per month). Justifying why these patterns occur is difficult. While the first drop may reflect the speed at which expert players are done with new game content and leave the game, the final drop may only reflect the seasonal effect of the Summer holidays.

Around 82% of the total player base report having ever taken a break of at least one month before November 2011. These players churn at nearly 5% per month. On the other hand, the players who report having never taken a break before churn at 2% per month on average. Figure 4.8 illustrates these numbers. Several explanations are possible. It could be that players who never stopped before started playing the game more recently. Beginners have more things to do in the game, such as leveling up and exploring, than veterans, so they churn slower. Another explanation is that there are two segments of players: the more transient stops more often and churns faster than the more stable.

Summary: CN players churn faster than US players. Gender and age make no difference. 15% of the player base churns in the first three months.

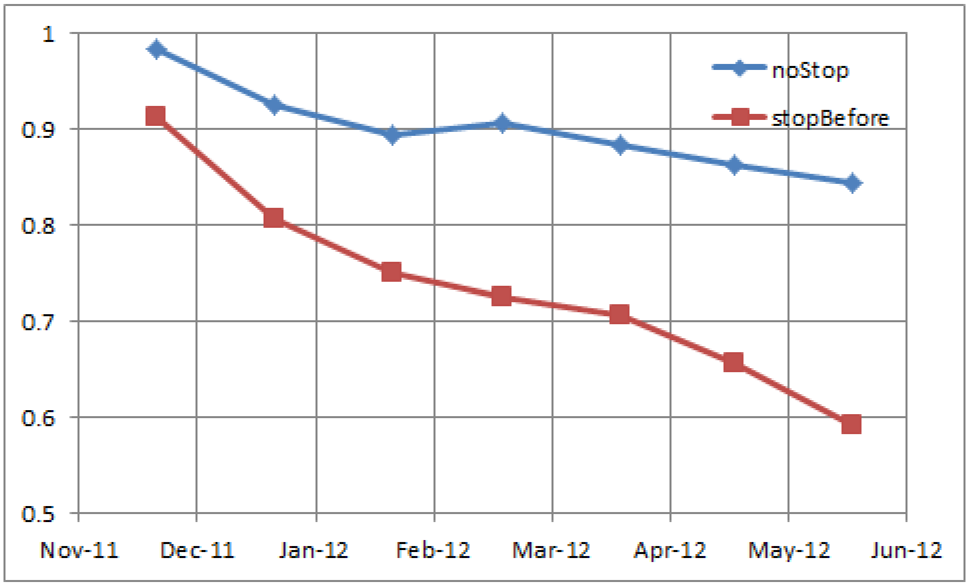


Figure 4.8: Ratio of active players, split by "having taken a break before November 2011".

4.5.2 Raid progression and replays over time

Dragon Soul contains eight boss monsters to be defeated sequentially. Morchoc is the first and easiest, Ultraxion fifth, and Madness of Deathwing last and most difficult. Figure 4.9 plots the ratio of active CN and US players who killed these bosses at least once for each month.

In both regions in December, a third of the active player base manages to kill Morchok, 20% Ultraxion, and 7% Madness in normal difficulty. These percentages may seem low, but the Armory API only returns the successes at defeating a boss, not the failures. Since groups fail 10 times before succeeding [9], the rate of players who actually tried defeating normal difficulty bosses may be much higher. Moreover, the ratio of players defeating a particular boss first increases, then peaks, and finally falls in both regions. In heroic difficulty, these peaks may happen after June. Unsurprisingly, the more difficult the boss, the later it peaks. For example, in normal difficulty, Morchok peaks in January, Ultraxion in March, and Madness in April.

We identify three major differences between US and CN players. First, all normal bosses, as well as heroic Morchok, peak around 40% in CN. For US players, the peaks decrease with the boss difficulty (or with time): 40% for normal Morchok, 35% for normal Ultraxion, 30% for normal Madness, and 23% for heroic Morchok. Second, active US players defeat more and harder bosses in December than CN players. Moreover, the slope of a boss indicates how fast the active player base manages to defeat it. These slopes are steeper for CN than for US players. Thus US players start better, but progress slower. And finally, normal Morchok drops abruptly after March for CN players, but more smoothly for US. In June, nearly 25% of US players defeat it, versus only 12% of CN players. Since Blizzard reduces the damage of DS monsters by 5% every month starting in January, difficulty should not be an issue. It could be that US players find more replay value in re-defeating easier bosses.

Looking at the raiding behavior in the FL dungeon tells us how much players enjoy replaying it. Players may raid FL for gold, equipment (even though DS bosses drop better equipment), to gain an achievement, or just for fun. With the better equipment dropped in DS and sold in the auction house, FL should be easier in December than in November. Therefore, and as shown in Figure 4.10, it is not surprising that around 30% of CN and US players manage to defeat normal Baleroc in December. In both regions, fewer and fewer players defeat normal FL bosses over time.

The ratio of players defeating heroic Alysrazor and Baleroc rises from 5 to 10% for CN players, but stays flat at 3% for US players. Similarly, the ratio of players defeating heroic Ragnaros rises from 0 to 5% for CN players, but remains at 0% for US players. These findings complement what we previously found for DS: CN players find more replay value in challenging raids than US players.

Summary: Nearly half of the active player base raid every month. The number of raids entered by active players is halved in seven months. CN players take on more challenging bosses than US players.

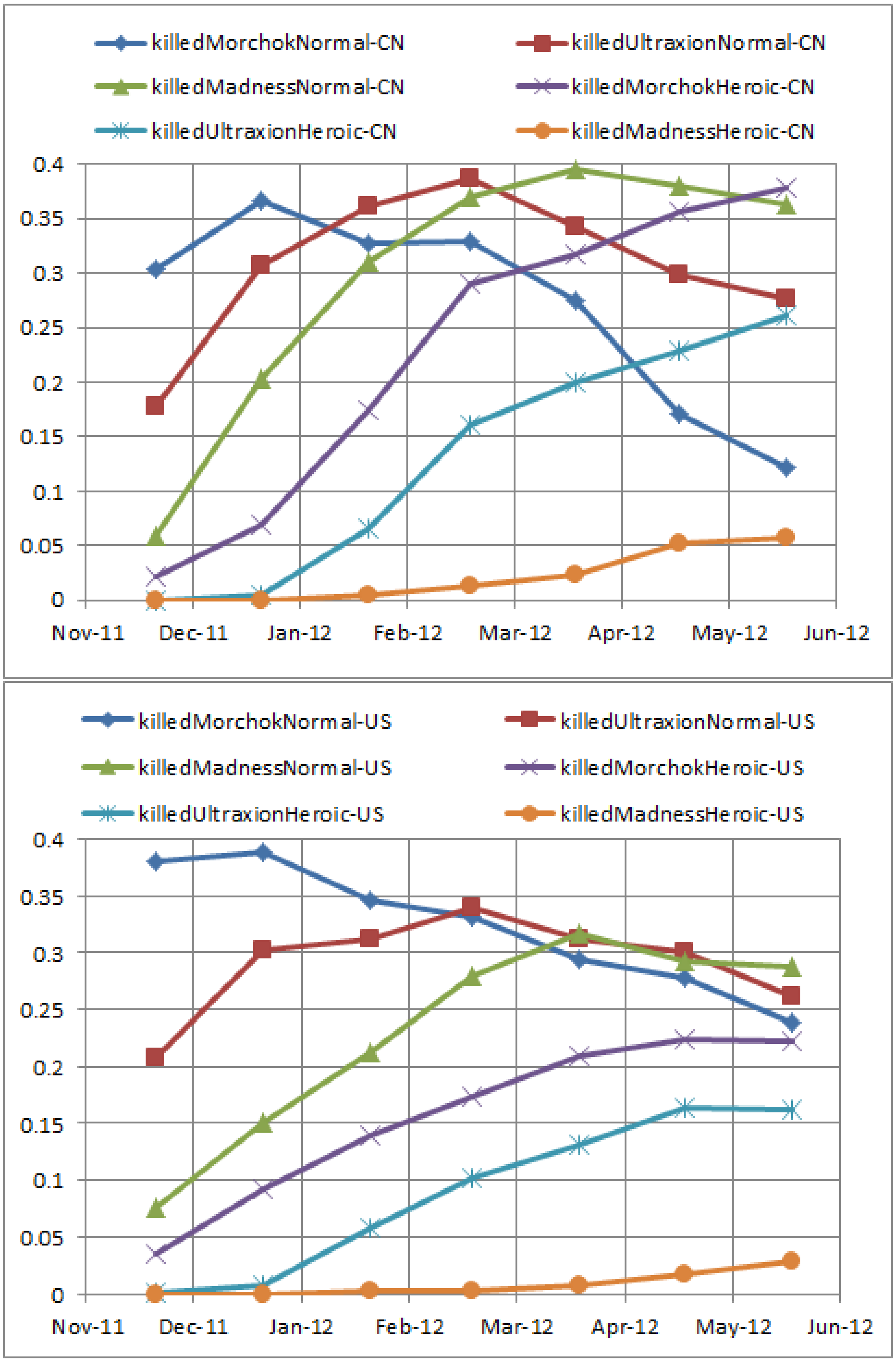


Figure 4.9: Active player base defeating each boss in Dragon Soul, split by region.

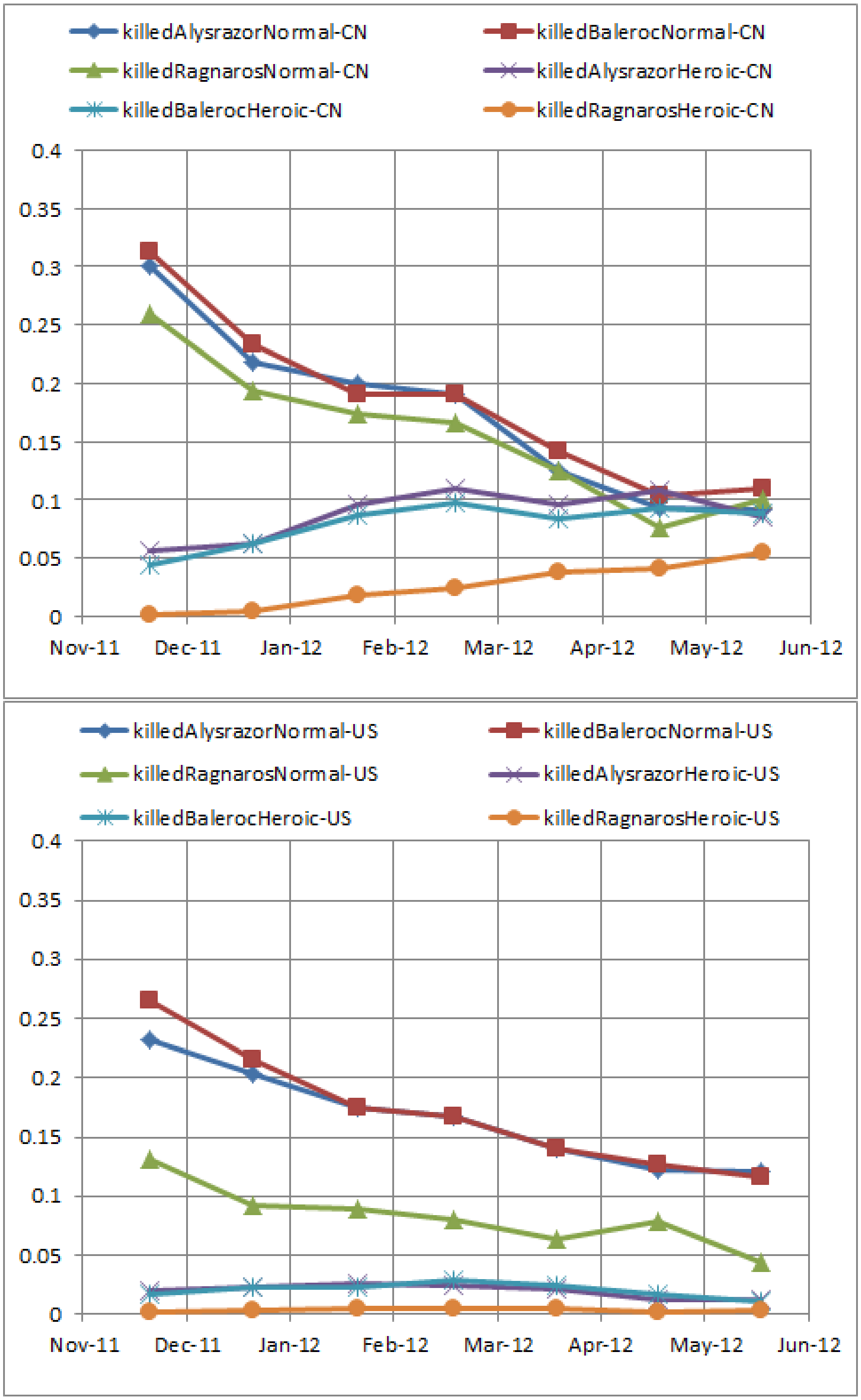


Figure 4.10: Active player base defeating each boss in the Firelands, split by region.

4.5.3 Explaining presence and churn

MMO companies may want to explain why, or predict if, a player will play next month. The naive approach focuses on presence: what makes a player login next month or not. We show that this problem is relatively trivial, and that the more challenging problem is churn: among the active player base, which players will leave the game?

Predicting presence: For each player, dataset 2 provides 8 survey variables: age, gender, region, number of years playing WoW, having taken a break, and normalized achievement, immersion, and social motivation scores. For each player, dataset 2 also provides 3 game-play data aggregates: number of characters played, total PvE occurrences, and total PvP occurrences. The baseline model is the simplest: a player will play next month if she played this month. We put ourselves in the place of Blizzard in January 2012. Applying this simple model using the January data of the total player base to predict the February presence gives 89% precision and 95% recall. These precision and recall set the bar very high. Using the May data to predict presence in June gives similarly high precision and recall.

The second and more complex model is a step-wise multivariate logistic regression. The regression takes as input a train set containing the 11 variables of December about each player. January presence is the label to predict. For each variable, the regression outputs a log of the odds ratio (log OR) for two groups differing in one unit of that feature, and a significance level. If the log OR is positive (negative), it means the feature is positively (negatively) correlated with presence the next month. During the testing phase, when given the data of a particular player as an input, the model will output a number between 0 and 1. The closer to 1, the more confident the model is that the player will play the following month. We keep the cutoff at the default value of 0.5: the model considers that a player with a value output of 0.5 or above will play the following month.

The resulting model is described in the left-most column of table 4.7. It achieves 91%

Variable	Jan. Presence logOR	p	May Presence logOR	p	Jan. Churn logOR	p	May Churn logOR	p
chars. played	.44	<.001	.87	<.001	-.25	<.001	-.30	<.001
num. raids	.05	.<.001	.10	<.001	-.05	<.001	-.09	<.001
region is US	-.36	.06	-.28	.08	.77	.001	–	>.1
stopped before	–	>.1	-.50	.04	–	>.1	1.10	.004
num. PvP	–	>.1	.025	.07	–	>.1	–	>.1
achiev. score	–	>.1	-.14	.08	–	>.1	.18	.10
age	–	>.1	–	>.1	-.05	<.001	–	>.1
other 4 vars	–	>.1	–	>.1	–	>.1	–	>.1

Table 4.7: Log odds ratios from regression models explaining presence and churn.

precision and 93% recall, which is not much better than the baseline. Only three variables are significant and kept in the model. The first feature is the number of characters played in December: the odds of playing in January are $exp(.44) - 1 = 55\%$ higher per character played in December. Put more simply, the more characters played in December, the more likely to play in January, independently of the ten other variables. Similarly, more raiding in December means more chance to play in January. And US players are $exp(-.36) - 1 = 30\%$ less likely to play in January than CN players. None of the remaining eight features are significant, and are therefore excluded from the model. Similarly, the model predicting May presence has 6 significant predictor variables. It achieves 90% precision and 87% recall, which is good, but again comparable to the baseline.

Transfer behaviors: To understand why the regression model gains nothing compared to the baseline model, we need to break down what precision and recall mean in the baseline model. When predicting February presence in January, each player can fit in only one of four transfer behaviors: either 1) they play in January and will play in February (they will "stay in"), or 2) they play in January and will not play in February (they will "churn"), or 3) they did not play in January and will play in February (they will "come back" to play), or 4) they did not play in January and will not play in February (they will "stay out"). These definitions help us explain the great performance of the baseline model. A precision of 89% means that 11% of the total player base churned from January to February. A recall of 95%

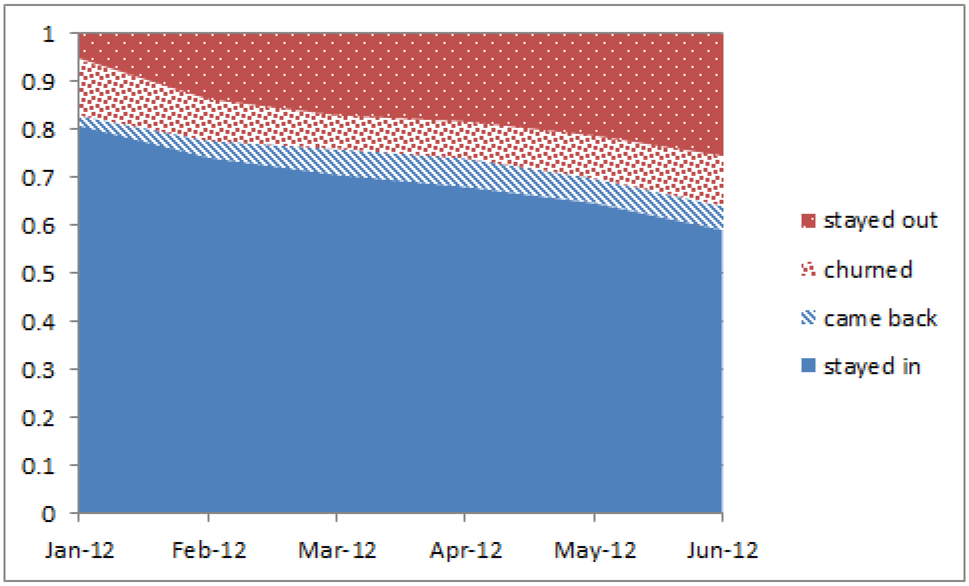


Figure 4.11: Month-to-month transfer behavior of the World of Warcraft player base.

means that 5% of the total player base came back.

For all months, the percentage of players who churn or come back is small compared to the proportion who stay in or stay out, as shown in Figure 4.11. In January for example, the baseline was right for $100 - 11 - 5 = 84\%$ of the total player base. That is how we obtain high precision and recall using a very dumb model. Note, however, that for all months, the percentage of players who churn is roughly twice the percentage of players who come back. Since the average player in our sample has been playing for nearly five years, the ratio of newcomers seems small, and the game is therefore "leaking" players every month. This leak means that more and more players are staying out.

Predicting churn: Predicting presence is not such an interesting problem after all. Instead, we focus on predicting who will churn among the active population. Such a model may be more actionable to game companies. They can focus some of their attention on churners to identify potential flaws in their design, or populations that their game is not targeting well.

The baseline model mentioned earlier does not make any sense anymore. It predicts all active players to stay in the game, thereby achieving a recall of 0. We abandon that model. We

Prediction	Model	Precision	Recall
February Presence	Same as January	.89	.95
	Regression	.91	.93
June Presence	Same as May	.85	.92
	Regression	.90	.87
February Churn	Regression	.28	.90
June Churn	Regression	.19	.89

Table 4.8: Precision and recall of predicting monthly presence and churn.

rebuild the regression models in the same fashion as for presence, except that we only use data from the active player base for a given month, and that the label is whether a player will churn the following month.

As shown in Table 4.7, the January model for churn is the opposite of the January model for presence: US players are more likely to churn in January, and the more characters played and raids in December, the less likely to churn in January. However, age has become significant: older players are less likely to churn independently of all other variables. Similarly, the May model for churn is the opposite of the May model of presence, except the number of PvP occurrences and region have become insignificant.

After manual tuning using the December-January data, we set the cutoff to 0.1. We obtain very modest precisions of 28% and 19% to predict February and June churn respectively, but recalls remain around 90%. This means the models miss out on very few churners, but also flag as churners lots of players who actually stay in. Although their precision is poor, these regression models are simple: they focus on only four features, all obtainable in-game except age. These models provide a good basis for future more complex models.

Summary: Predicting churn is more difficult than predicting presence. Every month, for two players who churn, one is coming back.

4.6 Summary

Several variables reflect the diversity of World of Warcraft players: age, gender, region, play motivations, and play companions. Asian players tend to be more dedicated, play longer per week, and be more tolerant to difficulty than American players. Compared to women, men are more achievement-oriented, and nearly twice more likely to buy gold. On the other hand, players who play with a family member are less likely to buy gold, maybe because they do not need it, or maybe because they worry about the repercussions of their family members knowing about their purchase. Overall, the findings from this chapter lead to the following design recommendations: 1) When targeting a wide spectrum of players, not everyone can be satisfied, and churn is to be expected. 2) Release new content faster. Players come back when new content is released, but drop out at a linear rate of 5% per month afterwards. 3) Provide ways for players who are coming back to the game after a break to catch-up with others. This "re-boarding process" should give players a tour of the features most recently added to the game, maybe provide tasks with a lower level of difficulty, but should not be a hand-holding tutorial. More than 3 out of 4 players reports having taken a break from the game and coming back. 4) Provide an anonymous and confidential way for players to purchase game gold with real money. There is a strong demand for it: despite being against the terms of service, 14% of players report buying virtual gold with real money. The recently-released game token allows players to trade subscription time, paid for with real money, for game gold anonymously through the in-game auction house. More work is needed to assess if the token addresses player demands or if players are still using third-party websites to buy their gold.

Category	**Finding**
Challenge	Achievement-oriented players play longer per week.
Time	Every month, 10% of players churn and 5% come back.
Time	Boss kills peak and fall over 3-4 months.
Time	Dungeon activity is halved 7 months after release.
Segments	Engagement and gold buying differ between region, age, and gender.
Segments	Asians play harder but churn faster.
Segments	Playing more characters decreases the likelihood to churn.
Segments	Casual players take fewer breaks than hardcore players.
Segments	Achievers, men, and older players are more likely to buy gold.
Segments	Players motivated by immersion play longer per week.
Social	Players playing with real-life friends buy less gold and churn less.
Social	Players playing with only in-game friends buy more gold.
Social	Socially-motivated players play longer per week and churn less.
Social	Players in a guild play longer per week and churn less.
Money	25% of players pay their subscription but do not play for 3-6 months.
Money	Players are more likely to buy gold when coming back from a break.
Misc	Playing more characters decreases the likelihood to churn.

Table 4.9: Summary of findings for WoW.

Chapter 5

Forza Motorsport 4

Forza Motorsport 4 (FM4) is a racing game released in 2011 for the Xbox 360. As of 2016, the game has sold 4.6 million copies[1]. FM4 lets players customize the game's difficulty by providing several assists. Assists are game mechanics such as the trajectory line or automatic gear shifting that help a player race. Not all players have the same skill, so players can enable or disable assists as they see fit. We look at how these assists are used, and how they are disabled over time, for more than 200,000 players. This allows us to determine which assists are useful to beginners, unforgiving to the average player, or disabled only by veterans. By looking at the trends in players' racing data, we can model which factors determine when players can safely disable an assist without being bored or frustrated. This chapter relies on a study we conducted in 2013 [42]. We first provide a quick overview of the game with a description of each of the assists, then present the methods and an overview of the data. We explore in depth how players use and disable assists, and propose two models predicting if a player is ready to disable an assist.

[1] See `http://www.vgchartz.com/game/45656/forza-motorsport-4/`

5.1 Gameplay

Forza Motorsport 4 (FM4) is an Xbox 360 racing game developed by Turn 10 Studios. FM4 is the fourth installment in the series of Forza games.

5.1.1 Cars and Modes

FM4 is a complex racing simulation where each car has a particular weight, steering rate, drive type (e.g., all-wheel drive, AWD, versus rear-wheel drive, RWD), and so on. To allow players to compare cars at a glance, the game summarizes the performance of a car in a score called the Performance Index (PI) ranging from 0 for small city cars to 1 for purpose-built racing cars. Players can acquire cars by spending in-game credits or as a level-up reward. These credits and experience points are awarded when a player completes a race. Races can take place in several modes:

- In career mode, players progress through ten divisions, each containing a dozen races. Higher division races involve cars with higher PI controlled by more difficult AI.
- In quick race, players can select any track and car they want. Quick races are offline and involve up to two players sharing the same screen.
- In online mode, players can participate in multi-player activities ranging from the usual lap races pitting up to 16 players to drag races or even car soccer.
- In rivals, a single player races against the time that another (remote) player previously set for the race.

5.1.2 Assists and Bundles

Assists are game mechanics that help the player drive. The FM4 telemetry logs six assists, each of which can take two or three levels. A seventh assist, steering, makes turning easier, but was not recorded. Turn 10 Studios designed most of the assists to make players more consistent and competitive[2].

Stability control prevents the car from spinning when cornering too fast. **Traction control** prevents it from spinning when accelerating. When the game detects that the car starts spinning, these two assists will slow the car down. Each of these two assists can be turned on or off.

The **braking** assist supports players when they brake or should brake. It can take three levels: in assisted with Anti-lock Braking System (ABS), it automatically slows the car down when approaching a turn. In non-assisted with ABS, it prevents the car from drifting when braking in a straight line, but may increase the braking distance. In non-assisted without ABS, players have complete control of the brakes.

The **shifting** assist helps players in passing gears. It can take three levels. Gears can be shifted automatically, much like in real-life automatic cars. Gears can also be shifted manually, in which case players have to press a button to shift gear up, and another button to shift gear down. Manual shifting gives players control over the gear ratios, which, if used properly, can result in faster acceleration. Gears can also be switched manually with clutch, in which case players must press the clutch trigger to switch gears.

The **line** assist overlays the optimal trajectory to follow on the track. (The optimal trajectory does not take into account the presence of other cars.) The assist can also take three levels. In full, the line shines green when players should accelerate and red when they should brake.

[2] In a press conference, one of the game's designers mentioned: "Assists make you slower, but they make you more consistent". See `https://www.youtube.com/watch?v=xMUGP4SKAyg`

<table>
<tr><th></th><th>Easy</th><th>Medium</th><th>Hard</th><th>Advanced</th><th>Expert</th></tr>
<tr><td>Stability</td><td colspan="2">ON</td><td colspan="3">OFF</td></tr>
<tr><td>Traction</td><td colspan="3">ON</td><td colspan="2">OFF</td></tr>
<tr><td>Braking</td><td colspan="2">Assisted with ABS</td><td>ABS</td><td colspan="2">OFF</td></tr>
<tr><td>Gear Shifting</td><td colspan="3">Automatic w/o clutch</td><td>Manual w/o clutch</td><td>Manual w/ clutch</td></tr>
<tr><td>Line</td><td colspan="2">Full</td><td colspan="2">Brake</td><td>OFF</td></tr>
<tr><td>Damage</td><td>Cosmetic</td><td colspan="2">Limited</td><td colspan="2">Simulation</td></tr>
</table>

Table 5.1: Configuration of assists in built-in bundles.

In brake, only the red portions are displayed. The line can also be completely turned off.

The **damage** assist determines how much the performance of the car can change during the race. In cosmetic, collisions only leave visible traces on the car. In limited, collisions can reduce the car's performance. In simulation, collisions reduce the performance of the car a lot more, tires wear off, and the player has to think about refueling.

We call bundle a configuration of assists. The game ships with five built-in bundles: easy (selected by default), medium, hard, advanced, and expert. Table 5.1 describes the value of assists in the built-in bundles. A custom bundle is a configuration of assists that is not built-in but player-defined. Players can change the bundle or the value of a particular assist in a menu before any race. In career and online modes, each disabled assist increases the credit rewards by 10%, giving players an incentive for disabling an assist.

Figure 5.1 and a short scenario provide a picture of the skill necessary to drive using the easy vs expert bundles. When approaching a turn in easy, the player stops accelerating and steers so as to follow the full line. The player needs coordinating only two controls together: the right trigger to accelerate and the left stick to steer. In expert, the player guesses the optimal trajectory, shifts gears down in sync with the clutch, and taps the brakes lightly so as not to skid. This involves five simultaneous controls and much more anticipation of the game overall.

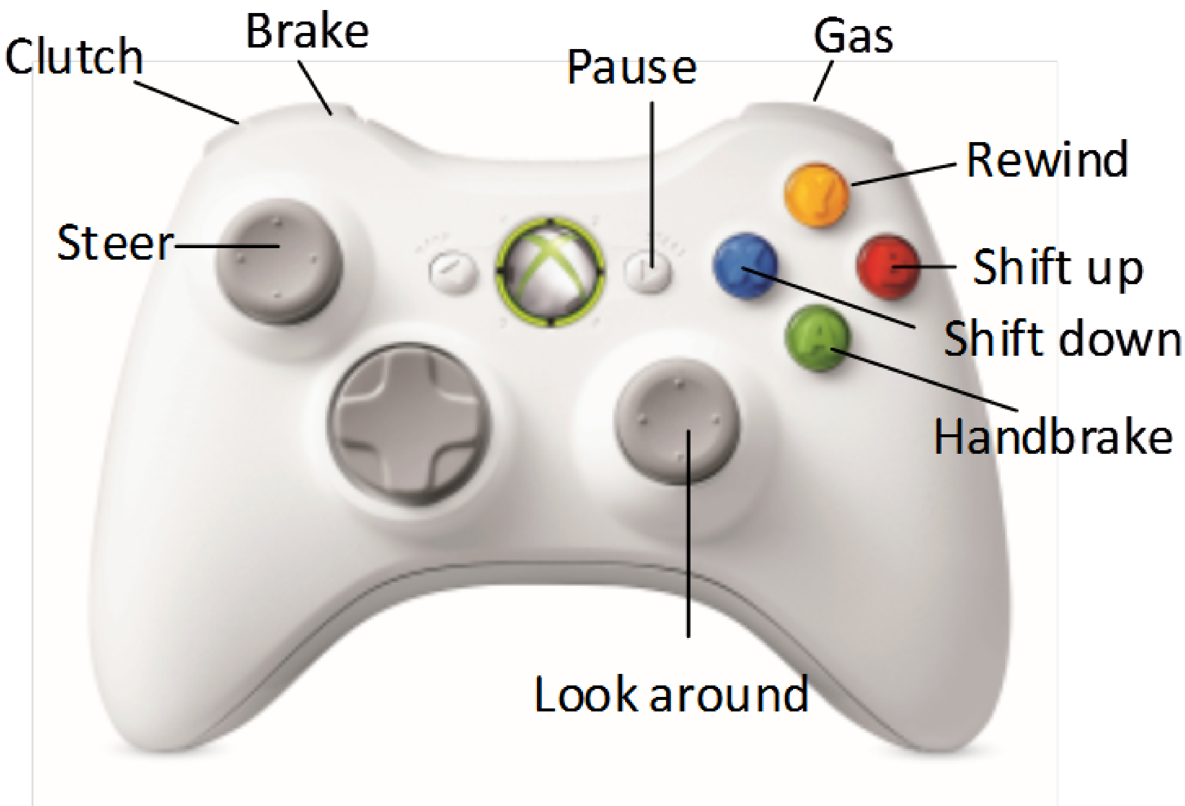


Figure 5.1: Mapping of the controls on an Xbox 360 controller.

5.2 Methods

5.2.1 Race and Achievement Datasets

This chapter combines two datasets. The main dataset consists of the race entries from a random sample of 5% of the whole player base. The 220,000 players generated 24.5 million race entries from the FM4 launch in October 2011 until July 2013. A race entry is created when the player leaves a race, whether completed or abandoned. Each race entry contains the value of the aforementioned assists, and other metrics such as the player's rank at the finish line and the car's PI. The game telemetry logs each of the 3-value assists into two binary assists (on or off). For example, the 3-value braking assist translates to two binary assists called ABS and autobrake, which are partially dependent: the players who use autobrake also use ABS, but those using ABS do not necessarily use autobrake.

Therefore the six assists described in the previous section are represented with ten binary assists in the remaining of this chapter: stability, traction, autobrake, ABS, autoshift, clutch, full line, brake line, cosmetic damage, and limited damage. This abstraction reduces the complexity of the analysis, as now all assists are binary. The telemetry tracks neither the input device (Xbox 360 controller vs steering wheel) nor the view during the race (cockpit vs wheel level vs above car).

To complement this racing data and get a sense of player progression through the game as a whole, we also look at the Xbox Live achievements that players unlock over time. Players unlock a particular achievement when they reach certain levels, complete certain missions, or perform a certain action in the game. These achievements are tracked on the Xbox platform, which is independent from the game's telemetry. Players can be cross-referenced between the race and achievement datasets because they have the same identifier in both.

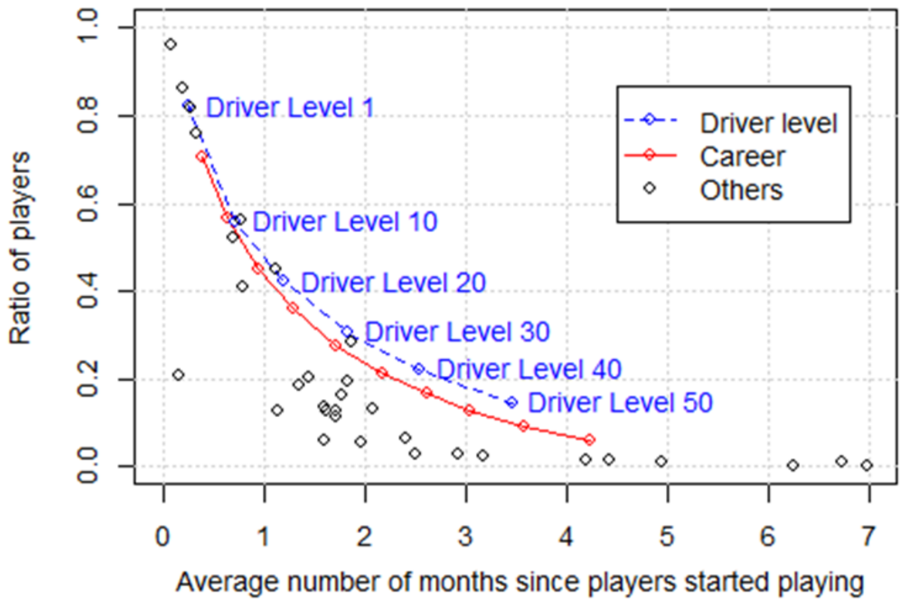


Figure 5.2: FM4 achievements tracking player progression over time.

5.2.2 General Trends and Distributions

Players unlock an achievement for reaching each of driver levels 1, 10, 20, 30, 40, and 50 (the maximum level). They also unlock an achievement for completing each of the 10 divisions of career mode. Figure 5.2 plots the number of players who unlocked each achievement, and the average time it took them. Half of the players reach driver level 20, where moderate-performance cars are offered as level-up rewards. Around 15% reach level 50, where the fastest cars are awarded. More than 20% import data from FM3, the previous game in the franchise, and they do so, on average, a few days after they start playing.

Around 48% of all races take place in career mode, 30% online, 16% in quick race, and 6% in rivals. Given this distribution, we place more focus on the career and online modes than on the other modes. Not all players race equally. The median player has raced 29 times, and the top 5% players more than 434 times. These top 5% players account for half of all races. Players race three times per day on average (median 1.2).

5.3 Patterns of Assists

5.3.1 First-race Bundle

The game ships with the easy bundle selected by default. Thus it is not surprising that 85% of players race for the first time using the easy bundle. This also means that 15% of players race for the first time using a bundle different than easy (medium, hard, advanced, expert, as well as custom). Note that the first race, albeit full-fledged with AI opponents and moderate-performance cars, is a one-lap hands-on tutorial in quick race mode, aimed at explaining the controls to the player. Thus it is understandable that all the assists are enabled by default.

Among the players who start with the easy bundle, most eventually change the assists, but a small proportion of players never touch them. The players who eventually change the assists race on average 147 times (median 57) over 163 days (median 81). On average, they change an assist for the first time after 9 races (median 3). Most of the players who never touch the assists race less than 10 times. We call this segment of players the "samplers", since they only experienced a sample of the game.

5.3.2 Most Frequently Used Bundles

Before looking at individual assists, we look at the bundles commonly used by players. With ten binary assists, there are around a thousand possible bundles. However, only the eight reported in Table 5.2 are used at least once by 10% of players of any mode. In terms of players, the Easy bundle is by far the most used bundle in career mode, and the third most used in online mode. The large segment of "samplers" mentioned before helps explain this.

Among the other built-in bundles, only Medium is somewhat popular in career mode, but

Bundle	Career Mode % players	Career Mode % races	Online Mode % players	Online Mode % races
Easy	57.9	18.5	29.9	3.0
Stability, traction, clutch, cosmetic damage	28.2	1.8	49.5	5.0
Stability, traction, clutch	21.0	2.1	12.5	0.5
Stability, traction, clutch, limited damage	13.7	0.8	4.0	1.1
Medium	10.5	3.2	0.8	0.0
All but autobrake	7.6	4.0	12.5	2.8
All but autobrake and full line	4.2	2.7	10.0	2.6
Clutch and cosmetic damage	1.2	0.5	31.8	20.4

Table 5.2: Frequency of bundles across modes.

most players seem to customize the difficulty at the assist level rather than at the bundle level. Moreover, the popular bundles in online mode keep damage at the cosmetic level, probably because players know they will collide with each other often and do not want their performance to suffer from it. And finally, clutch is the only assist that is enabled in all of the bundles listed in Table 5.2.

The player-defined bundle with stability, traction, clutch, and cosmetic damage assists enabled is used by almost a third of career players (28.5%) and half of online players (49.5%). Yet no built-in bundle resembles it.

5.3.3 Assist Progression

How many races does the average player need to feel confident about disabling an assist? Once again, we distinguish between the career and online modes. We plot the evolution of assist usage over the number of races in Figure 5.3. The graph is cut at 1,000 races because less than 1% of players ever raced more than that in either mode.

In their first career mode race, around 70% of players enable the full line, autobrake, and cosmetic damage, 95% enable the clutch, and 80-85% the other assists. Over time, autobrake and full line are disabled twice faster than other assists, suggesting they are the easiest to

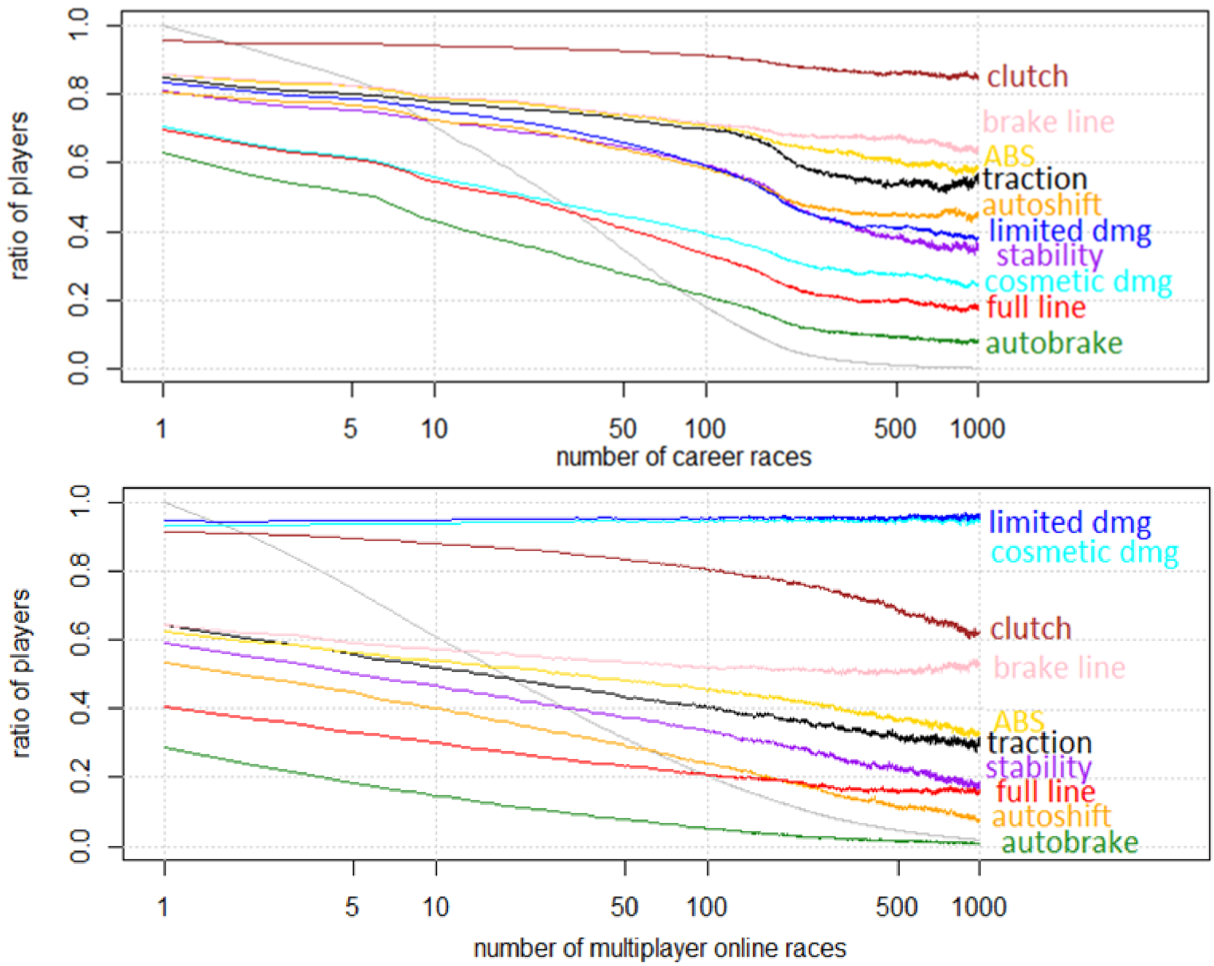


Figure 5.3: Ratio of players enabling an assist over their races in career and online modes.

disable, or the least enjoyable. Clutch stays flat, suggesting it is the hardest to disable. The slight dip around 200 career races marks the final race of career mode; beyond that point, players replay career races they already completed.

Assists follow the same trends in online mode with several exceptions. First, players start their first online race with more assists disabled than in their first career race. For example, autobrake is used in only 30% of the first online race versus 60% of the first career race. Second, clutch is disabled faster and by more players in online mode. Players who race online may have practiced in career mode beforehand, and seem to be more skilled than career mode players. Yet after a thousand online races, 20% of online players still use the full line, and 50% the brake line. Figure 5.3 also confirms that players do not want to be penalized by colliding with others: both cosmetic and limited damage stay flat at 95

5.3.4 Assist Transitions

Once a player disables an assist, several scenarios can happen. (1) The player can keep it disabled forever, in which case the increased difficulty matches the player's skill, and we call the disabling a "success". (2) Otherwise, the player eventually re-enables the assist. If the player never disables the assist again, racing without the assist was probably so hard or displeasing that the player never does it again. In this case, we call the disabling of the assist a "failure". (3) If the player re-enables and re-disables the assist once or more, we call the disabling a "yoyo". Figure 5.4 summarizes these terms and highlights the race before and the race after the first time an assist is disabled.

For each assist, Figure 6 provides the percentages of successes, failures, yoyos, and players who never disable. Autobrake is easiest to disable with 35% of success, while clutch the most difficult with less than 1% of success. Overall, the players who "fail" at disabling an assist re-enable it after very few races. For example, 38% of the stability failures re-enable

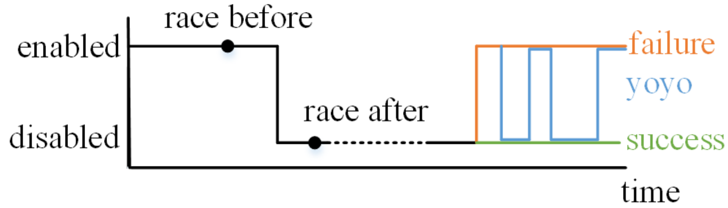


Figure 5.4: Scenarios of success, failure, and yoyo after disabling an assist for the first time.

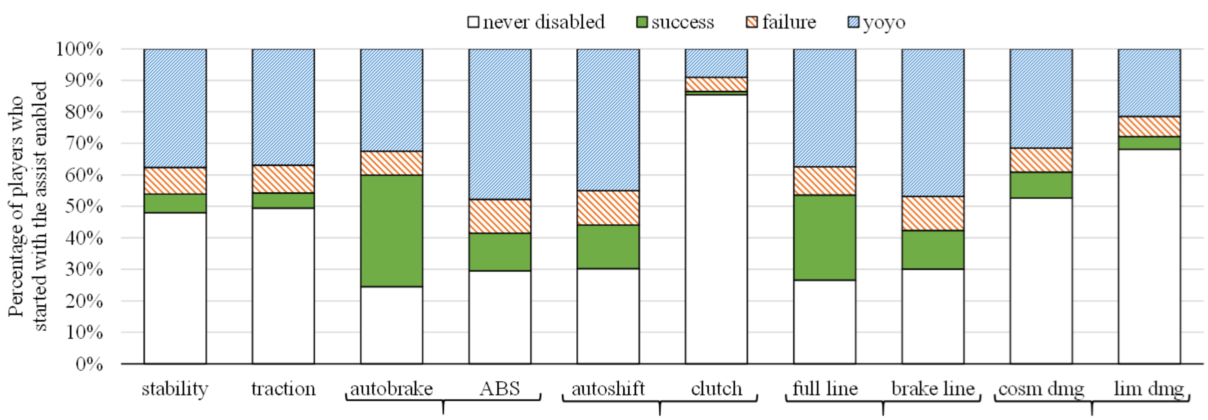


Figure 5.5: Rate of success, failure, and yoyo after disabling an assist for the first time.

it the very next race. This percentage rises to 56% for clutch, 69% for autobrake, and 80% for brake line. Disabling an assist seems to be an immediate hit-or-miss.

5.4 Modeling Assist Transitions

To model the success or failure of players in disabling a particular assist, we restrict our sample to the players in the success and failure categories for that assist. We investigate two possible scenarios:

- Scenario 1: A player finished a race. Can we predict if the player is ready to successfully disable an assist?
- Scenario 2: A player finished a race, disabled an assist, and then finished a second race. Can we characterize what makes a successful disabling of the assist?

We use one dataset for each of the ten binary assist. Each dataset holds racing data from the races before and after the assist was disabled. The features we select aim at measuring in-game racing skill. They are drawn from our experience playing the game and from the literature on game skill.

For each assist and scenario, we run a multivariate logistic regression using the player's race statistics (see below) as independent variables, and the success of disabling an assist as the dependent variable (also called outcome or label). For each independent variable, the regression computes the log of the odds ratio and a significance level. For example, a negative log odds ratio for the number of races means the odds of successfully disabling an assist are higher for players with fewer races, holding all other independent variables constant. In other words, the sooner players disable an assist, the more likely they are to succeed.

Once a model is built, it contains one coefficient (the log odds ratio) per variable. We can then input the data of a player who was not used to build the model to predict whether she will disable an assist successfully or not. The models that are built for Scenario 1 can be used as input for recommender systems suggesting which assists to disable; models from Scenario 2 can shed light on the factors contributing to successfully disabling an assist.

5.4.1 Scenario 1: Predicting the Success of Disabling

To assess if a player were to successfully disable an assist after the completion of a race, we measure player skill using metrics generalizable to any racing game. We use metrics from only the race before the assist was disabled.

We saw that the more races played, the more likely to disable an assist; the **number of races** is our first metric. The **position at the finish line** is common to all racing games, and finishing last is a strong signal of under-performance; this is our second metric. When we

Assist	Num races	Races/day	Career	RWD	Car PI	Position
Stability	−	−		+	−	−
Traction	−	−		+	−	
Autobrake	−	−	−	+	−	−
ABS	−	−	−	+	−	−
Autoshift	−	−	−	+	−	−
Clutch	−	−		+	−	
Full line	−	−	−	+	−	−
Brake line	−	−	−	+	−	−
Cosmetic Damage	−	−		+	−	−
Limited Damage	−	−		+	−	−

Table 5.3: Sign of the log-odds ratios of a logistic regression in scenario 1.

played the game, we found **career mode** more appropriate for training than online mode. It is only in career mode that the player can lower the AI's difficulty or rewind the race when missing a turn. A previous work noted large differences between the number of games and the number of days one has been playing for [69]. So we add the **number of races played per day**. We also add two car metrics: the **car performance index** (PI) and whether the car's drive type is **Rear-Wheel Drive** (RWD). Cars with higher PI are generally more sensitive and demanding in terms of reflexes and skill. An FM4 developer admits using the traction assist when driving a RWD car[3]. Cars with high PI numbers are generally RWD.

Table 5.3 describes the influence of each factor on the model. We expect the players who race only a few times to be less skilled than players who race many times, and thus less likely to keep an assist disabled they disable too early. However we find the opposite in the data: the odds of successfully disabling an assist are higher for the players who participate in fewer races. In other words, the players who disable an assist early are more likely to keep it disabled. We saw earlier that roughly 20% of the player base played FM3 before. These players already know which assists they need. On the other hand, players who race for a long time with an assist enabled may become used to it.

[3]FM4's creative director wrote: "I still switch back to the Normal Steering setting when I move into some of the more nutter cars (i.e. 599 GTO with no TCS [traction], STM [stability] or ABS [braking]) or when I'm just looking to drift.". See `http://forzamotorsport.net/en-us/news/underthehood3`

5.4.2 Scenario 2: Characterizing a Successful Disabling

We saw earlier that most of the players who re-enable an assist do so the race immediately after they disabled it. In order to assess the factors in the race that influence the re-enabling of an assist, we start with factors from model 1: career mode, position at the finish line, number of races so far, and number of races per day so far. Previous literature shows that skill degrades quickly over time [69], so we also add the number of **days elapsed between the race before and the race after** disabling the assist. RWD was also included in the early stages of Model 2, but became insignificant when we added to the model whether players **switch from RWD to FAWD** (front or all-wheel drive). Switching from FAWD to RWD is also insignificant for all assists, and thus not reported. The car performance index is carried over from model 1, complemented with the **difference in car PI** between the race before and the race after. A positive PI difference indicates that the player races a car that requires more skill than the car used in the previous race. And finally, we add to the model whether the player **finishes the race**, as opposed to abandoning it.

Table 5.4 lists the sign of the log-odds ratios for all significant variables retained by the model. Controlling for all other factors in the model, the odds of maintaining the assist disabled are higher for players who finish the race. This is not surprising: a player who has no fun without the assist can (and probably should) abandon the race and re-enable it.

In Scenario 1 the odds of successfully disabling an assist are lower for the players who race in career mode before disabling the assist. Here in Scenario 2, the odds of success are higher for players who race in career mode after disabling the assist. In other words, the players who are the most likely to successfully disable an assist are those who race in online mode, disable the assist, and then race in career mode. Career mode is indeed better for practice.

Assist	Num races	Races /day	Days diff	Career Mode	RWD to FAWD	Car PI	PI diff	Fini -shed	Posi -tion
Stability	−	−	+	+	+			+	−
Traction	−	−	+	+	+				−
Autobrake	−	−			+			+	−
ABS	−	−	+		+			+	−
Autoshift	−	−	+		+		−	+	−
Clutch	−	−	+	+	+				−
Full line	−	−	+		+		−	+	−
Brake line	−	−	+	+	+			+	−
CosmDmg	−	−	+	+	+	−	+	+	−
LimDmg	−	−	+	+	+	−		+	−

Table 5.4: Sign of the log-odds ratios of a logistic regression in scenario 2.

	Number of players who	Disable Success	Scenario 1: race before		Scenario 2: before & after	
Assist	disabled	(%)	Precision	Recall	Precision	Recall
Stability	23,915	46	61	73	65	64
Traction	21,066	40	61	58	63	53
Autobrake	73,244	82	84	97	88	95
ABS	42,535	50	66	73	81	65
Autoshift	44,314	54	70	80	83	74
Clutch	11,653	22	–	–	63	29
Full line	62,955	74	78	95	88	89
Brake line	46,094	51	68	76	83	68
Cosm. Damage	31,212	51	61	81	64	80
Ltd. Damage	20,872	38	60	50	60	62

Table 5.5: Precision and recall for each assist in both scenarios.

5.4.3 Evaluating the Models

In order to assess the efficacy of our approach at predicting whether a player will keep an assist disabled, we train models for each scenario and each assist on two thirds of the data (picked randomly), and test on the last third. For each test instance, the regression outputs a likelihood between 0 and 1. The cutoff point is set at 0.5. We repeat this process 50 times, and report the average precision and recall for each assist and each model in Table 5.5.

In our case, precision matters more than recall. A low precision means that the model overestimates the skill of a lot of players. If the system tells the player that she is ready, when she actually is not, disabling the assist may lead to an unpleasant game experience. On the other hand, a low recall means that the model underestimates the skill of a lot of players, which is acceptable: if the game is too easy, players can disable the assists themselves, as is presently the case.

Overall the precision is between 60% and 90% for Scenario 1, suggesting that in at least 6 out of 10 races we can correctly predict and recommend when a player should switch. For Scenario 2, predicting whether a user should keep the assist disabled after they have switched, we have higher precision values compared to Scenario 1 because more information is available to predict a successful disabling.

5.5 Summary

By analyzing 24 million races from 200,000 players, we identified patterns of assist usage, and highlighted noticeable differences between player segments, notably the veterans from Forza Motorsport 3 and the "samplers" who race very little. The simple models we built achieve great precision and recall for the easiest assists such as autobrake and the full line. Their performance is more modest for the more difficult assists such as traction or clutch. The findings from this chapter allows us to draw the following recommendations for design: 1) Provide players mechanics, such as assists, that let them adjust the difficulty themselves. On a tortuous track, for example, savvy players may re-enable the brake-line assist to help them anticipate the track. This is shown in the yo-yo pattern of certain assists. 2) In competitive multi-player activities, set defaults to prevent toxic behavior. In this case, it means setting the damage assist to cosmetic by default in online mode. This would prevent players driving solid jeeps to ram into fragile sports cars and break them, leaving their driver unable to complete a race. 3) Nudge players to make the game more difficult earlier than later. Player behavior seems to follow an early-or-never pattern, where they get used to using a particular

Category	**Finding**
Challenge	A third to half of players yoyo their assist transitions.
Challenge	Disabling are more permanent if followed by easier races.
Time	Easier assists are disabled faster.
Time	Players tend to disable an assist early or never.
Time	Disabling are more permanent if players have not raced in a while.
Time	50% of players churn within a month.
Segments	20% of Forza 4 players are Forza 3 veterans.
Segments	13% of players change an assist before their first race.
Segments	Top 5% of players contribute half of all races.
Social	Damage assist is cosmetic when racing other players online.
Infrastructure	Assists like clutch may be disabled only with certain controllers.

Table 5.6: Summary of findings for FM4.

set of assists after a while.

Chapter 6

Jelly Splash

Jelly Splash (JS) is a free-to-play (F2P) mobile game released in August 2013. The game quickly reached number 1 on Apple's AppStore in the US, Germany, and several other countries [114]. It was installed 25 million times between August 2013 and April 2014, and had 8 million monthly active players in December 2013 [118]. JS is still running as of 2016.

F2P mobile games aim at attracting as many players as possible. They make money through a small percentage of their players spending real money in virtual items or currency. Table 6.1 lists mechanics that F2P games to generate revenues. JS uses three of these mechanics: occasional spikes in level difficulty, chapter gates every 20 levels, and a limited number of lives regenerating slowly over time.

In this chapter, based on previous published work [38], we ask the following questions: Which gating mechanics drive revenues? Which are most likely to make players churn? Are there trade-offs? We first describe the methods used to analyze the data, and then investigate each gating mechanic one by one in more details.

Mechanic	Example in games	Focus
Time-based	Speed-ups in Clash of Clans	Time
Energy points	Stamina in Puzzle and Dragons	Time
Luck-based	Crafting gear in Angry Birds: Epic	Time
Collecting/Gacha	Egg machine in Puzzle and Dragons	Time
Advertisement	End of a round in Words With Friends	Time
Expansions/DLC	Naxxramas Wings in Hearthstone	Space
Chapter gate	Stars in Plants vs Zombies 2	Space and Challenge
Continues	Arcade and endless runner games	Challenge
Power-ups	Rare gems in Bejeweled Blitz	Challenge
Customization	Skins in League of Legends	None

Table 6.1: Common F2P monetization mechanics and their focus.

6.1 Gameplay

Jelly Splash is a tile-matching game similar to Bejeweled or Candy Crush. As shown in figure 6.1, the game consists of connecting 3 or more same-colored tiles, called jellies, to blow them up and make new ones drop from the top of the board. The core loop is designed to be 70% luck and 30% skill [118]. The game shipped with 140 levels, and 20 new levels have been added roughly every other month in 2014. Each level is a puzzle where the player has to complete a certain goal in a limited number of moves to unlock the next level. Goals include reaching a certain score, making 2 to 6 diamonds fall to the bottom of the board, or clearing a certain number of gray cells.

Jelly Splash has three monetization mechanics. First, when players run out of moves during a game, they can spend 70 in-game coins (approximately $1) to receive 3 **extra moves**, or they lose a life. These **extra move requests** (XMR) are a continue mechanic tailored to the luck-based core loop of the game [118]. Purchasing extra moves is a direct counter to the game's difficulty.

Instant **life refills** are a second monetization mechanic. Players have at most 5 lives, and one life regenerates every half hour. To refill all their lives immediately, players can spend

Figure 6.1: Screenshots of Jelly Splash.

100 coins. They can also login through Facebook, the only social platform supported by the game, and ask their friends for lives. This mechanic is both time-based and social. However, refilling lives is a consequence of losing lives, which means this mechanic is also related to the game's difficulty.

Players can also spend 70 coins to unlock each of the **gates** stationed every 20 levels starting at level 40, or ask three of their friends for gate keys. This is purely a spatial restriction of the content available. It is not related to difficulty, but it is interesting to compare to the other challenge-based mechanics.

6.2 Methods

6.2.1 Dataset

We select 5% of the Jelly Splash iOS players based on their device id, a number assigned by the hardware manufacturer and free of any sampling bias. This sub-sampling makes data analysis tasks more tractable, while retaining a very strong statistical validity [51]. An update in November 2013 significantly altered the chapter gates, so we restrict our data to the 10 weeks from launch on August 22 until October 31, 2013. The in-game telemetry tracks when users install the game, login, finish a level (whether won, lost, or abandoned), and purchase virtual currency. The telemetry also tracks when players log into the game with their Facebook account, and how many of their friends are playing. The telemetry does not track any personally identifiable data such as a player's email address, IP address, or Facebook account.

Real-life telemetry data can be messy. We have to clean it up. First, we discard data that is 1) erroneous, such as some players purchasing a billion coins, 2) inconsistent, such as some players finishing a level before they installed the game, and 3) lost, such as some players having won level 6, but having no data for level 5. Second, we searched Jelly-Splash-related forums for known exploits, cheats, and technical issues. An early exploit allowed some players to obtain a potentially infinite number of virtual coins by re-installing the game. Therefore we discard data from players who installed the game more than once. Some players manipulated their device clock to refill their lives faster. Thus we ignore data from players who have no Facebook friends playing the game and who managed to refill their lives without spending coins. After cleanup, we have data from 273,819 players. These players played 25 million games in 7 million sessions, and made 37,170 purchases.

This study is exploratory and descriptive. We want to find trends happening in the game, not predict when players are going to churn, or how much they will purchase. This is what is referred to as descriptive data mining [53]. To this end, we use basic statistical tools and graphs. Unless reported otherwise, all results are significant at the $p < .001$ level. For correlations, we often cannot report Pearson's linear correlation coefficient, because a lot of our exploratory plots show obviously non-linear relationships between variables. Thus we report Spearman's rank correlation coefficient ρ instead. We follow [32] when estimating effect sizes: when ρ is between .1 and .3 we consider the effect weak, between .3 and .5 medium, and above .5 strong.

6.2.2 Data Overview

The average game session lasts 7 minutes (sd 14, median 4, 99% below 31). Users play 3 sessions per day on average (sd 3). The time between two sessions is very short: 20% of inter-session times are below a minute, and 99.2% below a week. Previous work on Diamond Dash, another mobile game, found that 98% of players stay away from the game for less than 14 days, and thus defined churn as 14 days of inactivity [108]. We map this definition to our study, and consider that people who do not play for a week have taken a break from the game (vacation, tired of the game, and so on). With this definition and its limitations in mind, players leave the game on average 13 days after install (sd 16, median 8).

Figure 6.2 plots the ratio of users still playing the game against the number of days after they installed the game. The average level reached by active players is indicated for days 0, 1, 2, and multiples of 10 since install. For example, 21% of the player base is still playing the game 10 days after having installed it, and these players have reached level 35 on average. Around 5% of players install the game but never play. Half of players churn within a day after having installed the game, reaching only level 15.

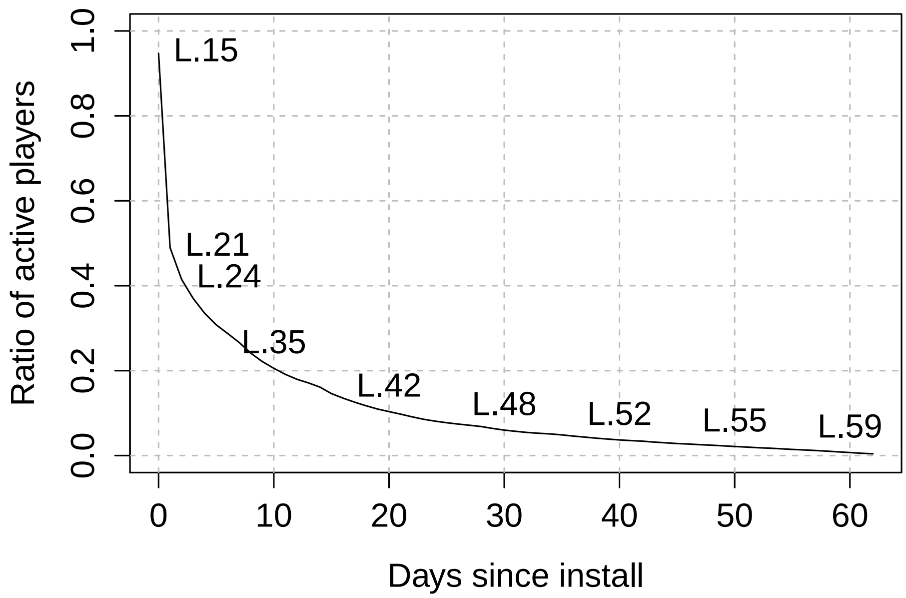


Figure 6.2: Ratio of active players X days after install.

6.3 Difficulty Spikes

Traditionally, game difficulty ramps up progressively as players learn new skills, and higher levels are generally harder than lower levels [3]. The designers of Jelly Splash took a different approach. In Jelly Splash, difficulty is not designed to increase monotonically, but rather to spike at certain levels [118].

6.3.1 Level Difficulty

The Jelly Splash designers [118] measure the difficulty of a level by the average number of tries that players take to win it:

$$Difficulty = avg(number\ of\ tries\ until\ won)$$

We look at data aggregated by level. Each of the 140 data points is indexed by level number, and contains aggregates such as the number of players who ever reached the level, the number of players who leave at that level, the average number of tries for players to win the level, the total number of coins bought at that level, and the total number of extra move requests (XMR) made at that level.

With this dataset, the average level takes 8 tries to pass, or 4 hours worth of lives. Level 97, the most difficult level in the game, takes the average player 71 tries to pass. The blue bars in Figure 6.3 show the spikes in difficulty from levels 10 to 30. The spikes in churn coincide with the spikes in difficulty. Although the figure does not show levels below 10 or above 30 to remain readable, the spikes continue to visually coincide until level 140.

Despite the spikes, the overall difficulty increases as the levels go by ($\rho(138) = .41$). Therefore, we have to control the level number when assessing how player behavior relates to difficulty. To do so, we use partial correlations instead of regular correlations. Partial correlations measure the association between two variables while holding constant a third variable. In our case, controlling for level number, we observe very strong partial Spearman correlations between difficulty and churn ($\rho(138) = .71$), between difficulty and the number of coins purchased per user ($\rho(138) = .81$), and between difficulty and the number of XMR per user ($\rho(138) = .76$). This is intuitive: players are more likely to purchase coins and spend them in extra moves in more difficult levels, where they need them.

6.3.2 Level Hopelessness

To differentiate "good" difficult levels from "bad" difficult levels, the Jelly Splash designers [118] measure a level's hopelessness as:

$$Hopelessness = avg(\frac{number\ of\ tries\ until\ won}{number\ of\ tries\ nearly\ won})$$

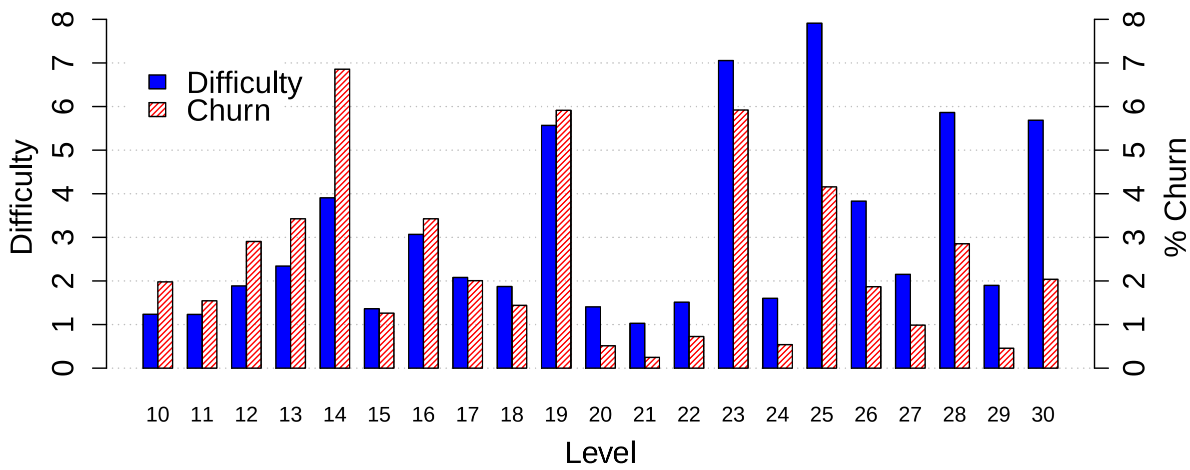


Figure 6.3: Difficulty and churn for levels 10-30 in Jelly Splash.

When players lose a difficult level with high hopelessness, they have usually completed only few of the level's goals. For example, they only bring one diamond down, out of the six required. When players lose a level with low hopelessness, they have usually completed most of the level's goals, such as five out of six diamonds. Levels with low hopelessness are "good" because they give players feedback that the goal is within their reach [118].

However, the designers' definition of "nearly won" is imprecise. Does it mean that players, without spending coins in XMRs, complete 50% of the level's goals? Or 90% of the level's goals? Determining this threshold is probably more art than science. For this study, we arbitrarily consider a try to be "nearly won" when the player completes at least 75% of the level's goals without using extra moves.

As a level metric, hopelessness has several limitations. First, hopelessness for a particular level is undefined for players who never "nearly win" that level. Second, the 75% threshold is arbitrary and difficult to justify. Third, hopelessness only makes sense for difficult levels. Players pass an easy level in so few tries that they probably do not have time to notice its hopelessness. A fourth limitation is that the hopelessness of some levels can not be calculated. Levels 1 to 4 are tutorial levels with infinite moves, so they can never be lost. Level 132 has only one recipe to make in 15 moves, which means it can be 100% won or

100% lost, but never “nearly won”, whatever threshold we pick.

With these limitations in mind, the average level hopelessness is 4, meaning that the average level makes players nearly win once every 4 tries. According to the Jelly Splash designers, a “good” difficult level has a hopelessness below 10 [118]. Hopelessness, much like difficulty, spikes occasionally, and increases with level ($\rho(132) = .47$). Hopelessness also increases with difficulty ($\rho(132) = .64$). When we control level number and difficulty on the relationship between hopelessness and churn, we find a very modest partial Spearman correlation ($\rho(132) = .13$, $p < .05$). Similarly, controlling level number and difficulty, hopelessness is strongly correlated with XMR per user ($\rho(132) = .54$) and coins bought per user ($\rho(132) = .55$). In other words, for two equally difficult levels, the most hopeless one has more purchases, more spending, and also more churn.

6.3.3 XMRs and Board Rerolls

Two game mechanics can help players mitigate luck in difficult levels: XMR and board rerolls. We just saw that XMR are more likely to happen in difficult and hopeless levels. To see how helpful XMR are to players, we look at the won or lost but not canceled games after the tutorial ends, at level 5. Our dataset comprises 23 million games. We find that games in which players do not request any extra moves are won 34% of the time. For games with one XMR, the win rate is 38%, two XMRs 69%, three XMRs 83%, and 10 XMRs 87%. This progression suggests two take-aways. First, players hardly increase their winning chance with only one XMR. Difficult levels may be designed to be reliably won with 2 or 3 XMRs, but not with one. Second, players seem to severely misjudge their winning chance when using extra moves. Why spend 70 coins (worth $1) in an XMR if it increases the chance of winning so little?

Players can cancel a game without losing a life if they have made no move. We call this a board reroll. In a game designed to be 70% about chance [118], this mechanic should be very useful to players. Aggregating by players the dataset of 25 millions games, we find that 18% of players ever reroll a board. And among the players who reroll, 60% reroll only once, which may be by mistake. In short, players do not seem to be aware of a mechanic that could increase their odds of winning a level. This may be a user interface issue, since there is not really a button for it.[1]

Summary: Hopelessness and difficulty are correlated. Both spike, increase with levels, and are positively correlated with churn and purchases. XMR and board rerolls mitigate luck, but players use them ineffectively, if at all.

6.4 Chapter Gates

Chapter gates happen every 20 levels, starting at level 40. Passing a gate requires 70 coins or asking 3 Facebook friends for keys. Thus we have to look at retention and purchases, and also Facebook logins. To drive our analysis, we try to confirm or reject the following hypotheses: 1) Gates have a negative influence on retention. 2) Gates are responsible for a large portion of all revenues. 3) Gates are effective at making players login with Facebook.

We focus in particular on the gates at level 40 and 60 for 2 reasons. First level 40 is the first gate that players encounter, and the only one that can be unlocked with the 70 coins players start with. Level 60 is the first gate where all players have to use friends or purchase coins to pass. Second, gates at level 80 and above have much fewer players than the first two. With smaller N, our conclusions are weaker.

[1]The ability to reroll may sound more like an exploit than a mechanic. Yet the developers never fixed it, so we consider it a mechanic.

Outcome	β Level number	β Lvl difficulty	β Lvl is gate	Adj. R^2	F
Churn (in %)	-.04	.10	11	.50	47
Num. purchases	.002	.009	.093*	.59	68
Qty purchased	1.0	3.5	45*	.52	51

Table 6.2: Linear regressions explaining churn and purchases in JS.

6.4.1 Retention

The average churn rate for non-gate levels is 1.4% (median 5%, max 9%). Gate levels have much higher churn: around 31% of players who reach level 40 churn at the first gate, and 28% who reach level 60 churn at the second gate. At first glance, gates are terrible for retention.

We saw in the previous section that level number and difficulty are correlated with churn. So if we want to measure the relationship between churn and gates, we need to control for level number and difficulty. We perform a linear regression using the level data with 140 rows. The dependent variable (also called outcome) is churn, measured as a percentage of the player base that ever reached the level. The independent variables are the level number, the level difficulty, and whether the level is a gate. The results of this regression are shown in the first row of Table 6.2. Note that all p-values are below .0001 except for * where p is .092. The results can be interpreted as follows: controlling for level number and difficulty, churn in gate levels is 11 points higher than in non-gate levels.

Being stuck at a gate does not prevent players from replaying previous levels. One could argue that players can replay the levels below 40 and stay forever. The data suggest players actually leave quickly. Among the players who churn at the level-40 gate, 12% do so right when they reach the gate (they have 0 replays), and 50% leave before 15 replays. Among the players who churn at the level-60 gate, 13% do so right when they reach the gate, and 50% leave before 14 replays. So the argument that players can replay previous levels is moot: players churn quickly and do not replay much of the previous levels when they reach a gate.

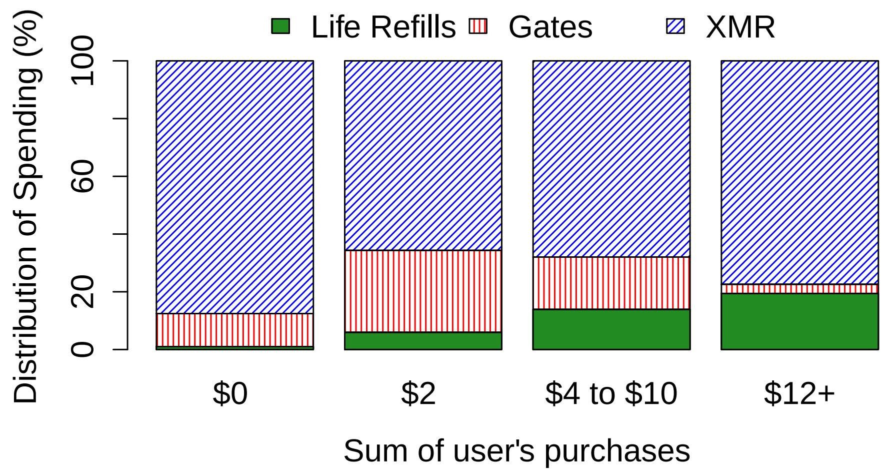


Figure 6.4: Distribution of spending for free players and small, medium, and big buyers.

6.4.2 Purchases and Spending

Similarly to churn, we perform two linear regressions using the 140-row level data. In the first, the dependent variable is the number of purchases made by the average player at a given level. In the second, the dependent variable is the number of coins purchased by the average player at a given level. In both, the independent variables are the level number, the level difficulty, and whether the level is a gate. The results are listed in Table 6.2. Controlling for level number and difficulty, gate levels increase the number of purchases by .093, and the number of coins per purchase by 45, compared to non-gate levels.

Gates are an effective conversion mechanic. In the e-commerce industry, a conversion is the moment when a website visitor makes a purchase [68]. The free-to-play game industry uses the metric as well, but only for the first purchase, not for the subsequent ones, if any. In Jelly Splash, 42% of all conversions happen at the level-40 gate, and 14% at the level-60 gate. These two gates are responsible for more conversions than all other levels combined.

We can also look at spending patterns through Figure 6.4. We call non-buyers the players who do not purchase coins, and small buyers those who ever make a single $2-purchase of 140 coins. Both types of players start with 70 “free” coins. Only 11% of non-buyers spend the 70 coins they start with to unlock the level-40 gate. Maybe they do not anticipate that there is going to be a gate to unlock. Or maybe most churn before reaching the gate. Small buyers are very different. First, 28% of all the coins they spend go into gate unlocks. Second, 44% of small buyers pass exactly 1 gate using 70 coins, and 22% pass exactly 2 gates using 140 coins. Therefore gates are effective at making small buyers spend their coins.

6.4.3 Facebook Logins

Gates are a critical milestone in the game in terms of retention and purchases. But the designers may have introduced them in the game to foster virality as well. To analyze the whole impact of gates on players, we look at Facebook logins, since gates are the first time the game explicitly asks players to login with Facebook and use their Facebook friends.

To ask for keys at level 40, players need to login or be already logged-in with Facebook. Nearly a quarter of players login with Facebook at some point during the game. Are gates responsible for these Facebook logins? Three quarters of Facebook logins happen before the level-40 gate, and only 12% of all Facebook logins happen at the level-40 gate itself. Players who login with Facebook do so on average 3 days after installation (median 5 hours). So at first glance, it seems that players login early or never. But much fewer players reach level 40 compared to level 5. If we control for the number of players who reach a level, figure 6.5 shows that the level-40 gate triggers 9% of the players who reach it to login with Facebook. Facebook logins at the level-60 gate also spike. Hypothesis 3 is confirmed: gates are effective at making players login with Facebook.

If the level-40 gate encourages players to login with Facebook, does it also make logged-in players ask their friends for keys? Table 6.3 compares players who are logged-in with Facebook versus those who are not. For both categories, it lists the percentage of players who pass the gate by asking friends for gate keys, pass by spending the 70 coins offered at the

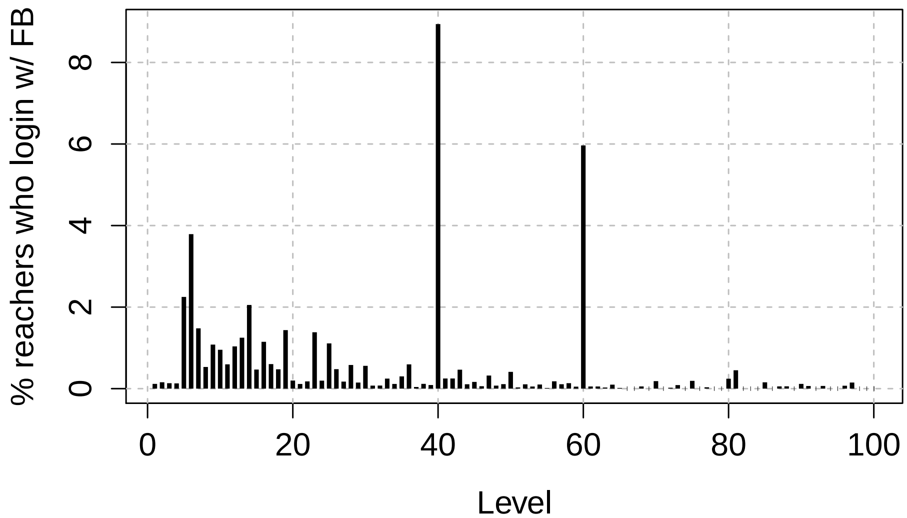


Figure 6.5: Percentage of players who login with Facebook for each level.

	Ask friends	**Use starter coins**	**Purchase coins**	**Never pass**
FB Login	30%	34%	9%	27%
No FB Login	–	49%	5%	46%

Table 6.3: Level-40 gate behavior.

beginning of the game (no purchase required), pass by spending coins they just purchased, and do not pass the gate. Only 30% of the players logged-in with Facebook manage to unlock the gate with friend keys. This could be explained by the fact that a quarter of the players who login with Facebook have no friends playing the game. This itself could be explained by the fact that we only look at the first 10 weeks after the game launched.

Summary: Gates at levels 40 and 60 are effective at converting free players into spenders, and at encouraging players who reach them to login with Facebook. However, they represent a very small fraction of big buyers' spending, and have high churn rates.

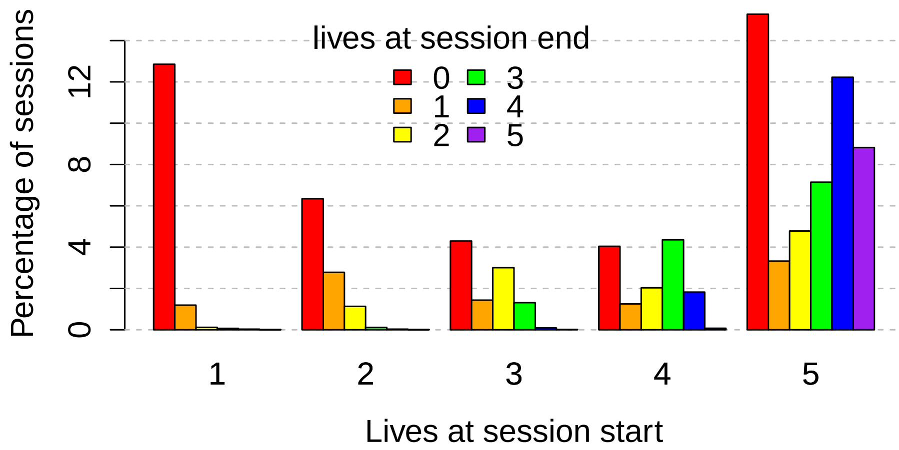


Figure 6.6: Histogram of sessions, split by number of lives at session start and end.

6.5 Energy/lives

We have three hypotheses: 1) Lives do gate player activity. 2) Players are more likely to churn when they reach 0 lives than when they have 5 lives. 3) Players are more likely to buy coins when they reach 0 lives than at 5 lives.

According to Figure 6.6, 43% of sessions end with 0 lives. This confirms hypothesis 1. Moreover, this number also shows that players very frequently run out of lives, and are often presented the possibility to refill their lives using coins. Looking at churn, lives do not seem to matter. When people churn, 20% churn with 0 lives, 55% with between 1 and 4 lives, and 25% with 5 lives. This rejects hypothesis 2. Regarding purchasing, 23% of purchases happen when players have 0 lives, 57% between 1 and 4 lives, and 23% at 5 lives. So we have to reject hypothesis 3 as well.

To better understand the impact of lives on players, we can look at the impact of lives on sessions. Having only 5 lives caps the amount of play per session. Players who lose five times in a row cannot play for half an hour. Much like arcade games of old, players have to spend money to refill their lives and continue playing. Thus in the rest of this section, we look for session factors related to churn and purchases.

6.5.1 Sub-Sampling and Mixed-Effects Regression

We saw above that both chapter gates and spikes in difficulty influence churn and purchases. So we have to include level-specific data such as difficulty when analyzing sessions. The problem is that most sessions involve lower levels (because players churn way before they reach level 140). So if we want to include level variables, we need a way to control for level. Statistically speaking, this means we need a mixed-effects model. Moreover, our outcomes of interest are binary: will the player churn at the end of the session or not, and will the player purchase during the session or not. We use one logistic regression for each of these two binary prediction tasks. Thus, our models are mixed-effects logistic regressions.

To be able to compute these models in a reasonable time, we use a sampling technique known as partial deep telemetry [51]: we decrease our sample size from 5% to 0.1% of the player base, but keep all information concerning these players. If we look at sessions from 5% of the player base, we have more than 7 million sessions, ie 800-MB worth of data. To sub-sample 0.1% of the player base without introducing a bias, we select players whose device id modulo 1000 is zero. Shrinking the dataset to 0.1% results in 5,119 players and 139,017 sessions, or 15-MB worth of data. This smaller dataset makes computations much more tractable while retaining a large-enough sample.

The two resulting models are shown in Table 6.4. In the first regression, the dependent variable is churn likelihood at the end of the session. In the second, it is the purchase likelihood during the session. A + or – indicates that an independent variable is positively or negatively correlated with the dependent variable, controlling for all other variables. Session variables are listed on top, level variables at the bottom. Empty cells have $p > .001$. To run the mixed-effects regressions, we use the package lme4 in the statistical environment R.

Dependent variable	**Churning**	**Purchasing**
Lives at start of session	+	+
Lives at end of session	+	−
Level reached at session end	−	+
Ratio of wins during session	−	+
Session number	−	−
Levels unlocked during session	+	+
Days since install	+	
Games played during session		−
Session duration		
Difficulty of level reached	+	−
Hopelessness of level reached		−
Level reached is a gate	+	+

Table 6.4: Sign of the log-odds ratios from two mixed-effects logistic regressions.

6.5.2 Interpreting Churning

Table 6.4 shows that, controlling for all other dependent variables, players are more likely to churn at the end of a session the more lives they have at the end of the session. This rejects our second hypothesis again. It could be that players do not churn when they run out of lives, but simply when they get bored with the game, maybe after they replayed easier levels in which they did not lose any life. It could also be that players who use all their lives by the end of a session are more engaged by the game, and less likely to churn. In any case, lives as a gating mechanic does not seem to harm retention.

Table 6.4 also presents several findings that are unrelated to lives, but confirm what we saw previously. For example, the likelihood to churn at the end of a session is higher in the first sessions than in later sessions, and it is also higher in earlier levels than later levels. Basically, players either leave the game right away or get hooked. We saw this early-or-never pattern earlier in assist selection in Forza Motorsport 4.

6.5.3 Interpreting Purchasing

As shown in Table 6.4, players are more likely to purchase during sessions that start with many lives and end with few lives. It could be that players who purchase coins use all their lives in one session, and eventually refill and use all their lives again. This reflects a more intense play style. Sessions in which players do not purchase coins may consume only one or two lives, which reflects a slower-paced play style.

As mentioned before, the purchase likelihood increases with the level number. Purchases are also more likely when the level reached is a gate. We saw earlier that purchasing and spending coins happen at the same time. The purchasing model confirms this: when players purchase during a session, they win and unlock more levels. Surprisingly, they also play fewer games, maybe because sessions without purchases involve more replays of easier levels.

Compared to extra move requests, life refills are a luxury item. First, they cost more: 100 coins compared to the 70 coins for an extra move request or to unlock a gate. Second, Figure 6.4 shows that the proportion of spending in life refills increases with the amount of coins purchased. While small buyers spend 6% of their coins in life refills, big buyers spend 19% of their coins in life refills.

Summary: Lives do gate players, but do not seem to have an impact on churning or purchasing. Life refills are a luxury item probably used mostly by the most intense players.

6.6 Summary

This chapter contributes an exploratory study of retention and revenues in Jelly Splash, a free-to-play mobile game. The game uses three main mechanics to gate player progression: spikes in level difficulty, chapter gates every 20 levels, and a limit of 5 lives, regenerating

Category	Finding
Challenge	Difficult levels see more churn.
Challenge	Hopeless levels see more churn.
Challenge	Players rarely replay previous levels when stuck at a gate.
Challenge	Churn decreases with the win ratio during a session.
Time	Half of players churn within 24 hours.
Time	Players churn less at higher levels.
Time	Players churn over time, controlling for level and difficulty.
Time	Lives slow progression, but do not increase churn.
Segments	iPhone users play more sessions, but churn faster than iPad users.
Social	Most Facebook logins happen at the first gate.
Social	Players venting their frustration on social networks creates virality.
Money	Difficult levels see more purchases.
Money	Gates are responsible for 42% of first purchases but cause churn.
Money	90% of coins are spent on extra moves.
Money	Purchases are more likely in difficult levels.
Misc	Churn increases with number of lives at session start and end.

Table 6.5: Summary of findings for Jelly Splash.

slowly over time. We show that chapter gates are the main mechanic making players purchase coins for the first time, that level difficulty is the main driver for player spending, and that lives do gate players, but seem to have little impact on revenues and retention. Our findings suggest the following recommendations for game design: 1) Reducing the difficulty of the most difficult levels may reduce churn. The drawback is that it may also decrease revenues. 2) Players will drop out, the question is when. Players churn both very early and as time goes by, not just because of difficulty. 3) Avoid hard-gating mechanics such as the gates where players have to invite friends or spend money to continue progressing.

Chapter 7

Undisclosed Mobile Game

The last game we look at is another free-to-play mobile game. The data we acquired for that game comes from its Android soft-launch period in 2015. As part of our agreement with the game's developer, we do not disclose the names of the game and its developer. Suffice to say that millions of users installed the game on Android and iOS during the first week of its release. The game is still running as of 2016. Soft-launch data is particularly interesting because it shows how developers iterate to tackle issues of churn, particularly churn happening within the first few minutes of the game.

This chapter has two goals: 1) Replicating the findings of Chapter 6, particularly the influence of level difficulty on churn, and the influence of time. 2) Investigating churn happening before and during the game's tutorial.

7.1 Gameplay

The game's core gameplay consists of acquiring characters, building teams of characters, sending teams in player-versus-environment (PvE) missions or against other players (PvP), and expanding a town. Missions are sequential, and winning one unlocks the next. Each mission consists of three to six parts, and players need to complete all the parts to win. Character rarity ranges from common to rare and legendary, with the rarest characters being the strongest. Players acquire common characters by completing PvE missions, rare characters by completing challenging weekly events, and rare to legendary characters by spending virtual currency in a virtual gacha machine. Characters level up and evolve by consuming other characters or special items.

During PvE and PvP fights, characters attack their opponents in a turn-based fashion. In PvP, players are matched against other players of their ranking, with characters usually of similar power. Winning against other players steals some of their resources, which can be used to expand the town, acquire common characters, and upgrade characters.

7.2 Methods

The game's soft-launch started in mid 2015, and culminated in the game's global launch in late 2015. The game company we worked with gave us access to telemetry data collected during the soft-launch of the game. In the case of mobile free-to-play games, a soft-launch is a period of several months during which a beta version of the game is released only to a particular country. Within few months, developers iterate over several versions to perfect the game before releasing it globally.

7.2.1 Telemetry Data

The game telemetry records the following events of interest in analyzing the tutorial: install, login, tutorial, and level up. All events logged are accompanied with the game's current version, the current time, and the player's device id. This device id is a number assigned by

Event	Data	Rows
Install	d-id, time, version, device model	78k
Login	device id, time, version	2M
Tutorial	device id, time, version, step	4M
Level up	device id, time, version, level	300k

Table 7.1: Structure of the main dataset.

the device manufacturer. It guarantees players' anonymity while making it possible to cross-reference players across tables. The dataset holds no personally identifiable information such as IP or email addresses. Table 7.1 lists for each table the variables used in this study.

We look at data collected during the roughly four months of soft-launch in 2015. During these four months, 11 versions were released, and a total of 78,244 players installed the game. A breakdown of the number of installs per version is provided in Table 7.3.

7.2.2 Defining Churn

Players churn when they do not play the game for at least 14 days. We arrived at these 14 days by following the empirical method used previously in two similar mobile games [108]: If players login on four different days, they contribute three data points. For example, if they login on days 0, 1, 4, and 10 after they installed the game, their periods of inactivity are 0 days (between days 0 and 1), 2 days (days 2 and 3), and 5 days (days 5 to 9). In our sample, 83% of these periods of inactivity are 0 days, 97.8% are 7 days or less, and 99.2% are 14 days or less.

Following the definition of churn above, players churn on day N if they do play N days after install and do not play $N + 1$, ..., $N + 14$ days after install. To reduce right-censoring, we exclude the players who installed the game at least $14 + 1$ days before the cutoff. This decreases the number of installs from 78,244 to 56,876. Figure 7.1 plots the percentage of installs still active up to 6 weeks after install. This graph ignores players who are still playing

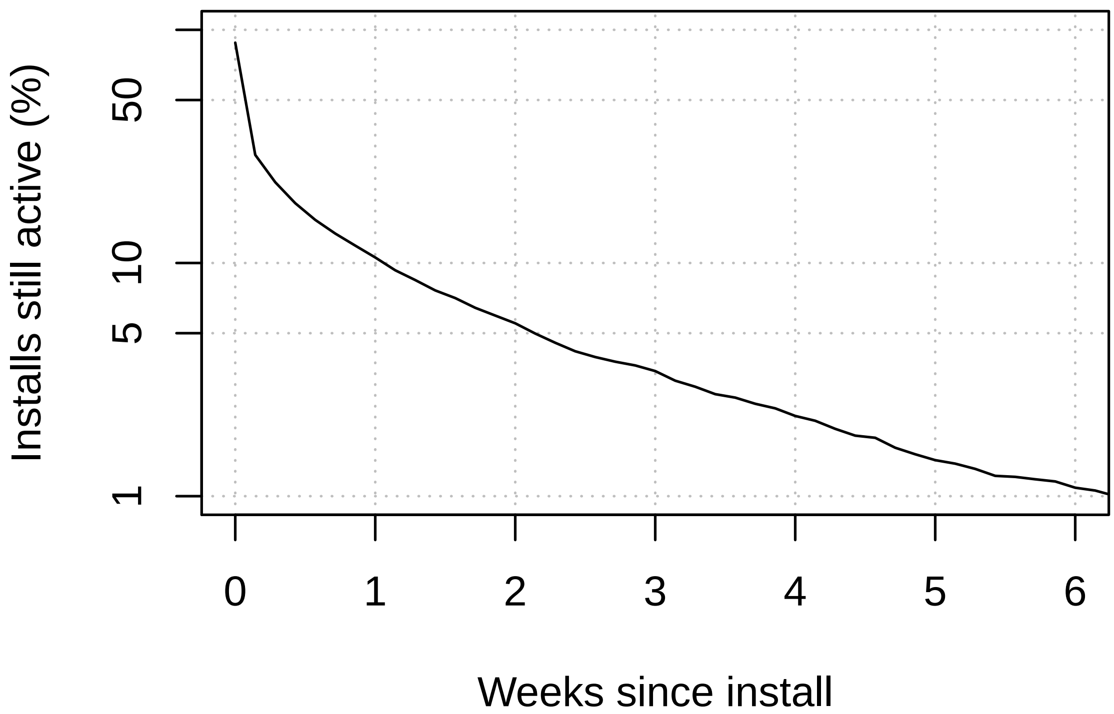


Figure 7.1: Percentage of users still playing against time since install.

at the cutoff, and whose lifetime remains right-censored. Yet it shows that 90% of churners drop out within a week after install, 95% within two weeks, and 99% within 6 weeks.

Figure 7.1 shows a large drop on the first day. Table 7.2 breaks this drop down into 3 steps: before tutorial, during tutorial, and after tutorial. For 100 players who install the game, 89 load the tutorial, 72 complete it, and 45 remain on day 1. In other words, 11% of players churn before starting the tutorial, and 17% during the tutorial. In the remaining of the paper, we first delve into pre-tutorial churn and tutorial churn, and suggest possible explanations for these numbers. Then we replicate the study described in Chapter 6 to correlate difficulty with post-tutorial churn.

Milestone	N	% of installs
Install before cutoff	56,876	100
Load tutorial	50,840	89
Complete tutorial	40,681	72
Play on day 1	25,781	45

Table 7.2: Percentage of players reaching each beginner milestone.

Version	Days Available	Number of Installs	Unique Models	Pre-tutorial Churn (%)	Median Start Time
1	16	12,445	2,928	31	0:21
2	22	3,991	335	3	0:15
3	6	1,992	210	2	0:14
4	8	3,026	357	3	0:16
5	6	1,716	308	5	0:17
6	25	9,340	682	5	0:30
7	5	2,202	369	9	1:01
8	18	18,566	894	6	0:51
9	10	12,672	775	7	0:49
10	3	6,283	459	19	0:17
11	5	5,952	467	5	0:55

Table 7.3: Number of installs, models, churn, and start time across versions.

7.3 Pre-Tutorial Churn

This section focuses on the 11% of players who installed the game and never started playing it. While many explanations are possible, we present two that are suggested by previous literature and supported by the data. First, players cannot start playing the game when their device is not fit to run it. Second, longer loading times make players churn more.

7.3.1 Inadequate Devices

Table 7.3 shows two particularly abnormal numbers for pre-tutorial churn: 31% for version 1, and 19% for version 10. This sub-section builds the argument that both numbers can be attributed to devices unfit to run the game.

In version 1, the number of unique device models helps explain churn. The 12,000 players who installed the game used nearly 3,000 distinct device models. Out of these 3,000 models, 1,720 (57%) represent only one player. In comparison, version 9 had as many installs, but 4 times fewer models (775), and with fewer models representing only one user (183, or 24% of all models that installed version 9).

Developers can use the app manifest to specify filters preventing certain devices to see the app on Google Play[1]. These filters include screen resolution, version of the Android SDK, version of a given software library such as Open GL, and hardware such as GPS, camera, or light sensor. The large variety of device models in version 1 might have happened because some filters were missing. Therefore, many inadequate devices installed the game, could not run it, and are labeled as churning. Filters were likely added in version 2, so that the number of distinct models dropped to a few hundreds, and churn fell to 3%.

In version 10, the churn rate is 19%, but the variety of models is in line with other versions. Taking a closer look, out of 19% churn, 16% comes from a single device model, the Odroid C1, a low-cost single-board computer similar to Raspberry Pi and capable of running Android and Linux for TVs. This model was only seen in versions 9 (4 users) and 10 (1,020 users). This sudden spike in C1 devices may be attributed to the launch of version C1+ around the time that version 10 of the game was released. The remaining 3% churn in version 10 comes from models other than C1. This 3% is a number in line with those found in other versions (barring version 1). The C1 device was not found in version 11, suggesting that another filter was added to the app manifest of version 11. This reduced churn to 5% in version 11.

[1] See `http://developer.android.com/google/play/filters.html`

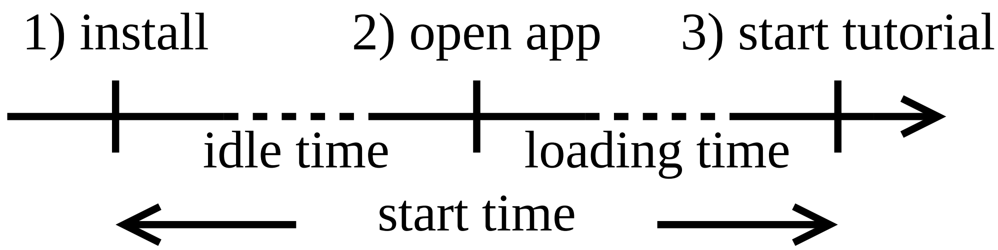


Figure 7.2: Diagram of install-related steps happening before the tutorial.

7.3.2 Longer Loading Times

To be able to link the game's loading time with churn, we face two problems. First, we do not have access to the actual loading time, but only to a proxy for it. As illustrated in Figure 7.2, there are three steps before the tutorial actually starts: 1) the user installs the game, 2) the user taps the icon to start the game, and 3) the user faces the beginning of the tutorial. The loading time is the time elapsed between steps 2 and 3. Unfortunately, the telemetry tracks only steps 1 and 3, not 2. We call start time the duration between steps 1 and 3, and idle time between steps 1 and 2. The idle time can differ between users: some may install the game and play within seconds, others after days. So obtaining the loading time from the start time is impossible. However, it is safe to assume that the median idle time stays constant between sufficiently large cohorts of users. Therefore we can use median start times as a proxy to compare median loading times.

The second problem to link loading time with churn is that the telemetry records no loading time for churners. By definition, churners do not load the game, so they have no loading time. But we can compare the churn rate to the start time. We first exclude versions 1 and 10 for the reasons mentioned in the previous section. A quick glance at Table 7.3 suggests that churn is 2-5% for median start times of 15-20 seconds, and 5-9% for 50-60 seconds. But with only 9 versions to compare, this is only a hint that starting time and churn may be correlated.

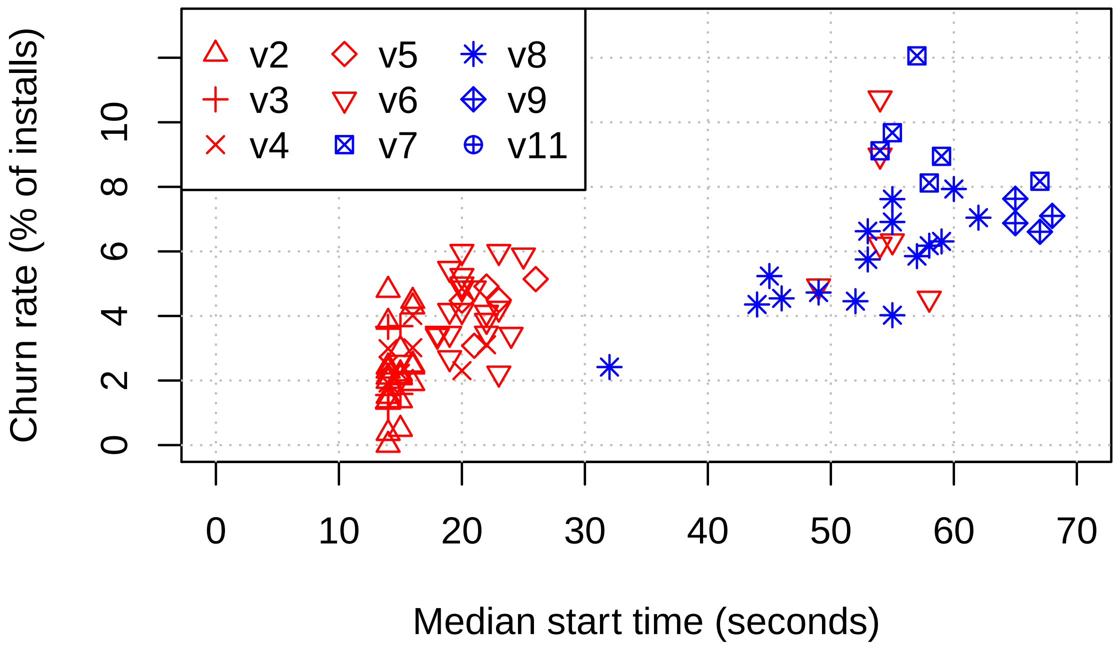


Figure 7.3: Percentage of installs that churned against median start time.

Instead of grouping players by version, we group them by day of install. We ignore versions 1 and 10, and keep only the days with more than 100 installs. Figure 7.3 plots churn against the median start time for these 92 days. Excluding versions 1 and 10, each point represents a day with at least 100 installs. The dot plot shows that the distribution of median start times is bimodal. Up to version 6, the median player takes around 20 seconds to start the tutorial after install. In versions 7 to 11, around 55 seconds. The extra delay added by version 7 can be attributed to patching. Where most other versions were bug fixes, version 7 introduced new content that had to be downloaded and patched when starting the game. This downloading and patching happens in each following version, and adds 30 seconds to the median start time. After splitting the data by day rather than by version, we find a very strong correlation between the median start time and pre-tutorial churn (Spearman $\rho = 0.836$, $p < .001$).

7.4 Tutorial Churn

7.4.1 Step by Step

In the game we are studying, the tutorial is unavoidable and mandatory. There is no tutorial option to select from a menu. When the game finishes loading, players are immediately faced by a character introducing the story. The tutorial is also entirely linear. For each step, players either follow a predetermined sequence of sub-steps, or they churn. At every sub-step, a clear marker highlights the button that players should tap to progress to the next sub-step. Players can never be lost figuring out what to do next.

After reading the introduction to the story, the first battle starts against very easy opponents. A pop-up explains how the basic battle mechanics operate. After the player makes one character attack, the first step of the tutorial is over, but the battle continues. The second step of the tutorial centers on character special powers, with several other pop-ups to read. Using the special power finishes the battle, and concludes the second step. The tutorial goes on. The third step is another battle. In the fourth, players are asked their name and introduced to the town. In steps 5 and 6, they construct buildings and collect resources from these buildings. Step 7 takes players to the gacha screen and gives them a token for a pull. This pull always awards the same rare and somewhat powerful character. In step 8, players reach player level 2, which rewards them two basic gacha tokens, which they must use right away. The two gacha pulls give two common characters. Players then load the campaign screen, and build their team of characters to tackle their first mission in the ninth and last step. The tutorial ends when players start that mission.

Figure 7.4 plots the percentage of player base remaining at each substep. It is a graph frequently found in dashboards in the industry to analyze tutorial churn [43]. Each color represents a different version. We set 100% as the number of players who loaded the tutorial,

Step	Sub-steps
Battle 1	Tutorial loaded, tap character
Special Power	Load and use special power
Battle 2	Tap character, battle ends
Town 1	Intro story, input name, building 1
Town 2	Building 2, collect building 1
Town 3	Building 3, info pop-up
Gacha Screen	Open, receive character, close
Rewards	Mission rewards, level-up
Mission	Load screen, tap stage, build team

Table 7.4: Breakdown of tutorial steps.

ie we are ignoring players who churn between install and start of the tutorial. For example, out of 100 players starting the tutorial in version 10 (dark blue), 96 reach the end of the Battle 1 step, and 80 reach the end of the Rewards step. There are two take-aways from the graph. First, despite minor variations, all versions follow a similar trend, which is why we do not indicate them in Figure 7.4. Second, 8 to 10% of players leave the game during the Town 1 step. By itself, this step is responsible for half of the churn in the whole tutorial.

7.4.2 Churn Over Time

The abnormally high churn happening during the Town 1 step is explained by a single confounding factor: time. Figure 7.5 plots the churn rate of the 9 steps of the 11 versions against their median duration. Although all 9 steps are found in all versions, their sub-steps underwent some changes: some sub-steps were added, removed, shortened, lengthened, or shuffled around. This causes small variations in the duration of a given step across versions. The Town 1 step from version 1 is excluded because it is somewhat of an outlier, with a median step duration of 72 seconds and 12% churn. Later versions increased the median duration of Town 1 by 10 to 30 seconds.

Despite these minor variations, the general trend is that players churn more during longer

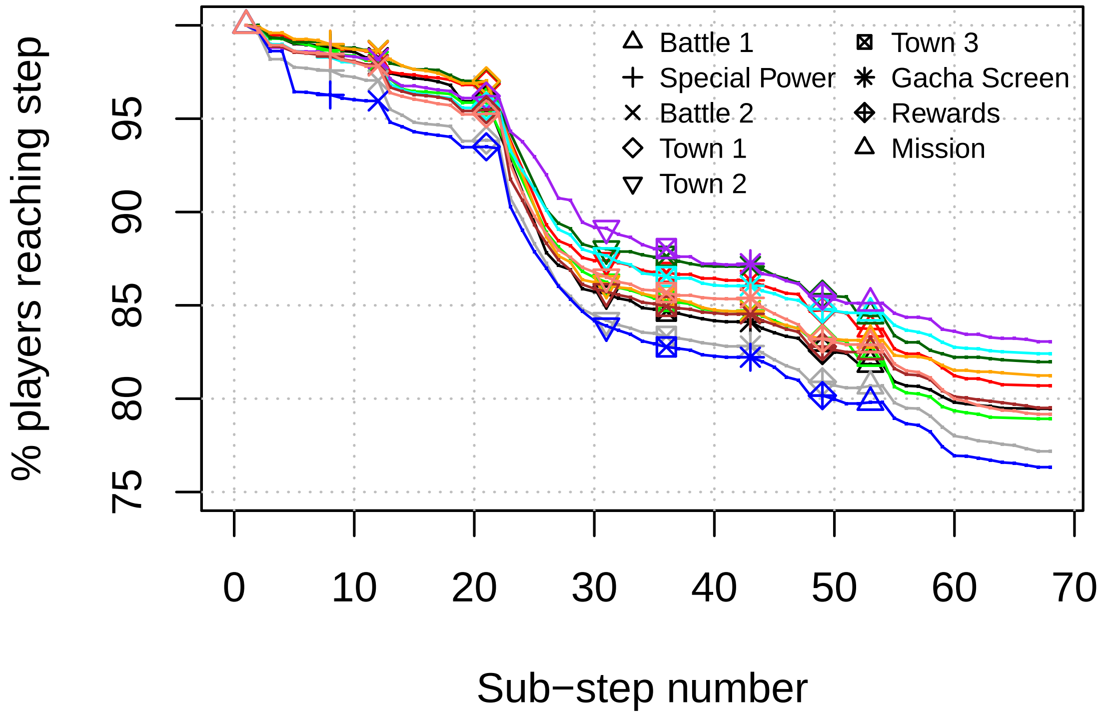


Figure 7.4: Dashboard graph of the players reaching each tutorial sub-step.

steps. In fact, step duration explains 88% of the churn variance between steps (Spearman $\rho = .88$, $p < .001$). The dashed red line in Figure 7.5 represents a linear fit of churn rate against median step duration ($R^2 = .90$, $p < .001$, $\beta = .095$). The beta coefficient of .095 means that 0.095% of players churn for every second spent in the tutorial. Put more simply, for every 10 seconds of tutorial, nearly 1% of players churn.

The residuals plotted in Figure 7.5 show that this linear regression is not perfect: there are three identifiable clusters. The blue cluster is made of the steps Special Power, Gacha Screen, and Rewards. They last less than 20 seconds, and their residuals are slightly above the regression line, which means the model is slightly underestimating churn during these steps. The black cluster is made of the steps Battle 1, Battle 2, Town 2, Town 3, and Mission. These steps last between 20 and 60 seconds, and their residuals are mostly negative, meaning the model slightly over-estimates their churn rate. The red cluster consists of the Town 1 step. The median player takes more than 80 seconds to complete it, and its residuals are mostly

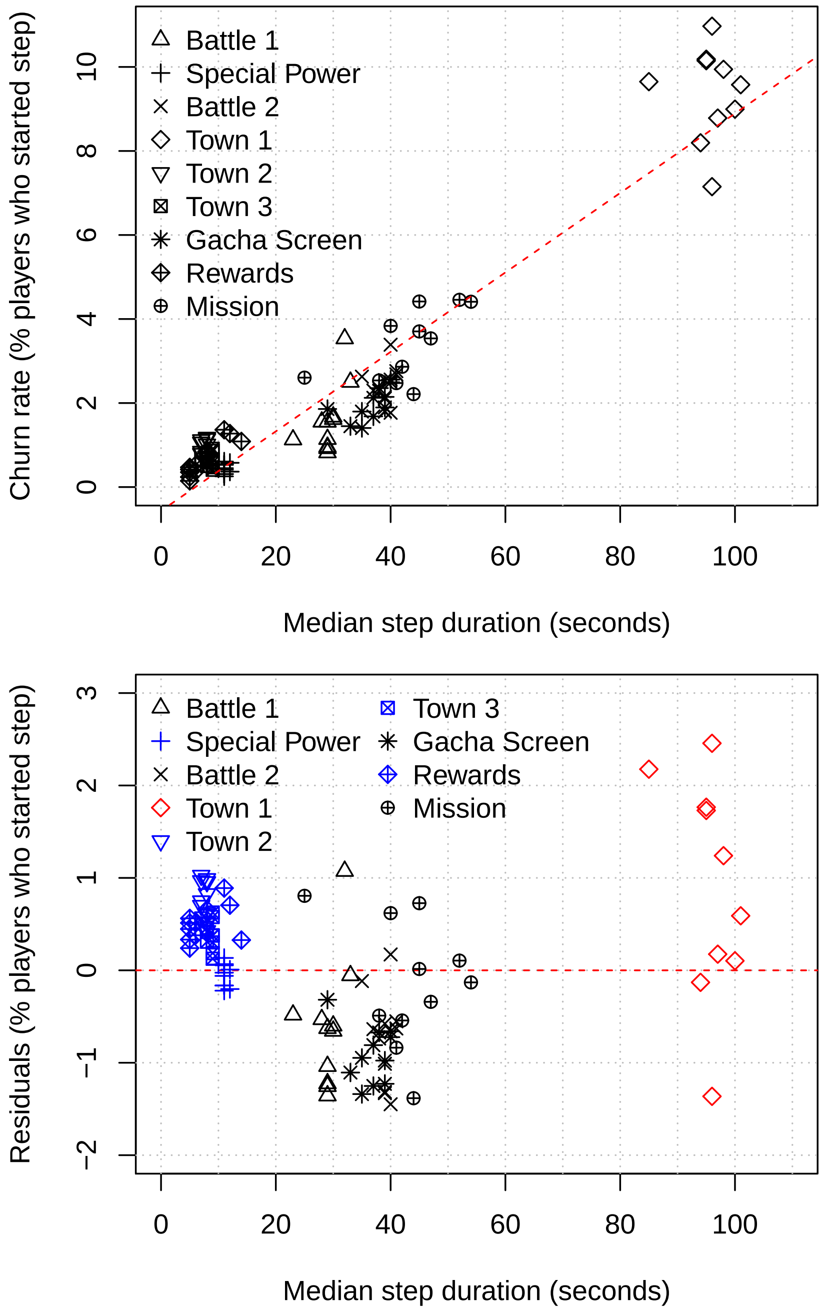


Figure 7.5: Linear regression of churn against median step duration, with residuals.

positive and the largest in absolute value. Knowing why the blue cluster is above the black cluster is difficult. It could be that the blue steps interest players less, and therefore churn slightly more players per second. An explanation for these clusters could help understand the remaining 10% of variance unaccounted for by the linear regression. Overall, this fit is a decent first approximation showing that churn happens over time, and independently of the mechanics introduced.

7.5 Modeling Tutorial Churn

7.5.1 Filtering and Cleaning-up

Based on the previous sections on tutorial behavior, we use the following variables to explain churn on the day of install. Three of these variables deal with player presence in the tutorial. The first one is the **number of sessions** taken to complete the tutorial. The second is the (wall-clock) time taken to complete the tutorial, or **wall-clock completion time**. This is obtained by subtracting the timestamp of the beginning of the tutorial from the timestamp of the end of the tutorial. Note that players who log out of the game before completing the tutorial have to retake it from the start the next time they login. The third presence variable is the **time spent inside the game**. This number is similar to the wall-clock completion time if the user takes little or no break during the tutorial. This number tends to be higher for players who take breaks, because the tutorial restarts from the beginning if players did not complete it the first time they tried.

We also include two variables that are not related with presence in the tutorial. The first is the **start time**, ie the delay between installing the game and starting the tutorial for the first time. We saw earlier that the start time is shorter before version 7. The version number, in particular before vs after 7, could be a confounding factor. So we add a boolean

Variable	Filter	% excluded
Start time	$\leq$12h	2.6
Wall-clock completion time	$\leq$12h	4.9
In-game completion time	$\leq$1h	1.9
Number of sessions	$\leq$4	0.1
Number of restarts	$\leq$4	0.002

Table 7.5: Filtering criteria, and ratio of dataset excluded by each.

checking whether the **version is before 7**.

Before clean-up, the dataset consists of 40,681 players who completed the tutorial. To prevent bias towards outliers, we remove them using the filtering criteria listed in Table 7.5. After applying these criteria, the dataset consists of 36,765 rows, which is roughly a 9% cut. The overall rate of day-0 churners in this cleaned-up dataset is 39%.

7.5.2 Logistic Regression

We want to build a model that assigns 1 to players who churn the same day they install, and 0 to the others. The method we pick is logistic regression (LR) because it is easy to interpret. It is also a simple model with few parameters, which makes it easy to tune and apply to very large datasets. And last, it has been used recently in similar studies of churn and purchases in games [38, 117]. LR's major drawback is that its classification is monotonous. Unless the label progresses monotonously with the input variables, many data points will be misclassified.

The resulting model is in Table 7.6. Holding all other variables constant, a one-minute difference in terms of in-game completion time decreases the odds of churning by 3.7%. Given that the median time spent in-game to complete the tutorial hovers around five minutes, the effect size is very small. Nevertheless, it may suggest a difference between players who take their time, probably reading the text more carefully and appreciating the graphics, and

Variable	Odds Ratio	p
In-game time (minutes)	0.963	$< .001$
Wall-clock completion (hours)	0.978	.026
Version before 7	1.195	$< .001$

Table 7.6: Odds ratio and significance of factors explaining day-0 churn.

those who blaze through the content and skip reading the story and tooltips.

Holding all other variables constant, taking one wall-clock hour longer to complete the tutorial decreases the odds of churning by 2.2%. This effect is small and barely significant, but it may reflect the possibility that players who complete the tutorial in one session churn more than those who complete it in two or more sessions (39% vs 36%, $\chi^2 = 3.52$, $p = .061$). After all, the players with 2 or more sessions liked the game enough during their first session to come back for another one. So it makes sense that players taking 2 or more sessions to complete the tutorial would churn somewhat less than those taking only one session.

The last row of Table 7.6 suggests that holding the in-game and wall-clock time constant, the odds of churning on day zero for players who install the game before version 7 are 19.5% higher than for those who install the game on version 7 or later. There are several possible explanations. Version 7 introduced PvP tournaments. It could be that players particularly enjoyed that feature, and it motivated them to come back. Another reason could be that bugs fixed in earlier versions do not appear in later versions. Players may be more likely to come back if the game has fewer bugs. A third possibility is that by version 7, the game has reached a large-enough player base that makes socializing easier and competing more meaningful. These three possibilities may all contribute to a lower churn rate in versions 7 and above.

Other variables listed in the previous subsection, such as the number of sessions, are not significant in the model, probably because they are redundant. The correlation network in Figure 7.6 helps understand these redundancies. It displays correlations between all six

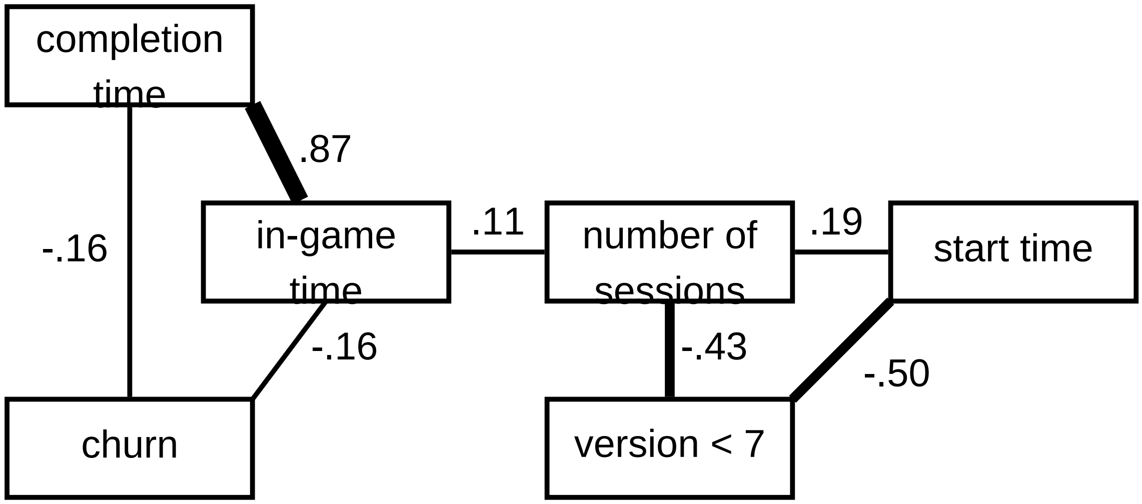


Figure 7.6: Correlation network between variables from the undisclosed game.

variables. Edge labels indicate Spearman correlation's ρ between two nodes. Edge thickness is proportional to the absolute value of ρ. Only correlations with $|\rho| \geq .10$ and $p < .001$ are displayed.

The correlations between (wall-clock) completion time, number of sessions, and in-game time can be explained easily through personas illustrating common behaviors. For example, Alice, a baseline player, starts the tutorial a few minutes after install, and completes the tutorial in one session and in 6 minutes. Her in-game completion time is the same as her wall-clock completion time. Bob's behavior is different. He installs the game and starts playing right away, but for some reason (eg class, meeting, or appointment), he stops playing after 3 minutes, and does not complete the tutorial. Four hours later, he comes back to the game. The tutorial restarts at the beginning, and this time, he completes it in 6 minutes. Bob's two sessions result in a wall-clock completion time of 4 hours and 9 minutes, and an in-game time of 9 minutes. These two personas exemplify why the number of sessions, the completion time, and the in-game time are all very highly correlated ($\rho = .44$, .34, and .87, all $p < .001$). The wall-clock completion time is enough to distinguish Bob's behavior from Alice's behavior. Since the in-game completion time and number of sessions are redundant, their influence in the model is reduced or canceled.

Model	accuracy	precision	recall
Random Guess	50	38	50
Logistic Regression	54	43	56

Table 7.7: Performance of models predicting churn on day of install.

The start time is not included in the model probably because of its large correlation with the version number ($\rho = -.50$, $p < .001$). We mentioned it earlier: version 7 introduced new content that has to be downloaded and patched when the game loads. Since the game always takes longer to start in versions 7 and above, the LR model only needs to know whether the version is below 7. Interestingly, the correlation between start time and churn is significant, but too small to be worth considering ($\rho = .01$, $p < .001$).

To measure the performance of the LR model and see how much extra information it provides compared to a random classifier, we run a ten-fold cross validation. For both models, Table 7.7 reports the accuracy, precision, and recall averaged over the ten folds. These metrics can be interpreted as follows: An accuracy of 56% means that the LR model classifies 56% of all players correctly, and 44% incorrectly. A precision of 43% means that among 100 players classified as churners by the model, 43 actually are churners, and 57 are false positives who actually do not churn. A recall of 56% means that for 100 actual churners, the model detects 56 of them, but misses out 44 (false negatives).

The model performs around 5 points better than a random classifier for all three metrics. This is not much better, but it is comparable to other similar works that reduce error by 8% compared to the baseline [117]. The model also shows that the five variables displayed in Figure 7.6 explain only a small amount of all the churn happening on install day. But along with the correlation network, this model helps characterize typical patterns of beginner churn.

7.6 Post-Tutorial Churn

This section replicates the process correlating difficulty with churn, described in Chapter 6. First we define and correlate level difficulty and level hopelessness with player churn. Then we build a model to investigate in more details which session variables correlate with churn.

7.6.1 Difficulty and Hopelessness

In this subsection, we correlate the level difficulty and hopelessness with churn. These metrics are defined as follows. The churn rate of a given level is the number of players who churn at that level divided by the number of players who reach that level. The level difficulty is computed in the same way as in Chapter 6: it is the average number of tries that players take to win the level for the first time. For example, a level with a difficulty of 2 means that on average, players need 2 tries to win a level – they lose the first try and win the second. Level hopelessness is consistent with Chapter 6 as well, and remains the average of the number of tries to win the level divided by the number of tries "nearly won" until winning the level. In the case of the undisclosed game, levels have between 3 and 6 goals, so we consider that "nearly won" consists of completing all but one of the goals of the level. For example, a level with a hopelessness of 3 means that near-wins happen once every three losses on average. In other words, for three losses on this level, one was at the last goal ("nearly won"), and two were far from completing the last goal. The churn rate, difficulty, and hopelessness, are computed for each of the 100 or so levels.

Difficulty increases with the level number (Spearman $\rho(99) = .21, p = .04$), so we need to control for level number when correlating difficulty with churn. A partial Spearman correlation of level difficulty with churn, controlling for level number, indicates a strong correlation between difficulty and churn (Spearman $\rho(99) = .66, p < .001$). Figure 7.7 plots

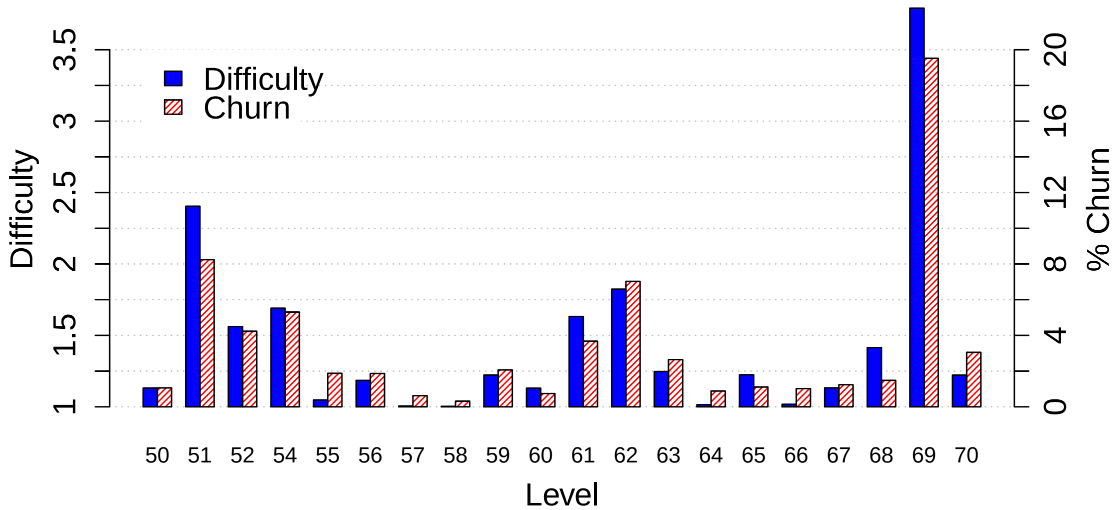


Figure 7.7: Difficulty and churn for levels 50-70 in the undisclosed game.

difficulty and churn against level number for levels 50 to 70. It shows that the spikes in difficulty and churn coincide. Note that the y-axis for difficulty starts at 1 because it is the minimum possible – players need at least one try to pass any level. Also note that level 53 is not displayed in the graph because it is an outlier with abnormally high difficulty (more than 5 tries until won) and churn (more than 42%). Level 53 is included in all correlations and partial correlations nonetheless, since Spearman correlations are robust to such outliers.

Similarly, hopelessness is correlated with level number (Spearman $\rho(97) = .36, p < .001$) and difficulty (Spearman $\rho(97) = .82, p < .001$). Controlling for level number and difficulty, hopelessness is strongly correlated with churn (Spearman $\rho(99) = .40, p < .001$). Figure 7.8 plots hopelessness and churn against level number, and shows that the spikes in churn and hopelessness coincide. Level 53 is excluded again because of its abnormally high churn (42%) and hopelessness (players nearly win less than 1 in every 6 tries).

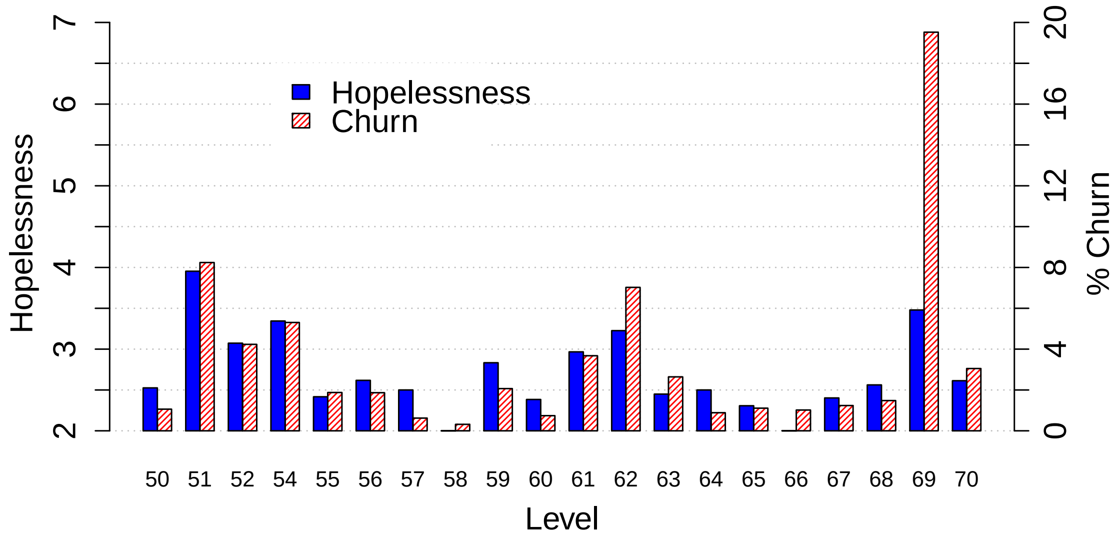


Figure 7.8: Hopelessness and churn for levels 50-70 in the undisclosed game.

7.6.2 Other Session Variables

Besides level difficulty and hopelessness, which other variables correlate with player churn? To answer this question, we look at player sessions recorded by the telemetry. We follow the same process as described in Chapter 6 to build a logistic regression predicting whether a player will churn at the end of each of his sessions. Since there are more than 2 million sessions, we sub-sample the sessions from a randomly selected 1/16th of all soft-launch players. This gives 30,568 sessions from 4,420 players.

For each session of each player, we compute several variables. We saw earlier that levels, particularly their difficulty and hopelessness, are correlated with churning, so we first add to the model **level reached at the end of the session**, **difficulty of level reached**, and **hopelessness of level reached**. Moreover, to check for the influence of time, we add the **session number**, **session duration**, and **number of days since install**. As days pass by, players have more opportunities to play, so session number and days since install are highly correlated (Spearman $\rho(30,566) = .81, p < .001$). We thought of adding the number of sessions played per day by dividing the session number by the number of days since install (plus one, to avoid divisions by zero). However, we saw earlier that 51% of players with at

Dependent variable	Churning
Level reached at session end	–
Ratio of wins during session	–
Session number	–
Levels unlocked during session	+
Days since install	+
Games played during session	–
Session duration	0
Difficulty of level reached	+
Hopelessness of level reached	

Table 7.8: Sign of the log-odds ratios from a logistic regression predicting churn.

least one session churn on their first day (25,781 out of 50,840). For this majority of players, the number of sessions played per day is the same as their session number. Therefore we choose not to include the number of sessions played per day.

Player performance during a session may also play a role in player churn, so we add the following variables. The **ratio of wins during the session** is a direct measure of player performance; if players win all their games, this metric is 1. The **number of games played during the session** is a measure of play intensity at the granularity of the session (as opposed to sessions per day, which is at the granularity of the player). The **number of levels unlocked during the session** characterizes whether players are progressing or stuck; while it is correlated with the number of games played (Spearman $\rho(30,566) = .33, p < .001$), some players may replay previous easier levels and win them all, in which case they play many games but do not progress.

Table 7.8 shows the resulting model. The dependent variable is the player's likelihood to churn at the end of the session, i.e. this session is their last. A + or – indicates that an independent variable is positively or negatively correlated with the dependent variable, controlling for all other variables. Session variables are listed on top, level variables at the bottom. Empty cells have $p > .01$. Note that the effect of session duration (in minutes) is 0.0002, meaning that controlling for all other variables, the odds of churning at the end of a session are 0.02% higher for sessions that last one minute longer. Given that 95% of sessions last less than 49 minutes, this effect is so small that it is practically zero and written as such in Table 7.8.

Table 7.8 also presents several findings that confirm what we saw with Jelly Splash in Chapter 6. For example, the likelihood to churn at the end of a session is higher in the first sessions than in later sessions, and it is also higher in earlier levels than later levels. In short, players either leave the game right away or get hooked.

7.7 Summary

This chapter is an exploratory study of churn using data collected during the soft-launch of an undisclosed free-to-play commercial mobile game. To analyze churn, we break the game in three parts: before players start the tutorial, during the tutorial, and after the tutorial. Pre-tutorial churn can be attributed to devices unsuitable to run the game. Tutorial churn is explained mostly by the passing of time, and not by the type of mechanic introduced, such as battle, reward screen, or story. Post-tutorial churn is correlated with level difficulty, but also with the number of days since install and the amount of progression during a session. The findings from this chapter suggest the following recommendations for game design: 1) Decreasing difficulty may reduce churn. 2) Players will churn at some point. The longer they play, the more likely they are to churn, and there is little we can do about it. 3) Players may churn out of impatience. This can happen for example during a particularly long loading screen.

Category	Finding
Challenge	Bug fixes and/or new PvP feature may increase day-1 retention.
Challenge	Difficult levels see more churn.
Challenge	Hopeless levels see more churn.
Challenge	Churn decreases with the win ratio during a session.
Time	Churn decreases with the level number. Churn early or never.
Time	Players churn over time, independently of difficulty.
Time	Half of players churn within 24 hours.
Time	88% of tutorial churn is explained by time, not mechanics introduced.
Time	During the tutorial, 1% of players churn every 10 seconds.
Segments	Players who take a break during the tutorial and come back churn less.
Infrastructure	Longer loading times make people churn more.
Infrastructure	Devices unsuitable to run the game make players churn.

Table 7.9: Summary of findings for the undisclosed game.

Chapter 8

Lessons Learned

The previous chapters list many perspectives and factors related to challenge, retention, or both. To organize these findings and group them in relevant categories, I built an affinity diagram as follows. First, I made a list of all relevant findings. These are listed in Table A.1 and Table A.2. Then I wrote down each finding on a post-it note, and hung these notes on a wall. These findings are grouped together iteratively, and categories and sub-categories emerge.

In this chapter, I first summarize the findings on the relationship between challenge and retention. Then I provide five broad aspects mediating this challenge-retention relationship: time, segments, social interactions, money, and miscellaneous.

8.1 Challenge and Retention

The literature reports that challenge is the main motivation to play games. This section shows that if players **perceive** that the amount of challenge is too much or too little for their taste, they will churn. When they are provided mechanics to adjust the difficulty themselves, they tend to use them and play longer.

8.1.1 Tolerance to Difficulty

Across all games, the trend seems to be that players with a higher tolerance to difficulty tend to be more engaged. For example in WOW, players more motivated by achievement play longer per week. Players who raid more in a given month are less likely to churn the next. In U, version 7 released PvP tournaments, which provided more incentives for players to compete against each other. When version 7 launched, day-1 retention increased.

8.1.2 Grind with Moderation

However, after having played the same content over and over, the game is not as challenging, and players will tend to drop out. In JS for example, the gates positioned every 20 levels prevent players from accessing the next levels unless they spend money or invite friends. When stuck at a gate, players do replay the previous levels, but only for a very short time before they churn. WOW and RO are MMOs, which means players are expected to grind for equipment by killing the same monsters again and again. In WOW, within 6 months after the release of new content, players have obtained the best equipment they could. By that time, they have either already churned, or are raiding half as much. In RO private servers, the drop rates are higher than on official servers, which makes players obtain equipment faster. This makes players somewhat happier, but may also make them churn faster.

8.1.3 Player Perception and Feedback

Moreover, if the game sends players a signal that they are doing poorly, or if players perceive that they are doing poorly, they are more likely to leave. In JS for example, we measured difficulty as the number of tries to pass a level. More difficult levels have more churn. The game also gives feedback to players in the form of how far they are from completing the level's goals – this was measured by a level's hopelessness. Controlling for difficulty, more hopeless levels have more churn. So not only do players churn more when the level is more difficult, but also when the game "tells" them so. Still in JS, players are less likely to churn at the end of a session if their win ratio throughout the session is high. Put simply, when players win more, they stay more. Another instance of players altering their long-term behavior based on performance feedback is in F4. Players sometimes disable an assist. This can make their racing more challenging. In the race just after players disable an assist, their position at the finish line influences whether they will keep an assist disabled or re-enable it.

8.1.4 Adjusting Challenge

Some game mechanics adjust the amount of difficulty to better fit players' situation. In doing so, these mechanics indirectly increase retention. For example, MMOs are designed for a large player base: some monsters can also be killed in a group, and the economy relies on a stable supply and demand. But RO private servers sometimes have a small player base, which makes the game more challenging, sometimes even frustrating. So private servers provide two NPCs and three commands well-received by players. The two NPCs make the game less frustrating for players who solo. The who command facilitates group building, which makes it easier killing tough monsters. The autotrade and whosell commands make it easier to find a buyer or seller for a given item. They boost the economy, somewhat stabilize prices, and also increase retention by making people login more often to adjust their prices.

In F4, a third to half of players yoyo their assist. They tune their assists, and thus the difficulty, to fit their situation: the traction and stability assists are re-enabled for twitchy cars, and the damage assist when racing against other players. Assist disables are also more permanent when they are followed by easier races, such as in career mode with rewind, or when driving a less twitchy car.

8.2 Time

Time is a construct heavily related to challenge. As players gain skill over time, the challenge offered by the game decreases. Pacing mechanics that increase the game's difficulty as players progress aim at maintaining that amount of challenge. As challenge differs widely between games, so does the pacing and the influence of time itself. Moreover, player preferences tend to manifest themselves early on, and players may get used to playing in certain ways.

8.2.1 Learning Curve and Pacing

What players consider challenging changes over time. As players spend time in the game, they learn new skills, and difficult content becomes easier. That's why designers try to pace the game's difficulty with players' skill progression. In F4 for example, easier assists are disabled faster. Some players may never disable the hardest assists. This longitudinal pattern is also found in WOW. Players defeat easier boss monsters quickly, which give them equipment enabling them to tackle harder monsters, dropping even better equipment, and so on. To facilitate this progression even more, the WOW designers decrease boss difficulty by 5% every month. In JS however, progression is controlled over time: when players lose all their lives on a difficult level, they have to wait for at least 45 minutes before they can try again. This gating does not seem to harm retention, maybe because players expect it

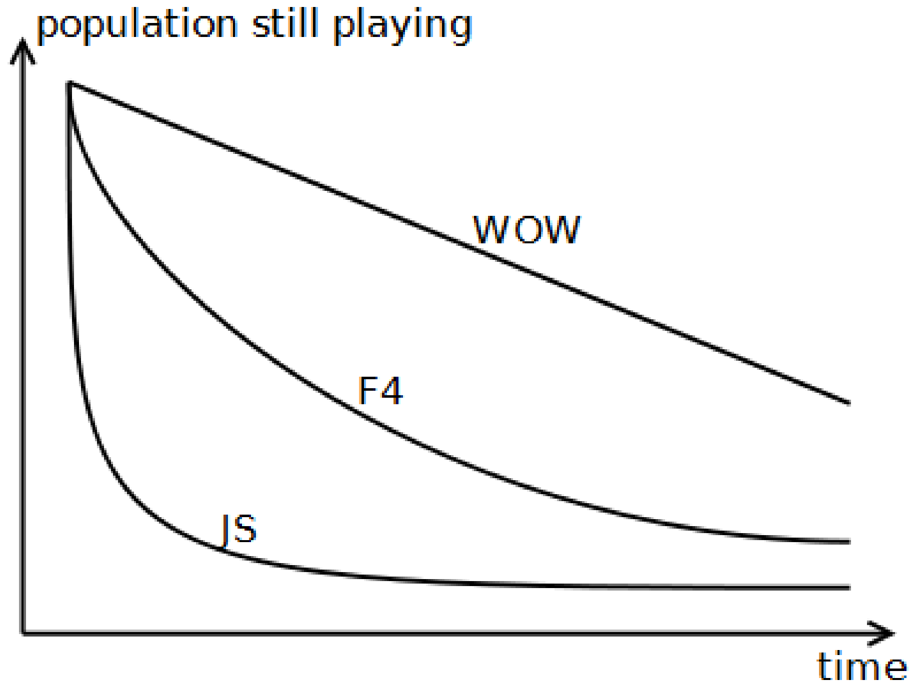


Figure 8.1: Shape of the retention curves for World of Warcraft, Forza 4, and Jelly Splash.

when they pick up the game.

8.2.2 Empirical Diversity

Games are very diverse in terms of game mechanics, player audiences, and business models. Figure 8.1 shows that their retention curves can also vary drastically. For example, we saw that the median player churns after a month in F4, but after a day in JS. It is tempting to conclude that F4 retains its players better than JS. But this is comparing apples to oranges. F4 is a $50 racing game for console, whereas JS a free-to-play tile-matching game for mobile. If anything, plotting together the retention curves of these games highlights yet another facet of video games' diversity.

	Transient	**Consistent**
Intense	The Cheetahs	The Rhinoceroses
Moderate	The Butterflies	The Elephants

Table 8.1: Retention personas accommodating various ways to segment players.

8.2.3 Early or Never

Another pattern noticed across several games are habits, or on the flip side, behavior that happens early or never. In JS, players churn more in earlier levels than in the later ones. Similarly in U, 88% of churn happening during the tutorial is explained by the time spent in it, not by a particular mechanic. In F4, players tend to disable assists early or never as well. The fewer races players take to disable an assist, the more likely this disabling will be permanent and not reverted. Assist disables are also more likely to be permanent the more time has passed since players last raced. This suggests that some players may get used to racing with certain assists. The later they try disabling them, the more difficult and frustrating it will be.

8.3 Segments

Most of the findings related to segmentation can be categorized along two dimensions: intensity (e.g. weekly play time) and consistency (e.g. taking breaks). Table 8.1 summarizes these two dimensions and provide animal-like personas that we explain in the rest of this section. These personas represent different user segments, but also different patterns of activity over time.

8.3.1 The Cheetahs

We call Cheetahs the players who play intensely for a short time and then churn. In WOW, the segments fitting this category are the Chinese, male, and achievement-oriented players: they all exhibit higher weekly play time but also churn faster than their counterparts. For the Cheetahs, the end justifies the means: they are more likely to buy gold to catch up with others or gain an advantage. This competitive behavior is also visible in the high-spenders of JS: most of their coins are spent in extra moves to progress faster. In RO, players of high-rate private servers fit the Cheetah persona: they seem to be more impatient and averse to grinding. They may also be younger. Looking at U's tutorial, the Cheetahs are the players who complete the tutorial faster, which makes them more likely to churn within a day.

8.3.2 The Rhinoceroses

The Rhinos are the stable and dedicated pillars of a game. They have a high tolerance to difficulty, churn little, and spend a lot of time inside the game every week. In F4 for example, the Rhinos are the top 5% of players who contribute half of all races. In JS, the Rhinos would be the players who spend a small amount of money or invite their friends to unlock the gates. In WOW, players who are strongly motivated by immersion fit the Rhino persona: they play more per week, and churn less. Similarly, players who use more characters in a given month spend more time in the game per week, and are less likely to churn the following month. When surveyed, these players label themselves as hardcore.

8.3.3 The Butterflies

Butterflies are the lightweights who spend very little time in the game and quit very quickly. In F4, the "samplers" who race less than 10 times and never change their assists fit this category. So do the nearly 30% of players who never finish the 6 minutes of U's tutorial, or the 50% of players who churn within a day and without spending any money in JS.

8.3.4 The Elephants

Finally, the Elephants are the players who play less intensely than the Rhinos, but are equally consistent. They are somewhat less competitive, and play when they find the time. In F4, 20% of players import data from the previous game in the franchise, Forza 3. They fit the Elephant persona. In WOW, women and American players fit this description too: they play somewhat less per week and churn less than their counterpart. In U, Elephant-like behavior can be observed in two ways during the tutorial. First, players who take their time in the tutorial are less likely to churn within a day. Second, players who take a break and come back during the tutorial are also less likely to churn within a day.

8.4 Social Interactions

The influence of interacting with other players on retention depends on the context and the nature of the interactions between players.

8.4.1 Boosting Retention

In some cases, playing a game with other people can be positively correlated with retention. In RO for example, players on smaller private servers report enjoying their tighter community more than players on larger private servers. In WOW, players who belong to a guild, or who score high on social motivation, may not play longer per week, but they take fewer breaks. In U, the introduction of PvP tournaments may have reduced day-1 churn. Even aggressive interactions between players can help retention.

Moreover, interactions need not happen only within the game to encourage players to play longer. In JS, the designers use the gates stationed every 20 levels to have players invite their friends from Facebook. Even though players can login with Facebook in the 40 levels before the first gate, most logins happen at that gate. JS designers also expect players to vent their frustration on social networks when stuck on difficult levels to foster virality and build a community.

8.4.2 Hell is Other People

In some other cases, people can get in the way of fun. In F4, the AI is somewhat predictable and polite. When racing against it in career mode, around 50% set the damage assist to cosmetic. This number increases to 95% in online mode when racing against other players, because humans are less predictable and their behavior can also be more toxic (e.g. blocking another player from racing by ramming a Hummer into their car). In certain RO private servers, the lack of players pushes administrators to launch the Wapra and Healer. These NPCs can be abused by players who solo camp boss monsters and prevent other players to access them. At best, the Warpra and Healer make solo players ignore the other players who want to play with them. Moreover, Hell can sometimes be restricted to some people and not others. In WOW, players who only play with in-game ties buy twice more gold and churn more than players who only play with real-life ties.

8.5 Money

Several findings indicate that money can play a role in the relationship between challenge and retention. By spending money, players can escape grinding and save time, or gain an advantage and reduce the game's difficulty. In WOW for instance, when players come back from a break, they need to find better equipment. They can buy it from the auction house with game gold. But this gold can only be obtained from completing quests and looting monsters. This takes time, which players may not have. To save time, skip the grind, and step right into the fun part of the game, players buy gold with real money. If they could not buy gold when they come back from a break, they might not come back to play. This phenomenon of acquiring the best equipment with real money is even more flagrant in RO. On free-to-play official servers offering players a subscription for game bonuses, some elitist guilds require their members to pay the subscription. On free-to-play private servers, sometimes a Paypal donation system is in place. Players who donate receive powerful gear. Some guilds require their members to obtain such powerful Paypal gear. In JS, players also spend most of their money in reducing a level's difficulty. More than 90% of coins are spent in extra move requests, and more difficult levels see more coin purchases.

Our findings also suggest that monetizing a game too aggressively can harm retention. In RO, players wondered about the fairness of players buying their spot on a guild's roster. Some may have quit the game for that reason. In JS, the more difficult levels see more revenues, but also more churn. The gates positioned every 20 levels are another example of the trade-off between retention and revenues. The level-40 gate provides the best conversion mechanic, triggering 42% of first purchases, but also increasing churn by 11 points.

8.6 Miscellaneous

Some of the findings raised in the previous chapters do not fit in any of the previous categories, but are related to challenge, retention, or both. In RO, players of the private server with an effective bot detection system said it contributed greatly to their enjoyment of the game. On the official server, lacking such a system, players complained that the server was plagued with bots, and would quit for that reason. In F4, the type of controller that players use to race may have an influence on the assists they use and their overall enjoyment of the game. Disabling the clutch assist, for example, may be frustrating with an Xbox 360 controller but as fun as driving a real car with the actual wheel and pedals. In U, longer loading times, possibly attributed to lower-end devices or the network, result in higher churn before the game even starts. Devices unfit to run the game also contribute to that churn.

8.7 Summary

Based on the several games I have studied, in this chapter I discussed how challenge influences retention, and how this influence is highly contextual. I show that not only challenge, but also players' perception and tolerance of it, impact their experience. Moreover, challenge is not the only factor influencing retention: time, social interactions, and money also play a role in player enjoyment and likelihood to churn.

The literature suggests that engagement is dichotomous: players are either engaged or not engaged. Through two dimensions of engagement, namely intensity and consistency, I have shown that engagement is a concept actually more nuanced. My dissertation also highlights that time, as well as certain key mechanics in a game, can have a strong influence on player behavior.

Most previous work concerning social interactions in games tend to focus on the positive and collaborative aspects of playing with others. This work presents another perspective of multi-player games: although the impact of social interactions on retention is often beneficial, it can also sometimes be detrimental.

Chapter 9

Conclusion

My dissertation relies on several studies, each providing a different context and shedding a different light on the influence of challenge on engagement. The study conducted in RO shows that the difficulty and repetition of official servers make players switch to private servers, where the game not only is easier, but also provides extra mechanics encouraging socialization or enabling solo play. The three studies centered on WOW introduce two metrics for engagement, namely the weekly play time and the duration of breaks. These metrics correlate only little, and suggest that engagement as a construct can be broken down into intensity (high weekly play time) and consistency (frequency and duration of breaks). The WOW studies also highlight differences in play behavior between various demographic segments. In F4, assists are a game mechanic that enables players to configure the difficulty by themselves. I find that players tend to decrease the difficulty as they race, and I identify a surprising early-or-never pattern, where assist disables are more permanent the earlier they happen. Assists disables are also more permanent if they are immediately followed by an easier race. The equivalent of enabling an assist in JS is spending (real) money on extra moves. In the most difficult levels, players do purchase extra moves, but they also churn more. This suggests that there may be a retention-monetization trade-off inherent to the design of free-to-play games relying on difficulty to generate revenues. The business model of free-to-play games also sets the lowest barrier of entry to new players. This can be a blessing and a curse, as a large amount of players churn within the first few minutes of the tutorial in U. This churn is not related to challenge or any particular mechanic; rather, it is mostly explained by time and players' impatience.

Previous academic work, games user research, and designer reflections already suggested that challenge plays a central role in player engagement. Using a variety of approaches on several games, this work confirms that indeed, *challenge influences retention, but context matters*: often, other factors such as social interactions, time, or money, can severely decrease player enjoyment and precipitate churn. This thesis is exploratory in nature, and opens the following avenues for future work in games research:

Broader scope and validation: This thesis makes claims based on five games. Although these games cover a wide range of genres and player segments, the claims made in this thesis can be strengthened, nuanced, or complemented by studies on other games, different genres (such as first-person shooters, which seem much more competitive, or real-time strategy games, which seem less social), more segments (such as children or professional gamers), and an even wider range of methods (such as physiological measurements or diary studies).

Identifying more dimensions of engagement: This thesis contributes two dimensions of engagement. Intensity represents how engaged players are over a certain duration at a certain point in time. Consistency represents how much players come back to the game over long periods of time. In Chapter 4 related to WOW, intensity was measured by their weekly play time when they answered the questionnaire. Consistency was measured by the duration of breaks that players ever took, also when they answered the questionnaire. Chapters 6 and 7 consider that players churn when they do not play for a certain number of days. Given that most players churn out of these games after several days, the number of days of inactivity is a metric related to consistency. In F4 however, as shown in Chapter 5, player lifetime is in the order of months, and so days of inactivity should be replaced by months of inactivity. Another work has suggested a different definition of engagement: the number of days that players logged in during the past 30 days [95], which seems like a measure of intensity rolling over time. There may be more ways to measure engagement and churn. It would be interesting to see if these metrics correlate with each other, or if they capture entirely different aspects of engagement.

Difficulty metrics tends to be game-specific. In games like JS, where players progress through levels one after the other, difficulty increases churn. That is, difficulty as measured by the average number of tries that players take to pass the level for the first time. A similar difficulty metric is the average win rate for the level. One problem with the win rate is that players may replay easier levels with much more skill than the first time they passed it. As a result, the win rate may under-estimate the difficulty of earlier levels. Another metric used in this thesis is hopelessness, or how far players are from completing a level when they lose it. Compared to the number of tries to pass, hopelessness carries an element of frustration. Frustration may also be adequately captured in games with sudden spikes in difficulty as the difference, in number of tries to pass a level, between a level and the previous one, as this metric would highlight sudden spikes and ignore plateaus.

Retention-monetization trade-off in free-to-play games: Chapter 6 shows that this trade-off is noticeable when game revenues come from challenge and progression mechanics such as level difficulty or gates. Some games generate revenues through social or aesthetic mechanics (such as skins and mounts in MOBAs). Given the importance of social capital in social games [34], do numbers show that these aesthetic or social monetization mechanics harm retention the same way that challenge mechanics do? Answering this question is important, because many free-to-play mobile games currently rely on monetization mechanics reducing difficulty or providing advantages against other players.

Bibliography

[1] Activision. Activision Blizzard announces better-than-expected first quarter 2012 financial results. `http://www.prnewswire.com/news-releases/activision-blizzard-announces-better-than-expected-first-quarter-2012-financial-results-150831155.html`, May 2012.

[2] E. Adams. The designer's notebook: Eight ways to make a bad tutorial. *Gamasutra*, June 2011.

[3] E. Adams and A. Rollings. *Fundamentals of Game Design*. Prentice-Hall, Inc., 2006.

[4] M. Ambinder. Valve's approach to playtesting. In *GDC*, 2009.

[5] E. Andersen, E. O'Rourke, Y.-E. Liu, R. Snider, J. Lowdermilk, D. Truong, S. Cooper, and Z. Popovic. The impact of tutorials on games of varying complexity. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 59–68, New York, NY, USA, 2012. ACM.

[6] C. Ang and P. Zaphiris. Social learning in mmo: an activity theoretical perspective. *Interactive Technology and Smart Education*, 5, 2008.

[7] S. Applecline. If you build it, they might come, part three: Encouragement. `http://www.skotos.net/articles/TTnT_89.shtml`, 2002.

[8] M. Ashton and C. Verbrugge. Measuring cooperative gameplay pacing in world of warcraft. In *Proceedings of the 6th International Conference on Foundations of Digital Games*, FDG '11, pages 77–83, New York, NY, USA, 2011. ACM.

[9] J. Bardzell, J. Nichols, T. Pace, and S. Bardzell. Come meet me at Ulduar: progression raiding in World of Warcraft. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, CSCW '12, pages 603–612, 2012.

[10] R. Bartle. Hearts, clubs, diamonds, spades: players who suit MUDs. *The Journal of Virtual Environments*, 1996.

[11] C. Bateman and R. Boon. *Twenty-first century game design*. Charles River Media, Inc., August 2005.

[12] S. Bateman, R. L. Mandryk, T. Stach, and C. Gutwin. Target assistance for subtly balancing competitive play. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2355–2364. ACM, 2011.

[13] C. Bauckhage, K. Kersting, R. Sifa, C. Thurau, A. Drachen, and A. Canossa. How players lose interest in playing a game: An empirical study based on distributions of total playing times. In *Computational Intelligence and Games (CIG), 2012 IEEE Conference on*, pages 139–146. IEEE, 2012.

[14] J. Bell, S. Sheth, and G. Kaiser. A large-scale, longitudinal study of user profiles in world of warcraft. In *Proceedings of the 22Nd International Conference on World Wide Web Companion*, WWW '13 Companion, pages 1175–1184, Republic and Canton of Geneva, Switzerland, 2013. International World Wide Web Conferences Steering Committee.

[15] M. W. Bell, E. Castronova, and G. G. Wagner. Surveying the virtual world-a large scale survey in second life using the virtual data collection interface (vdci). *German Council for Social and Economic Data (RatSWD) Research Notes*, (40), 2009.

[16] J. Billieux, M. Van der Linden, S. Achab, Y. Khazaal, L. Paraskevopoulos, D. Zullino, and G. Thorens. Why do you play world of warcraft? an in-depth exploration of self-reported motivations to play online and in-game behaviours in the virtual world of azeroth. *Computers in Human Behavior*, 29(1):103–109, 2013.

[17] J. Blow. Catching up with jonathan blow. `http://www.gamasutra.com/view/feature/6224/catching_up_with_jonathan_blow.php?page=3`, December 2010.

[18] Z. Borbora, J. Srivastava, K.-W. Hsu, and D. Williams. Churn prediction in mmorpgs using player motivation theories and an ensemble approach. In *2011 IEEE third international conference on social computing (socialcom)*, pages 157 –164, October 2011.

[19] Z. H. Borbora and J. Srivastava. User behavior modelling approach for churn prediction in online games. In *Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom)*, pages 51–60. IEEE, 2012.

[20] S. A. Boyer. *L337 Soccer Moms: Conceptions of "Hardcore" and "Casual" in the Digital Games Medium*. PhD thesis, Georgia State University, 2009.

[21] E. A. Boyle, T. M. Connolly, T. Hainey, and J. M. Boyle. Engagement in digital entertainment games: A systematic review. *Computers in Human Behavior*, 28(3):771–780, 2012.

[22] E. Brown and P. Cairns. A grounded investigation of game immersion. In *CHI'04 extended abstracts on Human factors in computing systems*, pages 1297–1300. ACM, 2004.

[23] S. L. Brown. *Play: How it shapes the brain, opens the imagination, and invigorates the soul*. Penguin, 2009.

[24] M. Busch, E. Mattheiss, R. Orji, P. Fröhlich, M. Lankes, and M. Tscheligi. Player type models: Towards empirical validation. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '16, pages 1835–1841, New York, NY, USA, 2016. ACM.

[25] A. Canossa, A. Drachen, and J. R. M. Sørensen. Arrrgghh!!!: Blending quantitative and qualitative methods to detect player frustration. In *Proceedings of the 6th International Conference on Foundations of Digital Games*, FDG '11, pages 61–68, New York, NY, USA, 2011. ACM.

[26] E. Castronova, D. Williams, C. Shen, R. Ratan, L. Xiong, Y. Huang, and B. Keegan. As real as real? macroeconomic behavior in a large-scale virtual world. *New Media & Society*, 11(5):685–707, 2009.

[27] G. Cheung and J. Huang. Starcraft from the stands: understanding the game spectator. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 763–772. ACM, 2011.

[28] G. K. Cheung, T. Zimmermann, and N. Nagappan. The first hour experience: How the initial play can engage (or lose) new players. In *Proceedings of the First ACM SIGCHI Annual Symposium on Computer-human Interaction in Play*, CHI PLAY '14, pages 57–66, New York, NY, USA, 2014. ACM.

[29] T. Chilton. Blizzard talks the mistakes of training players to hunt for quest markers. `http://www.pcgamesn.com/wow/blizzard-talks-mistakes-training-players-hunt-quest-markers-and-how-pet-battles-won-hearts-community`, August 2013.

[30] D. Chiu. Maximizing player retention and monetization in free-to-play games: Data, best practices and case studies. In *Game Connection Europe*, 2013.

[31] D. Choi and J. Kim. Why people continue to play online games: In search of critical design factors to increase customer loyalty to online contents. *CyberPsychology & behavior*, 7(1):11–24, 2004.

[32] J. Cohen. A power primer. *Psychological Bulletin*, 112, 1992.

[33] M. Consalvo. Hardcore casual: game culture Return(s) to Ravenhearst. In *Proceedings of the 4th International Conference on Foundations of Digital Games*, FDG '09, pages 50–54, 2009.

[34] M. Consalvo. Using your friends: Social mechanics in social games. In *Proceedings of the 6th International Conference on Foundations of Digital Games*, FDG '11, pages 188–195, New York, NY, USA, 2011. ACM.

[35] M. Csikszentmihalyi. *Beyond boredom and anxiety.* Jossey-Bass, 2000.

[36] J. P. Davis, K. Steury, and R. Pagulayan. A survey method for assessing perceptions of a game: The consumer playtest in game design. *Game Studies*, 5(1), 2005.

[37] B. De Schutter. Never too old to play: The appeal of digital games to an older audience. *Games and Culture*, 6, 2011.

[38] T. Debeauvais and C. V. Lopes. Gate Me If You Can: The Impact of Gating Mechanics on Retention and Revenues in Jelly Splash. 2015.

[39] T. Debeauvais, C. V. Lopes, N. Yee, and N. Ducheneaut. Retention and progression: Seven months in World of Warcraft. In *Proceedings of the International Conference on the Foundations of Digital Games*, FDG '14, 2014.

[40] T. Debeauvais, B. Nardi, D. Schiano, N. Yee, and N. Ducheneaut. If you build it they might stay: Retention mechanisms in World of Warcraft. In *Proceedings of the Sixth International Conference on the Foundations of Digital Games*, FDG '11, 2011.

[41] T. Debeauvais, B. A. Nardi, C. V. Lopes, N. Yee, and N. Ducheneaut. 10,000 gold for 20 dollars: An exploratory study of world of warcraft gold buyers. In *Proceedings of the International Conference on the Foundations of Digital Games*, FDG '12, 2012.

[42] T. Debeauvais, T. Zimmermann, N. Nagappan, K. Carter, R. Cooper, D. Greenawalt, and T. Soldberg. Off with their assists: An empirical study of driving skill in Forza Motorsports 4. In *Proceedings of the International Conference on the Foundations of Digital Games*, FDG '14, 2014.

[43] DeltaDNA. Six awesome dashboard charts for f2p games. `https://deltadna.com/blog/6-charts-games-need-on-dashboard/`.

[44] A. Delwiche. How will free-to-play business models affect the gaming landscape in the west? `http://terranova.blogs.com/terra_nova/2006/08/how_will_freeto.html`, August 2006.

[45] J. Dibbell. *Play Money*. Basic Books, 2006.

[46] J. Dibbell. The life of the Chinese gold farmer. `http://www.nytimes.com/2007/06/17/magazine/17lootfarmers-t.html`, June 2007.

[47] R. Dillon. *The Golden Age of Video Games: The Birth of a Multibillion Dollar Industry*. CRC Press, 2011.

[48] M. Doughty, L. Hopkins, S. Lawson, and C. Linehan. Are you the farmer? understanding social game player motivation and sustainability. In *Proc. of the Social Game Studies Workshop at CHI 2011*, pages 1–4, 2011.

[49] Y. Douglas and A. Hargadon. The pleasure principle: immersion, engagement, flow. In *Proceedings of the eleventh ACM on Hypertext and hypermedia*, pages 153–160. ACM, 2000.

[50] A. Drachen, A. Canossa, and G. N. Yannakakis. Player modeling using self-organization in tomb raider: underworld. In *Proceedings of the 5th international conference on Computational Intelligence and Games*, CIG'09, 2009.

[51] A. Drachen, A. Gagné, and M. S. El-Nasr. Sampling for game user research. In *Game Analytics*, pages 143–167. Springer, 2013.

[52] A. Drachen, J. Riley, S. Baskin, and D. Klabjan. Going out of business: auction house behavior in the massively multi-player online game glitch. *Entertainment Computing*, 5(4):219–232, 2014.

[53] A. Drachen, C. Thurau, J. Togelius, G. N. Yannakakis, and C. Bauckhage. Game data mining. In *Game Analytics*, pages 205–253. Springer, 2013.

[54] N. Ducheneaut, M.-H. Wen, N. Yee, and G. Wadley. Body and mind: a study of avatar personalization in three virtual worlds. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1151–1160. ACM, 2009.

[55] N. Ducheneaut, N. Yee, E. Nickell, and R. J. Moore. "Alone together?": Exploring the social dynamics of massively multiplayer online games. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, CHI '06. ACM, 2006.

[56] N. Ducheneaut, N. Yee, E. Nickell, and R. J. Moore. The life and death of online gaming communities: a look at guilds in World of Warcraft. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '07, 2007.

[57] Entertainment Software Association. Industry facts. `http://www.theesa.com/about-esa/industry-facts/`.

[58] L. Ermi and F. Mäyrä. Fundamental components of the gameplay experience: Analysing immersion. *Worlds in play: International perspectives on digital games research*, 37:2, 2005.

[59] J. P. Gee. Learning by design: Good video games as learning machines. *E-Learning and Digital Media*, 2(1):5–16, 2005.

[60] T. Hainey, T. Connolly, and E. Boyle. The differences in motivations of online-game players and off-line game players: A combined analysis of 2 studies at higher education level. In *4th European Conference on Games-based Learning (ECGBL)*, pages 21–22, 2010.

[61] J. Hamari. Perspectives from behavioral economics to analyzing game design patterns: loss aversion in social games. In *Proceedings of the CHI 2011 Social Games Workshop*, 2011.

[62] J. Hamari and A. Järvinen. Building customer relationship through game mechanics in social games. *Business, Technological and Social Dimensions of Computer Games: Multidisciplinary Developments. Hershey, PA: IGI Global, Forthcoming*, 2011.

[63] J. Hamari, L. Keronen, and K. Alha. Why do people play games? a review of studies on adoption and use. In *2015 48th Hawaii International Conference on System Sciences (HICSS)*, pages 3559–3568. IEEE, 2015.

[64] E. Harpstead, T. Zimmermann, N. Nagapan, J. J. Guajardo, R. Cooper, T. Solberg, and D. Greenawalt. What drives people: Creating engagement profiles of players from game log data. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play*, CHI PLAY '15, pages 369–379, New York, NY, USA, 2015. ACM.

[65] B. Harrison and D. L. Roberts. Using sequential observations to model and predict player behavior. In *Proceedings of the 6th International Conference on Foundations of Digital Games*, FDG '11, 2011.

[66] B. Harrison and D. L. Roberts. Analytics-driven dynamic game adaption for player retention in Scrabble. In *Computational Intelligence in Games (CIG), 2013 IEEE Conference on*, pages 1–8. IEEE, 2013.

[67] E. Hazan. Games User Research At Ubisoft Montreal. In *Games User Research Summit*, GUR, March 2015.

[68] B. Heathman. *Conversion Marketing*. 2012.

[69] J. Huang, T. Zimmermann, N. Nagapan, C. Harrison, and B. C. Phillips. Mastering the art of war: How patterns of gameplay influence skill in halo. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pages 695–704, New York, NY, USA, 2013. ACM.

[70] P. Hyman. Puzzle Pirates' James Reveals Free-To-Play MMO Revenue Specifics. `http://www.gamasutra.com/php-bin/news_index.php?story=23970`, 2009.

[71] J. Jansz and M. Tanis. Appeal of playing online first person shooter games. *CyberPsychology & Behavior*, 10(1):133–136, 2007.

[72] N. F. Johnson, C. Xu, Z. Zhao, N. Ducheneaut, N. Yee, G. Tita, and P. M. Hui. Human group formation in online guilds and offline gangs driven by a common team dynamic. *Physical Review E*, 79(6):066117, 2009.

[73] J. Juul. *A Casual Revolution*. MIT Press, 2009.

[74] J. Juul. *The Art of Failure*. MIT Press, 2013.

[75] A. S. Kahn, C. Shen, L. Lu, R. A. Ratan, S. Coary, J. Hou, J. Meng, J. Osborn, and D. Williams. The trojan player typology: A cross-genre, cross-cultural, behaviorally validated scale of video game play motivations. *Computers in Human Behavior*, 49:354–361, 2015.

[76] J. Kawale, A. Pal, and J. Srivastava. Churn prediction in mmorpgs: A social influence based approach. In *Proceedings of the 2009 International Conference on Computational Science and Engineering - Volume 04*, CSE '09, pages 423–428, Washington, DC, USA, 2009. IEEE Computer Society.

[77] J. H. Kim, D. V. Gunn, E. Schuh, B. Phillips, R. J. Pagulayan, and D. Wixon. Tracking real-time user experience (true): A comprehensive instrumentation solution for complex systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, pages 443–452, New York, NY, USA, 2008. ACM.

[78] C. Klimmt. Dimensions and determinants of the enjoyment of playing digital games: A three-level model. In *Level up: Digital games research conference*, volume 246, page 257. Utrecht: Faculty of Arts, Utrecht University, 2003.

[79] C. Klimmt, T. Hartmann, and A. Frey. Effectance and control as determinants of video game enjoyment. *Cyberpsychology & behavior*, 10(6):845–848, 2007.

[80] R. Koster. *A Theory of Fun for Game Design*. O'Reilly Media, 2013.

[81] E. Kraaijenbrink, F. van Gils, Q. Cheng, R. van Herk, and E. van den Hoven. Balancing skills to optimize fun in interactive board games. In *Human-Computer Interaction–INTERACT 2009*, pages 301–313. Springer, 2009.

[82] N. Lazzaro. The 4 keys 2 fun (presentation). In *Game Developers Conference*, GDC, 2004.

[83] Y.-H. Lee and D. Y. Wohn. Are there cultural differences in how we play? examining cultural effects on playing social network games. *Computers in Human Behavior*, 28(4):1307 – 1314, 2012.

[84] V. Lehdonvirta. Virtual item sales as a revenue model: identifying attributes that drive purchase decisions. *Electronic Commerce Research*, 9(1-2):97–113, 2009.

[85] V. Lehdonvirta. A history of the digitalization of consumer culture: From amazon through pirate bay to farmville. *Digital Virtual Consumption. New York: Routledge*, pages 11–28, 2012.

[86] F. Leroy. Mobile diary studies. In *Games User Research Summit*, GUR, March 2015.

[87] T. W. Malone. Toward a theory of intrinsically motivating instruction. *Cognitive science*, 5(4):333–369, 1981.

[88] B. Manero, J. Torrente, M. Freire, and B. Fernández-Manjón. An instrument to build a gamer clustering framework according to gaming preferences and habits. *Computers in Human Behavior*, 62:353–363, 2016.

[89] M. Morhaime. Earnings call transcript of q1 2011 results for Activision Blizzard. `http://seekingalpha.com/article/268912-activision-blizzard-s-ceo-discusses-q1-2011-results-earnings-call-transcript`, May 2011.

[90] L. E. Nacke, C. Bateman, and R. L. Mandryk. Brainhex: A neurobiological gamer typology survey. *Entertainment computing*, 5(1):55–62, 2014.

[91] B. Nardi and J. Harris. Strangers and friends: collaborative play in world of warcraft. In *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work*, CSCW '06, pages 149–158, New York, NY, USA, 2006. ACM.

[92] B. A. Nardi. *My Life as a Night Elf Priest : An Anthropological Account of World of Warcraft*. University of Michigan Press, 2010.

[93] B. A. Nardi and Y. M. Kow. Digital imaginaries: How we know what we (think we) know about Chinese gold farming. *First Monday*, 15, 2010.

[94] B. A. Nardi, S. Ly, and J. Harris. Learning conversations in world of warcraft. In *System Sciences, 2007. HICSS 2007. 40th Annual Hawaii International Conference on*, pages 79–79. IEEE, 2007.

[95] D. Nozhnin. Predicting churn: When do veterans quit? *Gamasutra*, August 2012.

[96] T. Pace, S. Bardzell, and J. Bardzell. The rogue in the lovely black dress: intimacy in World of Warcraft. In *Proceedings of the 28th international conference on Human factors in computing systems*, CHI '10, pages 233–242, New York, NY, USA, 2010. ACM.

[97] PapayaMobile. PapayaMobile Reaches 25M Registered Users, Takes Jabs At DeNA's Mobage. `http://www.insidemobileapps.com/2011/08/24/papayamobile-25-million-dena-mobage/`, 2011.

[98] R. Pardo. AGC: Blizzard's Pardo on WoW's success. `http://www.gamasutra.com/php-bin/news_index.php?story=10773#`, 2006.

[99] R. Pardo. Rules of Engagement: Blizzard's approach to multiplayer game design. `http://wow.joystiq.com/2008/02/20/gdc08-live-from-rob-pardo-talks-about-blizzards-approach-to-mm/`, 2008.

[100] C. Paul. Welfare epics? the rhetoric of rewards in world of warcraft. *Games and Culture*, 5, 2010.

[101] C. Pearce. The truth about baby boomer gamers: A study of over-forty computer game players. *Games and Culture*, 3, 2008.

[102] C. Pearce. *Communities of Play: Emergent Cultures in Online Games and Virtual Worlds*. Mit Press, 2009.

[103] C. Pearce, T. Boellstorff, and B. A. Nardi. *Communities of play: Emergent cultures in multiplayer games and virtual worlds*. MIT Press, 2011.

[104] L. Poisso. Returning to world of warcraft to play cataclysm. `http://wow.joystiq.com/2010/11/04/returning-to-world-of-warcraft-to-play-cataclysm/`, 2010.

[105] S. Reeves, B. Brown, and E. Laurier. Experts at play understanding skilled expertise. *Games and Culture*, 4(3):205–227, 2009.

[106] P. Relan. Interview with the Chairman of Crowdstar. http://lsvp.wordpress.com/2010/02/01/interview-with-chairman-of-crowdstar.

[107] W. R. Rice. Analyzing tables of statistical tests. *Evolution*, 43(1):pp. 223–225, 1989.

[108] J. Runge, P. Gao, F. Garcin, and B. Faltings. Churn prediction for high-value players in casual social games. In *Computational Intelligence and Games*, CIG, 2014.

[109] D. J. Schiano, B. A. Nardi, T. Debeauvais, N. Ducheneaut, and N. Yee. A new look at World of Warcraft's social landscape. In *Proceedings of the 6th International Conference on Foundations of Digital Games*, FDG '11, 2011.

[110] H. Schoenau-Fog. The player engagement process–an exploration of continuation desire in digital games. In *Think Design Play: Digital Games Research Conference*, 2011.

[111] I. Schreiber. Online class on game balance concepts. https://gamebalanceconcepts.wordpress.com/2010/07/07/level-1-intro-to-game-balance/, July 2010.

[112] A. F. Seay, W. J. Jerome, K. S. Lee, and R. E. Kraut. Project massive: a study of online gaming communities. In *CHI '04 extended abstracts on Human factors in computing systems*, CHI EA '04, pages 1421–1424, New York, NY, USA, 2004. ACM.

[113] M. Seif El-Nasr, A. Drachen, and A. Canossa. *Game analytics*. Springer, 2013.

[114] E. Seufert. Profitably Launching Jelly Splash to #1: A Marketing Postmortem (Presentation). In *Game Developers Conference*, GDC, 2014.

[115] B. Shelley. Guidelines for developing successful games. http://www.gamasutra.com/view/feature/131450/guidelines_for_developing_.php, August 2001.

[116] R. Sifa, A. Drachen, C. Bauckhage, C. Thurau, and A. Canossa. Behavior evolution in tomb raider underworld. In *Computational Intelligence in Games (CIG), 2013 IEEE Conference on*, pages 1–8, Aug 2013.

[117] R. Sifa, F. Hadiji, J. Runge, A. Drachen, K. Kersting, and C. Bauckhage. Predicting purchase decisions in mobile free-to-play games. In *Eleventh Artificial Intelligence and Interactive Digital Entertainment Conference*, 2015.

[118] F. Steinhoff. Jelly Splash: Puzzling your way to the top of the AppStore (presentation). In *Game Developers Conference*, GDC, 2014.

[119] D. Stuckel and C. Gutwin. The effects of local lag on tightly-coupled interaction in distributed groupware. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work*, CSCW '08, pages 447–456, New York, NY, USA, 2008. ACM.

[120] M. Tan, L. Q. En, and S. S. Lan. Social gaming: what attracts the most attention?: an investigation using an improved diary method. In *Proceedings of the 2008 international Conference on Advances in Computer Entertainment Technology*, pages 415–415. ACM, 2008.

[121] T. Taylor. Power gamers just want to have fun?: Instrumental play in a MMO. In *Proceedings of the 1st Digra conference: Level Up*, 2003.

[122] T. L. Taylor. *Play between worlds: Exploring online game culture*. Mit Press, 2009.

[123] J. Togelius, R. De Nardi, and S. M. Lucas. Towards automatic personalised content creation for racing games. In *Computational Intelligence and Games, 2007. CIG 2007. IEEE Symposium on*, pages 252–259. IEEE, 2007.

[124] J. Tuunanen and J. Hamari. Meta-synthesis of player typologies. In *Proceedings of Nordic Digra 2012 Conference: Games in Culture and Society, Tampere, Finland*, 2012.

[125] I. Van Geel and B. S. Woodcock. MMOG data: Charts. `http://mmodata.blogspot.com`.

[126] D. Williams. Social whales: Understanding a new kind of player. In *Casual Connect*, 2014.

[127] D. Williams, M. Consalvo, S. Caplan, and N. Yee. Looking for gender: Gender roles and behaviors among online gamers. *Journal of Communication*, 59(4):700–725, 2009.

[128] D. Williams, N. Ducheneaut, L. Xiong, Y. Zhang, N. Yee, and E. Nickell. From tree house to barracks: The social life of guilds in World of Warcraft. *Games and Culture*, 1:338–361, 2006.

[129] D. Y. Wohn. Spending real money: Purchasing patterns of virtual goods on an online social game. In *Proceedings of the 2014 Annual Conference on Human Factors in Computing Systems*, CHI '14, 2014.

[130] N. Yee. The Daedalus Project, vol. 3-5. `http://www.nickyee.com/daedalus/archives/pdf/3-5.pdf`, 2005.

[131] N. Yee. The demographics, motivations, and derived experiences of users of massively multi-user online graphical environments. *Presence: Teleoperators and Virtual Environments*, 15, 2006.

[132] N. Yee. Motivations of play in online games. *Journal of CyberPsychology and Behavior*, 9, 2007.

[133] N. Yee, N. Ducheneaut, L. Nelson, and P. Likarish. Introverted elves & conscientious gnomes: the expression of personality in world of warcraft. In *Proceedings of the 2011 annual conference on Human factors in computing systems*, CHI '11, pages 753–762, 2011.

[134] N. Yee, N. Ducheneaut, M. Yao, and L. Nelson. Do men heal more when in drag?: conflicting identity cues between user and avatar. In *Proceedings of the 2011 annual conference on Human factors in computing systems*, CHI '11, 2011.

[135] V. Zammitto, M. Seif El-Nasr, and P. Newton. Exploring quantitative methods for evaluating sports games. In *Proc. CHI 2010 Workshop on Brain, Body and Bytes*, 2010.

[136] A. Zook and M. O. Riedl. A temporal data-driven player model for dynamic difficulty adjustment. In *AIIDE*, 2012.

Appendix A

Compilation of Findings

The two tables below gather findings related to challenge, retention, time, money, segments, social interactions, and miscellaneous. These findings, as well as various graphs throughout the dissertation, are used to build an affinity diagram and surface higher-level lessons learned.

	Game	**Finding**
	WOW	Achievement-oriented players play longer per week.
C	RO	Warpra and Healer compensate for the lack of players.
H	RO	@whosell and @autotrade increase trade activity.
A	RO	Players prefer higher rates because it reduces grinding.
L	F4	A third to half of players yoyo their assist transitions.
L	F4	Disabling are more permanent if followed by easier races.
E	JS	Difficult levels see more churn.
N	JS	Hopeless levels see more churn.
G	JS	Players rarely replay previous levels when stuck at a gate.
E	JS	Churn decreases with the win ratio during a session.
	U	New PvP feature (and/or bug fixes) increase day-1 retention.
M	RO	Elitist guilds on free-to-play servers require members to pay.
O	WOW	25% of players pay their subscription but do not play for 3-6 months.
N	WOW	Players are more likely to buy gold when coming back from a break.
E	JS	90% of coins are spent on extra moves.
Y	JS	Purchases are more likely in difficult levels.
	JS	Gates are responsible for 42% of first purchases but cause churn.

Table A.1: Summary of findings related to challenge and money.

	Game	Finding
	RO	Higher rates may attract younger players.
	WOW	Engagement and gold buying differ between region, age, and gender.
S	WOW	Asians play harder but churn faster.
E	WOW	Playing more characters decreases the likelihood to churn.
G	WOW	Casual players take fewer breaks than hardcore players.
M	WOW	Achievers, men, and older players are more likely to buy gold.
E	WOW	Players motivated by immersion play longer per week.
N	F4	20% of Forza 4 players are Forza 3 veterans.
T	F4	13% of players change an assist before their first race.
S	F4	Top 5% of players contribute half of all races.
	U	Players who take a break during the tutorial and come back churn less.
	WOW	Boss kills peak and fall over 3-4 months.
	WOW	Dungeon activity 7 months after launch is half that of launch.
	WOW	Every month, 10% of players churn and 5% come back.
	F4	Easier assists are disabled faster.
T	F4	Players tend to disable an assist early or never.
I	F4	Disabling are more permanent if players have not raced in a while.
M	F4	Half of players churn within a month.
E	JS	Half of players churn within 24 hours.
	JS	Churn decreases with the level number. Churn early or never.
	JS	Lives slow progression, but do not increase churn.
	U	Half of players churn within 24 hours.
	U	88% of tutorial churn is explained by time, not mechanics introduced.
	U	1% of players churn every 10 seconds spent in the tutorial.
	RO	Warpra and Healer are popular, but incentivize solo play.
	RO	Players like active and personal community management.
S	RO	Groups encouraged through @who, group size, and level range.
O	RO	Customization and control panel allow players to show off.
C	WOW	Socially-motivated players play longer per week and churn less.
I	WOW	Players in a guild play longer per week and churn less.
A	WOW	Players playing with real-life friends buy less gold and churn less.
L	WOW	Players playing with only in-game friends buy more gold.
	F4	Damage assist is cosmetic when racing other players online.
	JS	Most Facebook logins happen at the first gate.
	JS	Players venting their frustration on social networks creates virality.
M	RO	Players acknowledge and appreciate bot detection tools.
I	F4	Assists like clutch may be disabled only with certain controllers.
S	U	Longer loading times make people churn more.
C	U	Devices unsuitable to run the game make players churn.
	JS	iPhone users play more sessions, but churn faster than iPad users.

Table A.2: Summary of findings related to segments, time, social, and miscellaneous.