JUMPING CAVEMAN: A TOOL FOR MANIPULATING PLAYER EXPERIENCE AND ANSWERING QUESTIONS IN GAMES RESEARCH

A Thesis Submitted to the
College of Graduate and Postdoctoral Studies
in Partial Fulfillment of the Requirements
for the degree of Master of Science
in the Department of Computer Science
University of Saskatchewan
Saskatoon

By

Rasam Bin Hossain

Permission to Use

In presenting this thesis in partial fulfilment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or part should be addressed to:

Head of the Department of Computer Science
176 Thorvaldson Building
110 Science Place
University of Saskatchewan
Saskatoon, Saskatchewan
Canada
S7N 5C9

Abstract

Standard tools exist for assessing player experience; however, there are few tools for inducing play experiences. Game researchers without the resources to operationalize an experimental factor of interest as an implemented mechanic in the design of a custom game are therefore limited in the types of controlled experiments they can conduct. Modifying an existing off-the-shelf game leverages the design and resources of game studio development, but researchers are limited in what type of questions they can ask due to the lack of access control on the source code. We present an open-source system that can be used by game researchers to manipulate player experience in a reliable way and at a finer time resolution than has previously been reported. We simulate the experience of success and failure by providing covert assistance or hindrance to a player, as this has been shown to reliably affect player experience measures. Through three studies, we show that the system manipulates experience in an intended and predictable way. With our system, researchers can also modify the experiment design through simple configuration interface - which allows them to quickly create experiment conditions even if they do not possess technical knowledge of game development. There are many research questions that revolve around the experience of in-game success or failure and our tool allows researchers to ask and answer interesting questions in games research through controlled experiments.

Acknowledgements

I would first like to express my gratitude to my supervisor, Professor Regan Mandryk for her guidance and input throghout my graduate studies. Completing this thesis would be impossible without her continuous support and relentless effort. Secondly, I am sincerely thankful to all other HCI lab members for their friendship and assistance. A big thanks to all the staff in the Department of Computer Science, University of Saskatchewan for being awesome and for helping me in all sorts of ways.

This thesis is dedicated to my beautiful wife, Rahin Sifat who stood behind all my decisions and gave me excellent mental support throughout my graduate studies.

Contents

List of Tables

List of Figures

List of Abbreviations

LOF	List of Figures
LOT	List of Tables
HCI	Human Computer Interaction
GUR	Games User Research
UX	User Experience
pX	Player Experience
PC	Personal Computer
2D	Two dimensional
3D	Three dimensional
IMI	Intrinsic Motivation Inventory
PENS	Player Experience of Need Satisfaction
GSAQ	Game Specific Attribution Questionnaire
HUD	Head-Up Display
URL	Uniform Resource Locator
DDA	Dynamic Difficulty Adjustment
PCG	Procedural Content Generation
AI	Artificial Intelligence
MTURK	Mechanical Turk
MANOVA	Multivariate Analysis of Variance

PREFACE

This particular portion is written to provide a brief description about other people's contribution to my thesis. The research idea, finalizing hypotheses for different studies, re-evaluating and re-constructing the manipulation techniques, different experimental conditions, procedures and data analyses - all of these have been supervised and finalized with the help of my dear supervisor Dr. Regan Mandryk, Professor, University of Saskatchewan. In addition to this, even though, I created the research tool from scratch, Dr. Mandryk helped me by providing her valuable advice at each step of my experiments. On the other hand, the online system that has been used to deploy all the studies (explained at 4.7.1), have been programmed by Colby Johanson, PhD student, University of Saskatchewan. The online system has been further modified by myself to accommodate different survey questionnaires, proper game condition/sequence appearance and other debriefing sessions. During the tool construction procedure, to verify the efficacy of the manipulation techniques and identify the level of game difficulty, I received feedback from other members of the Interaction Lab, University of Saskatchewan.

Chapter 1

Introduction

Although commercial games were released in the early 1970s [59], with the evolution of computers and gaming consoles, processing power has improved, and gamers can fully interact with detailed graphical environments in modern 3D games. Games are becoming more popular than ever and the video game industry has grown to a more than 23 billion dollar business in 2015 [102]. People spend more money on video games than they spend on movies and music combined[1] [29] . Video game products might range from gaming consoles, controllers, hand-held devices, PCs, downloadable content to digital games. At least 52% of the frequent gamers believe that video games provide more value for money than music and movies [102]. U.S. consumers spend 5% of their total household entertainment budget on games which surpasses many other notable entertainment media [82] and 65% of U.S households have a device that they use for playing games [102].

A closer look into these data will make us realize why research on game-play and experience in games is increasing in both prevalence and importance. One of the fundamental parts of successful game development relies on 'Games User Research' (GUR). The main objective of GUR is to understand player psychology by analyzing their behaviour using different techniques of play-testing, player analytics or user experience (UX) analysis. GUR researchers seek to identify players' motivation, how their engagement involves different types of game elements, how their actions can be explained, or simply to create new methodological approaches to capture player statistics to better design games. However, researchers without

[1]Although it might be hard to discriminate between these industries since some of them invest both in games and films. This also leads to some level of arguments about these reports while comparing the revenue of game industry vs the whole film industry [34].

the resources to operationalize experimental factors and implement them in custom game design are sometimes limited in the types of questions they can ask, or the approaches they can use to answer them.

1.1 Problem

One of the main obstacles that games researchers face is how to balance the trade-off between using experimental rigor to control the environment for internal validity (e.g., controlling confounding variables and random participant assignment) with the external validity[2] of having participants feel like they are engaging in play under their own volition, rather than participating in an experiment [11]. In the context of psychological experiments, the extent to which an activity in an experiment is similar to an activity one would encounter outside the laboratory or complete in everyday life is called mundane realism [3, 9] – the greater the mundane realism, the greater the study conforms to a real-life situation.

Some methodological approaches naturally prefer one side of the experimental control-mundane realism tradeoff - ethnomethodologies (such as Nardi's seminal study on World of Warcraft [80]) or the use of data mining (such as in McEwan's analysis of online game lounges [79]) are innately grounded in players' naturalistic play behaviours. However, there are situations in which researchers wish to ask questions that require greater experimental control than is possible using these methodologies. In these cases, researchers have taken differing approaches to balancing the experimental control-mundane realism trade-off. For example, many researchers conduct quasi-experiments in which players are not randomly assigned to groups, but divided based on an individual trait, such as personality [8] or a personal context such as life satisfaction [104], and asked to reflect upon their play experiences. This quasi-experimental approach generally puts more weight on the external validity of studying player experience in the context of real-life play and has the advantage of grounding the data in the actual experiences of the players. On the other hand, prioritizing internal validity and experimental control (including random assignment of participants to conditions) has

[2] the ability to generalize the study results to real-life situations beyond lab settings.

the advantage of isolating factors of interest and has led to many important research findings in games, such as investigating the effects of gamification elements on motivation and performance [75], exploring the effects of immersion and interaction with external affective measures [54], and evaluating momentary and long-term impact of serious experience [49].

When game researchers wish to carefully control experiments, they sometimes work within a game interface to vary a factor of interest, such as when Bowman and Tamborini [13] investigated the intervention potential of computer games for mood repair. Using off-the-shelf games leverages their design and resources, but researchers are limited in what kinds of questions can be asked due to the lack of access to the source code of these games. Alternatively, researchers sometimes take the approach of modifying an existing game through a modding interface, such as when Hoogen et al. [105] created a custom level in Half-Life 2 to investigate physiological responses to in-game deaths. Modding has the advantage of working with a commercial game and all of its resources and has more flexibility than working with what an off-the-shelf game provides, but lacks some flexibility in what can be achieved. Other researchers take the approach of operationalizing the experimental factor of interest as an implemented mechanic in the design of a custom game, such as when Miller et al. [76] created an infinite runner game to investigate the potential of touch pressure of distinguishing at-game versus in-game frustration. This last approach of designing a game to investigate a particular research question does have several advantages, e.g., increased experimental control, operationalization of a particular factor, and flexibility in research question choices; however, it can be difficult or even impossible for researchers who do not have the background or resources to design and implement their own games.

Although standard tools exist for assessing player experience (e.g. [24, 93, 101, 111]), the same attention has not been paid toward methods or tools for inducing player experience [11]. Researchers may wish to manipulate experience for myriad reasons; for example, how subjective[3] and objective[4] difficulty differently affect in-game performance, or to ask which

[3]how players reported their game-play experience

[4]calculated based on their in-game performance

patterns of success and failure foster resilience in players. Previous work on manipulating player experience has successfully used simulated leaderboards to manipulate perceptions of competence and enjoyment [11]. Simulated leaderboards are a subtle approach for manipulating the sense of success or failure in participants and were shown to result in small (explaining 3-9% of the variance) differences.

Even though simulated leaderboards were successful, they have two main problems. First, they worked because of social comparison and thus may not be as effective for people who are less susceptible to - or motivated by - social comparison. Although Bowey et al. [11] showed that the techniques worked as well for people regardless of demographic factors and play experience, other personality factors (e.g., resilience to failure) could moderate the efficacy of simulated leaderboards [85]. The timing of leaderboards at the end of a play condition is fairly low-resolution - you need one leaderboard for every moment that you want to manipulate - making it straightforward to manipulate, for example, experienced competence, but harder to manipulate, for example, experienced competence to study the role of peak-end theory in game experience [38]. Therefore, it was unknown whether the tools that exist previously could work as an effective solution for the games user researchers to manipulate player experience at every point of interest with complete experimental control.

As such the problem that I address in this thesis is as follows: there is currently no effective open source tool available to the researchers to manipulate and induce the feeling of success or failure in the context of a 2D game, in order to conduct controlled experiments.

1.2 Solution

As a solution to the problem addressed above, we present a system that can be used by game researchers to manipulate player experience through controlled experiments and at a time resolution much finer than can be done with simulated leaderboards [11]. Leaderboards worked because they faked the experience of success and failure for the participant by presenting a manipulated leaderboard at the end of each round. We also leverage this idea, but instead fake the experience of success and failure using the approach of providing covert assistance or hindrance to a player, as this has been shown to affect experienced competence, enjoyment, and many other aspects of play experience in several studies using different games; e.g. in the context of a 2D game [5], racing game [18], first-person shooter [27, 109]. Furthermore, the covert assisting of players prioritizes the mundane realism more than leaderboards, which do not exist in all game contexts.

We developed a custom parameterized 2D game named ‘Jumping Caveman’ using *Unity* - a cross platform game engine. The game is created based on the play mechanic used in an F2P[5] mobile game, ‘Spring Ninja’ [60], chosen because of its discrete mechanic of jumping from pole to pole. While developing the system we set three goals that we would accomplish:

1. First, the researchers, future developers and users of the system should be able to understand the game mechanics and in-game objectives with minimum effort,

2. Second, the difficulty manipulation of the system should include both overt and covert mechanisms that can be easily used by the researchers to accomplish the purpose of manipulating in-game success and failure in their experiments,

3. And finally, any researcher should be able to modify the system even if he or she does not posses technical knowledge of game development.

To attain our research goals, we applied covert assistance or hindrance to each pole (in-game objective) – that can vary the experience of the players over an entire level or within a part of the level. Based on specific research questions a researcher might ask, he can also control the resurrection point of the game avatar - controlling which part of the round gets re-played. In addition to this, we also designed a configuration interface that allows non-programmers to implement their custom game prototype with variable game difficulty and ask their specific questions without needing to access the source code[6]. Our system has been tested with approximately 160 participants and the assistance techniques have been refined through several studies to establish the system as an effective player experience manipulator.

[5]Free to play games - a large portion of the video game content is accessible without paying.
[6]Complete source code of the tool - https://github.com/rasambinhossain/Jumping-Caveman

With our system - a researcher would have to invest less time in modifying and creating new experimental conditions[7], considering that the research questions are closely related to in-game success and failure.

Figure 1.1: A snapshot of our 'Jumping Caveman' game. To view the manipulated trajectories, refer to 3.2.2

1.3 Steps to the Solution

There were several steps involved in achieving our research objective to create a system that will allow game researchers to have control over his or her experimental designs without losing the mundane realism of play.

1.3.1 Game Genre Selection

The initial step was to choose the right game genre as it correlates with in-game parameters to control success and failure. We chose a 2D platformer casual game. These games are indisputably popular[8] - mainly because of the following reasons: first, these games require little to no previous gaming experience, the game mechanics[9] are reasonably straightforward,

[7] the number of new game conditions depends on how they modify the configuration interface.
[8] popularity implies the number of downloads or sales record
[9] the set of rules or distinct configuration designed to interact with the game to make it engaging.

and anyone can easily learn the game controls in a few minutes. This appeals to a broad demographic of people and removes the limitations of choosing the right player for the experimental purpose. Secondly, they are playable in short bursts and people can spend a relaxed time without being completely immersed in game-play. They generally have a fairly common theme (move forward by avoiding obstacles and/or collect certain graphical elements like coins or gems) that makes it easier and simpler to play. This makes the system convenient for future games researchers who wish to run studies irrespective of particular demographics or previous gaming expertise.

1.3.2 Constructing the Game Environment with Precise Controls

Our second step involved developing the game environment that is comprised of minimal game controls. We used the play-mechanics of 'Spring Ninja' game since we could use that to manipulate in-game performance. In the original game, the poles are separated by random distances with a variable height. The game starts with a ninja avatar standing on the first pole. The poles are generated dynamically and the objective of the player is to deduce the jump distance and press on their touchscreen mobile device for a particular time period. We replicated these ideas in to our system but instead of creating the game contents programmatically, we developed a configuration file where we set different pole features.

1.3.3 Setting the Game Mechanics

Game Mechanics are specific set of rules that defines how each game elements would interact with the game or how the game would work; e.g., character movement, collision detection, enemy killing mechanism, projectiles they spawn or any other actions, how other game entities interact with each other or how win or loss conditions are decided. Since game performance directly co-relates to the subjective assessment of play experience [69, 70, 11, 99, 112], the next step was to design the difficulty level of our system by adjusting the configuration file parameters. We set the initial difficulty method to be noticeable to the players. Since the main objective of the player is to perform a jump prediction at each press, we configured the pole height and inter-pole distances at certain levels to construct three different difficulty designs. In addition to this overt mechanism, we also implemented a concealed technique to manipulate performance at each jump. Past research [18, 110, 108] have successfully manipulated player performance using different covert assistance and hindrance techniques with significant effects on players' perception of the game-play experience. The hidden assistance/hindrance are applied at each jump by correcting the avatar's trajectory either by keeping it closer to or away from the intended pole (check 3.2.2 for a detailed description).

All other features of the game were controllable through the configuration file. The file acts as an input resource file which can be used to alter other properties of the poles (discussed in chapter 3 - 3.2.3); e.g., pole magnetism (to create a magnetized pole to either attract or repel the players), pole friction (creating a variable of friction levels on the pole surface) and its visual material (altering the pole graphics). Moving forward, we created a database system that would store all the game log outputs which can be analyzed separately with statistical software. Based on the number of features of the system, we formed four separate studies to evaluate the strength of these features and identify whether the system works in regards to the intended experimental design.

1.4 Evaluation

In the primary stage of deployment, a feature called 'Pole Magnetism' [10] was developed as a means for our assistance or hindrance technique. To identify it's effectiveness, we performed a study (described in Chapter 5) by setting two distinct types of magnetism for every pole using the configuration file (the structure of the configuration file is described in Chapter 4) and received player feedback through surveys. We tried to determine if the pole magnetism gives us the desired outcome without raising player awareness. We had four conditions for each of the experimental settings and our participants completed a 17-item survey questionnaire created with two validated sub-scales from Intrinsic Motivation Inventory(IMI) scale [73, 96], one sub-scale from Player Experience of Need Satisfaction(PENS) [93, 88] and one item related to perceived difficulty. Aside from the 17-item questionnaire, they were also

[10] the avatar are attracted to or pushed away from the intended pole when it reaches near it.

asked to fill out their demographics and previous gaming experience related information in a separate form.

Although the perceived player experience were manipulable, the 'Pole Magnetism' feature was not completely concealed as expected and so we revised and further developed a separate type of assistance technique, called 'Trajectory Manipulation'[11] and conducted the next two studies to demonstrate that our system works for altering play-experience in addition to controlling their performance.

In Study 2, we demonstrated that manipulating assistance and hindrance in a covert manner affects player experience in the intended and hypothesized way (Chapter 5). The number of experimental conditions were three: one with added assistance (to aid the avatar to reach at its intended pole), a neutral round (with no assistance or hindrance), and one with high hindrance to push away the player whenever it gets closer. We followed up in Study 2B to show that it still works as well regardless of the difficulty of the game levels, demonstrating that the strength of assistance, and thereby the sense of success and failure can be scaled (Chapter 6). This study also had four conditions similar to study 1, crossed with the two factors of interest- 'Level Difficulty' and 'Degree of Assistance'.

Finally, in Study 3 (Chapter 7), we showed that the system works at a high resolution by manipulating the sequence of assistance. We created two levels with identical pole difficulty and assistance, varying only the placement of the assisted and hindered poles within the level where one condition finished with high assistance and the other one with high hindrance. We confirmed that differences in player experience depends on whether players finished with the assisted or hindered poles, demonstrating that the system can be used to manipulated experience in a fine-grained way. For each of these studies, we also collected the same information from the participants as we did for our first study and before performing any of the studies and participants who had completed a study were not eligible to participate in further studies.

[11] the initial velocity of the avatar is mathematically calculated and corrected for the intended pole.

We evaluated the performance metrics and survey questionnaire for each of the studies and identified whether perceived player experience measure matches with their game performance. In addition to all of the above conditions there was a separate training round in each of these studies to make the participants get comfortable with our game controls and environment.

1.5 Contribution

1.5.1 A Configurable Research Tool

The major contribution of this thesis is a simplified research tool that we verified through several studies. Future game researchers can consider this tool as a means to manipulate player experience at a high temporal resolution. This system could help them to investigate player behaviours, ask questions that revolve around the experience of in-game success and failure, and what type of assistance or hindrance leads to player enjoyment or motivation and how challenge or difficulty can be tailored to manipulate different aspects of player experience.

The research tool is configurable with an access to operationalizing particular factors of interest. This makes it easier to design the tool according to the research questions of a researcher. The configuration interface also removes the limitation of not possessing a particular level of expertise on game development or programming paradigm. Since we developed the tool with varying degrees of assistance and hindrance at a high temporal resolution (where both overt and covert features can be set at every pole), the pole properties can be adjusted to alter game performance and investigate player experience. As a result, researchers can control their experiments without losing the validity of game play experience. Game designers who mainly focus on casual games to test their research hypothesis can use this tool to thoroughly play-test their design (by controlling the configuration file), identify what is the current skill level of any player (based on player performance log), determine what would be the level of assistance or hindrance that a player might need and explore the learning curves of players (based on performance logs) and ask questions about it. There are many interesting research questions that revolve around the experience of in-game success or failure for example, why do some players feel motivated by failure and some feel disheartened? Can we identify a player with 'grit' by analyzing in-game play patterns? How do self-serving attribution biases [78] present in game play? Is challenge or difficulty the more important factor in characterizing peak-end experience in games? Our tool allows researchers with or without technical expertise to ask and answer these types of questions in games research.

1.5.2 How Overt and Covert Difficulty Interacts

The second contribution of this thesis is in our methods for identifying interaction between convert and overt difficulty methods. Researchers can explore the interaction to determine:

- How the player perceives a game as challenging and difficult in terms of achieving in-game objectives while there is both covert and overt difficulty,
- To ask what is the best choice between concealed and objective difficulty to manipulate the feeling of success and failure,
- To determine how players experience covert versus overt manipulation,
- To identify whether they feel any difference between the same level of covert and overt difficulty,
- What type of difficulty plays a significant role on player performance,
- How they interact when their in-game performance drops or gets better.

1.5.3 Additional Contributions

There are also some secondary contributions of this thesis:

- We provide a publicly-accessible research tool that can be used by other game researchers. Future game developers can also contribute to the tool by adding more features to manipulate player experience or altering previous functions to modify the system according to their requirement.
- This tool sends specific game logs that might be particularly interesting to evaluate how novice players react to individual game events like success and failure.

- This tool also logs data to determine whether or not the player would make a successful jump if their trajectories were not manipulated. By using these data, it is possible to detect the learning curve of novice players or the time to reach a desired level of expertise. This idea can be leveraged in to other game genres to develop certain assistance or hindrance techniques to increase player engagement.

1.6 Thesis Outline

Chapter Two presents a survey of related literature reviews on Human Computer Interaction (HCI) and games research. This includes a brief overview on previous research works with custom parameterized games with a focus on procedural content generation, game difficulty and how sequencing affects game-play. Following this we discuss different assistance techniques - mainly static and dynamic manipulation and covert assistance. Finally, we conclude our chapter by reviewing past research on how players' subjective assessment changes based on different set of criteria.

Chapter Three provides a complete description of the research tool that we have developed. This includes the game-play and interface, how a player re-spawns, configurable game features like level difficulty and different types of assistance mechanisms and other minor features (e.g., pole friction). A separate section also focuses on the structure of the configuration file.

Chapter Four describes a general overview of all the studies that we have performed to evaluate our tool. This includes an overview of the participant recruitment platform, training round, pre and post game questionnaires, debriefing sessions, and game performance measures. This chapter also gives a brief overview of the technical properties of the tool and the online system that was used for testing purposes.

From ***Chapter Five*** to ***Eight***, we present our studies to investigate and evaluate our system with online participants. Each of these chapters provides a brief overview of the experimental conditions, procedure, a short description about the participant demographics, hypotheses, data analyses, results, and a short discussion related to the particular study.

In ***Chapter Nine***, we discuss the most important results from chapters five to eight, the implications for game user research in terms of player response to failure, the interplay between covert and overt difficulty, and between player learning curves and players' choice of game. This chapter also provides a short description on the limitations of our work and future possibilities to enhance our tool with an emphasis on Dynamic Difficulty Adjustment (DDA) and multiplayer gaming.

Finally, ***Chapter Ten*** concludes the thesis, summarizing the research tool, its features, our findings and contributions presented in this thesis.

Chapter 2
Related Work

As a background for our work, we will briefly discuss four areas of research that influenced our investigation for the design and development of a custom research tool. At the beginning of this chapter, we will talk about previous research on parameterized games since research tool we developed, can be modified based on different parameters. This section includes the following discussions: 'Procedural Content Generation' or algorithmically generating game content, how game difficulty can motivate player engagement, and how it changes player experience. Next, since we implemented two different types of assistance techniques, we cover the research areas that emphasized different types of static and dynamic covert assistance techniques. Since we used our system to demonstrate that this can be used to analyze sequencing effects in games, we report some of the related background work of sequencing effects. To demonstrate the strength of our tool, we performed several studies with different groups of gamers and provide empirical data to establish the system as a reliable research tool. Therefore the final section of this chapter covers research topics in the area of psychology to understand what stimulates player enjoyment, what drives them to engage with a particular play-mechanic, and how they perceive the hidden assistance techniques compared to overt difficulty methods. This section mainly focuses on the validated scales and behavioural measures that we have used in our studies.

2.1 Custom Parameterized Games

Although we have not procedurally generated our game content through algorithms, previous researchers, who worked with parameterized games have done so using 'Procedural Content Generation'. As such, we review this construct here to consider how it could be applied in our context.

2.1.1 Procedural Content Generation

Using a random or pseudo-random algorithmic procedure to generate game content with limited or indirect human interaction to feed into a game level with desired variability and reliability is known as Procedural Content Generation (PCG) [103, 117, 67]. The content can be anything contained inside a game: characters, weapons, vehicles, levels, maps, rules, textures, items, quests, music, etc. except the game engine itself. A good level design makes the player immerse into a virtual environment and gives the player an opportunity to escape from his or her reality. Manually creating a game space is time consuming and storing the complete game space might consume a massive amount of memory [67]. PCG solves this problem with the aid of different algorithms. There are several reasons why we would generate game content with a tool instead of creating them manually [103]:

- The amount of time required to create the same amount of game content by human designers and artists is significantly higher than working with intelligent design tools and it might potentially delay the development process. A game development company would rather invest into creating an algorithmic method to speed up the content generation process instead of spending more for designers or artists. However, this does not necessarily mean human jobs are threatened with this replacement. Procedurally generating the game content helps indie game developers to generate content-rich games in shorter time spans.

- Game content can be tailored based on the feedback of players to increase game engagement and enjoyment. PCG can be combined with different neural networks to model player responses to create more player-adaptive games.

- Although humans are particularly creative, using the same design team might lead to repetitive content generation. A well-written algorithmic method would be able to create content much faster and without certain limitations that a human could have.

- Creating game content using PCG would help computer programmers to identify different design constraints to better understand game design problems.

PCG has been around for a long time, starting with a 1980s game called ‘Rogue’ [116, 66]; whenever a new game started the game content were recreated randomly. Although this algorithmic technique for game content generation has passed several decades, PCG is still used in several commercial games. 2D platformer games like Super Mario Bros [84] can be reconstructed based on different design patterns [23] whereas Spelunky [77], another popular indie game created different variation of game levels using PCG. The F2P game ‘Spring Ninja’ on which we were dependent on creating the game mechanics, also used a random technique to generate game levels with separate background each time the ninja dies. Although PCG-based games have benefits, they sometimes lack controllability. Researchers who wish to systematically control game content and design game parameters at certain order in order to investigate certain player behaviours might not benefit from algorithmic game content generation method like PCG.

In PCG, the design of level generation or instantiation of in-game content is unpredictable to the players; hence, they are not completely controllable. To address this issue with PCG, Yannakakis et al. [118] developed the Experience-Driven Procedural Content Generation (EDPCG) framework. The framework adapts to the player experience based on the cognitive and affective responses obtained through different physiological sensors. It also assess the quality of the game content, represents the game levels in vector dimension and links to the respective player model. The framework also searches through the game space for optimized player experience.

Compton and Mateas [20] proposed a level design algorithm with a four-layer hierarchy model to generate platforms of new game levels by looking into 4 type of patterns (basic, complex, compound and composite). Later on, Jennings-Teats et al. described the implementation of a 2D platformer ‘Polymorph’ game to create a distinctive player experience by structurally designing the levels using a machine learning-based strategy [55]. Polymorph game creates a model based on player’s current skill level from play traces. These traces were collected from a web-based tool where the same player rates multiple short level segments by completing them one at a time.

Furthermore, Shaker et al. [97] carried out a number of experiments in order to identify what game features could be extracted from game content to shape players' in-game behaviours, how these can be used to construct a player model, and how tailored game content generation is possible based on these results. The research identifies three emotional effects with high accuracies - engagement, frustration and challenge, reported by players. On the other hand, Isaksen et al. created a framework for a parameterizable game to create an optimal game space to explore the relationship between game design and player experience [52, 53]. Upon selecting different variants of game parameters, they performed automatic play-testing by generating and simulating a level by a player model called Monte-Carlo simulation. Based on exponential survival analysis, their techniques could search for unique game difficulty versions, create game visualization and even optimize parameters for next a simulation. However, their AI-based survival analysis looked only into the single dimension of score distribution that was independent from other game space parameters like time pressure or game speed. The accuracy of the framework did not have an impact on these other parameters to which a real-life player would definitely react to cope with the challenges faced.

Even though with all the advantages, the PCG technique might not be the best case for designing new levels since the expertise of players might vary and developing new skills will not be reliable anymore. Due to unmatched expertise, even the best-designed games become uninteresting to the gamers. Modern games tends to perform matchmaking to solve this issue. For example, Halo [22] uses a Bayesian algorithm to match players based on their skill or 'StarCraft II: Wings of Liberty' [10] selects opponents from same expertise tier.

Another technique to deal with this particular issue is to take player performance into consideration at a high resolution and tweak the game parameters to keep the challenge level optimum to player expertise. However, in some cases even they are designed by the game designer's intuition and focus only on player performance rather than responding to a specific play pattern. In addition to this, changing the parameters also alters the level of difficulty.

Several researchers have tried to address different issues with game difficulty to identify what type of difficulty relates to player engagement and other psychological aspects. The level of challenge can be either predefined or adaptive to player response. In a multi-player environment games are often unbalanced, which can be adjusted by providing assistance or hindrance to the players [109, 110, 108]. However there had been only a little attention paid toward the investigation of interaction between the level of difficulty and assistance or hindrance. Hence, we seek to investigate and find out the interplay between them.

2.2 Game Difficulty

By definition, the term 'Difficulty', can be exemplified by the average error or failure rate of any experimented item assigned to a particular population [62]. Although difficulty and challenge are often referred interchangeably [1], in this thesis, we will use the word difficulty as 'probability of failure to complete a task'. Similarly when it comes to accomplishing game objectives, the greater the probability of failure, the more challenging the game is. Humans are naturally more motivated to engage in to an activity that provides greater difficulty even if the reward of a less challenging activity remains the same [44, 45]. This motivation is driven by the need to experience competence, and a challenging task is more satisfying to complete. However, players will opt out of games that are too easy or too hard. This behaviour can be explained by Csikszentmihalyi's theory of 'flow' [1]. The theory describes a particular state where the gamers are completely focused in an experience in such a manner that they will ignore any other distractions and lose self-consciousness. It does not matter if the game is against other humans or in solitary mode, when it is well-matched with player's skill, he can fully immerse in that experience without feeling anxious or getting bored. This flow state had been widely embraced by several researchers of game enjoyment [71, 94, 100, 119].

In games, players often get to choose the difficulty level at the beginning of the game and the internal game mechanics (e.g., speed of the level, number of enemies in the scene, opponents' shooting accuracy, enemy search radius, players' health drop rate, presence of audio or visual cues, number of game objectives, in-game auto-save features) gets modified based on the se-

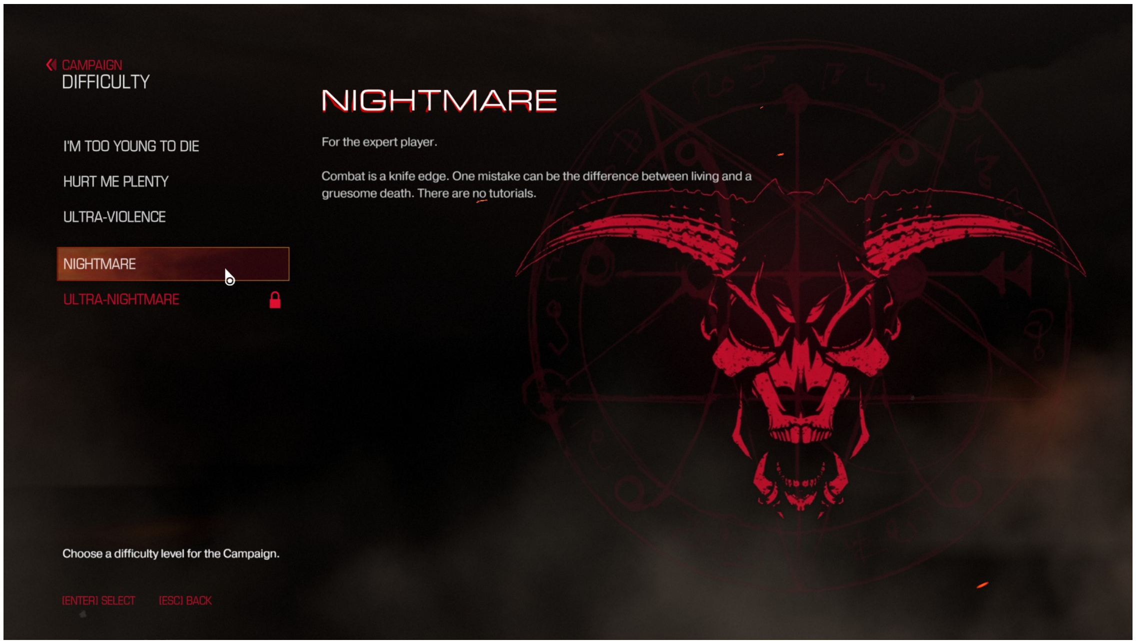


Figure 2.1: Game Difficulty selection screen in Doom (2016)

lected level. Since, it is the responsibility of the gamer to identify his correct difficulty level, an appropriate guess might not be always a feasible option. If a game is completely new to a player or he is unaware about his expertise level, he could end up choosing the wrong one [48]. Some games overcome this situation by offering a training round with medium difficulty and later on allowing players to choose their desired level of difficulty. To make the situation less complicated, recent games (e.g., Hitman [51], Doom (2016) [50], and The Witcher 3: Wild Hunt [17]) also let players change the difficulty during game-play. Even though from the development perspective, a harder level comes with greater challenge, a player's perception of difficulty and game enjoyment might vary depending on his adaptation to the game environment.

To investigate further, Spiel et al. [99] illustrated how game performance relates to a player's subjective assessment about game difficulty and perceived fun. They developed five algorithms that adjusted the difficulty of the game by choosing blocks of a Tetris game based on the current state, instead of adjusting game speed. In their experiment, eleven out of sixteen players reported the game as being more fun and enjoyable when they perceived it as less difficult.

Work of Wehbe et al. [112] looked into the effect of manipulating game parameters (scroll speed, target size, jump task complexity, perspective) in 2D platformer games. As expected, when the scroll speed intensified and the target size dropped, the players perceived the game as more difficult. Interestingly, increments in jump complexity also resulted into an initiation of increased difficulty but that related to player's adaption with the rhythm of the game.

Additionally, Lomas et at. [69] investigated several factors of game design on difficulty, motivation and learning. Their purpose was to identify if that generates the same curvilinear inverted U-shape relation between player enjoyment and difficulty as proposed by flow theory [1]. Surprisingly, their observations lead to a different result in which the maximum engagement occurred when the game was easier and people self-selected to play it for longer time. However, this condition did not support optimal learning. The biggest limitation was that the self-selection of game choice might be biased as players might prefer relatively easier conditions and abandon the hard games.

In addition to this, Lomas and his colleagues [70] performed three separate experiments on players' choices of difficulty, novelty and suspense to figure out if there are other factors of challenge that generate the inverted U shape. Flow theory [1] suggests that players optimal enjoyment depends on the difficulty of a game that is not too easy or not too hard. Lomas et al. [70] also suggested that games being easy is not the only cause of boringness, possibly the reason of failing is consistent repetitiveness and being uninteresting. Their evidence showed that letting players choose their game difficulty agrees with the flow theory and to maximize motivation it needs a moderate degree of novelty. Additionally, increasing difficulty consistently, decreases motivation where other motivational factors are controlled.

Although altering game-metrics is a common approach to change the game difficulty by controlling the in-game performance of players, Bowey et al. [11] used a different approach to manipulate player experience. They manipulated the feeling of success or failure using simulated leaderboards without altering the game performance. Their system worked despite previous gaming expertise or other personal traits. The experimental analysis showed that their approach of reinforcing the position in leaderboards had significant main effects on perceived competence, autonomy and presence [101]. On the other hand, emphasizing position with colour feedback had strong interaction effect on autonomy and enjoyment.

2.3 Assistance Techniques

Game difficulty and assistance are highly correlated, and assistance can be used to improve the performance of a weaker player. While difficulty adjustments are mainly used to slightly alter the game controls and other mechanics to reduce challenge [4, 18], assistance directly helps players to master skills and overcome challenges [5, 110, 109]. Many researchers have already devoted themselves to find out the implications of different type of assistance techniques and how these can be used to alter game experience. In this section, we will briefly discuss some background information on two types (static and dynamic) of assistance methods. However both subtle [5, 18, 110, 108, 109] and disclosed assistance [27] have successfully manipulated player performance and play experience.

2.3.1 Static and Dynamic Assistance

Static manipulation techniques are controlled by the initial game configuration and are designed based on the results of different tests performed on players with different expertise or by player simulations. Games, which use this technique frequently, are mostly targeted to be played in solitary mode. This type of assistance mainly changes the game parameters in a certain way that alters the difficulty level and makes the game objectives easier to accomplish.

On the other end of the spectrum, adaptive manipulations are done in real-time and depend on the participating player's in-game performance. Some games have attempted this technique of integrating adaptive assistance into their game. Dynamic adjustments have been around for a long time starting from the 1980s game Astromash [1]. A more recent game like

[1] A video game for the Intellivision game console, 1981.

Resident Evil 4 [16] rates its players on a scale of 1 to 10 and adjusts enemy behaviours or their damage level accordingly. However, even at the beginning, players get to choose a difficulty and later the dynamic adjustments are made (e.g., normal difficulty will lock the players at grade 4 - increased to grade 7 for doing well or decreased to grade 2 for performing poorly).

Researchers have tried to explore how static [61] or adaptive manipulation [41] techniques can maximize player engagement and persistence. To identify a relation between game difficulty and maximal game engagement Khajah et al. demonstrated their findings [61] by modifying two popular games - 'Flappy Bird' and 'Spring Ninja'. They moved beyond the popular well-known concept of 'difficulty affects engagement' by looking into the effect of covert vs. overt assistance. Furthermore, they also discovered that players attribute their improved performance resulting from the covert manipulation to their own competence, whereas overt manipulations alone are not sufficient enough to alter the level of engagement.

2.3.2 Covert Assistance

Alternatively, difficulty can be adjusted without the player's awareness. One of the popular techniques is to aid the weaker player(s) in a multi-player environment by either assisting him or by hindering the performance of the stronger player(s). A growing body of research has investigated concealed or hidden (covert) assistance techniques as a tool to influence player performance and subsequent player experience. Most of these studies used assistance and hindrance to balance play, effectively bridging the gap between two players with different skill levels. Studies investigating aiming assistance in first-person shooter games influenced player performance without being noticeable to players [5, 110, 108]. Assistance techniques have been shown to successfully balance a game between players by hindering the experts and assisting the novices [109, 110, 108]. Findings, also suggest that manipulating player performance through assistance increases weaker players perception of competence and fairness [109], and increases enjoyment for both players [37, 109]. Similarly, research with players of different expertise who competed against simulated drivers in a racing game found that combining several assistance and hindrance techniques successfully manipulated the players performance making the game more balanced [18]. Both experts and novices preferred the hidden balancing techniques. In addition to balancing between experts and novices, recent research has also shown that a combination of assistance techniques can overcome differences in physical abilities while also having the desired effect on player experience [36].

Literature on covert assistance techniques has previously discussed two possible downfalls to this approach. First, assistance of players might thwart their progression at becoming better at the game. However, a recent study investigating long term effects of assistance has found that assisted players gain the same amount of skill compared to unassisted players [39]. The second concern was the possible negative effects of players becoming aware of the assistance. The studies implementing assistance techniques appear to have successfully kept their players from noticing assistance [5, 18, 109]. Although some researchers claim that assistance disclosure could be harmful to player's experience [33, 18], Depping et al. carried out a study to investigate their conventional wisdom. They found that even if players are made aware of assistance, the positive effects of player balancing still prevail [27].

2.3.3 Sequencing Effects

Work on static and dynamic assistance generally applies aid to an entire condition. However, there are reasons to believe that performance within a play session affect play experience. The theory of peak-end effects tells us how people assess any experience that has a strong connection to the peak and end events of that experience. Peak and ending events of any task have an overriding effect on the retrospective and momentary evaluation of that affective experience. It is common for people to comprehend task completion time in a non-linear way [2] that includes various biases. Their perception of task duration can be explained by the peak moments during the task and the concluding experience [2, 47]. Kahneman and his colleagues showed that there is an overweighting on stimuli that had been perceived by participants in their final moments [58, 89, 90]. They performed a 'cold-pressor' experiment where they had their participants arms submerged into cold water for either short or long duration. In both of the conditions, participants submerged their hand in identical 14°C water. The only difference between these two conditions was, in the longer version they pulled out their hands after an additional of half minute but in this short period the water was slightly warmer (15°C). Even though the longer version was still painful, when asked, participants preferred the longer condition - where they could finish the experiment with less pain. However, besides these unpleasant ending moments, other researchers have investigated people's retrospective evaluation on other factors too [81, 68, 28]. Harrison et al. [42, 43] experimented with different type of progress bars and found out that a users choice of progress bars is exaggerated toward the end of the procedure (end moments) when the bar showed the highest acceleration of progress. In another study, Cockburn et al. [19] showed that users' subjective evaluation of an interface is influenced by the experiences they encounter at the most intense (peak) and end moments. To achieve this result, they combined both peak and end moments of that particular experience.

Subsequently, the effects of event sequencing have long been considered in the context of game design; for example, the placement of a boss fight at the end of a level, or use of cut scenes to pace experience [86]. However, there has been little systematic study to understand how sequencing of events affects player experience. Petralito et al. studied the experiences of players of Dark Souls III [87], showing that players' perceptions of positive experiences and enjoyment particularly peaked right after achievement and victories (defeating the boss), despite the struggle it took to get there. Gutwin et al. [38] systematically examined player responses to peak and end experiences in games by manipulating the sequence of difficulty and challenge in several games. The game sequences had an identical overall difficulty and equal possibility of success or failure over the whole experience. However, participants perception of their enjoyment, fun, and preference of replay were biased to series ending with positive outcomes or higher challenge.

2.4 Measuring Player Experience

In recent years, researchers have investigated the experience of players based on the subjective evaluation of a player's perception in the context of Games User Research (GUR); for example, in the context of different genres of games [6, 11, 18, 21, 109], different type of people during gameplay [8, 12, 35, 57], analyzing how player experience differs while playing in a social multiplayer setting vs solitary mode [106], interaction with strangers and collaborating in a short time [64], multi-player engagement with players of asymmetric roles [40], exploring different game input devices that leverages player interaction in ubiquitous settings [31] or using a combination of inputs [107], in the context of Virtual Reality [14], investigating the efficacy of different control schemes and effects on player experience while playing exertion based games [98] and leveraging gamification in the context of life experience [15]. There are many factors that are known to systematically affect the experience of players. Most of these come from studying how players were affected by a factor of interest - rather than how to affect players proactively for research.

It is no doubt that games are quite entertaining and have a massive audience who regularly invest into buying gaming products. Researchers have been investigating to find out why games are so compelling or engaging regardless of any physical or external reward and how well they can achieve the game design goals by observing and understanding player psychology through play-testing. In this section, several theories related to understanding and measuring player experience are presented.

2.4.1 Self-Determination Theory

In operationalizing effects on player experience (pX), many researchers tend to draw from self-determination theory (SDT) [25, 91]. Its primary focus is to differentiate between the ideas of intrinsic and extrinsic motivation, that describes how people are intrinsically motivated to perform an activity due to the enjoyment of the experience itself rather than seeking any external reward or separable outcome [92]. Since games are engaging and have a long-lasting effect on players [63], Ryan et al. further explored SDT in the context of video games [93] and showed what drives players to engage with games can be explained by these concepts. Based on the number of people playing games [83], it is presumable that video games are inherently appealing; players often engage with them solely because of their intrinsic motivation. However, players motivation might wane over time with a decrement in effort and enjoyment. But introducing a reward mechanism would react negatively and decrease the task performance for players who identified themselves as more-motivated at the beginning of the gamification task [7].

2.4.2 Intrinsic Motivation Inventory (IMI)

How participants assess their subjective experience related to any study performed on them can be often measured using a multi-dimensional measurement scale called the Intrinsic Motivation Inventory (IMI) [73, 96]. The complete list has a total 45 items with seven sub-scales in it. However, there are also some other versions of this IMI scale that has smaller number of items with three to five sub-scales. A brief description about the sub-scales are written below:

1. **Interest-Enjoyment**: This sub-scale identifies the level of enjoyment or fun a participant achieved during the execution of the task (the only sub-scale that measures self report of intrinsic motivation). The complete list contains 7-items under this sub-scale with two of them reverse-coded.

2. **Perceived Competence**: How satisfied the participant feels about their task performance or how skilled they think at the activity. The complete list contains 6-items under this sub-scale with one of them reverse-coded.

3. **Effort**: A separate measure for the participants to report on the level of invested effort. Participants report on how hard they tried or how much energy they invested in to this task. The complete list contains 5-item questionnaires under this sub-scale with two of them reverse-coded.

4. **Pressure-Tension**: The only negative measure for intrinsic-motivation. This identifies how much anxious or nervous a participant felt during the task completion. The complete list contains 5-items under this sub-scale with two of them reverse-coded.

5. **Value-Usefulness**: How participants internalize the activity as beneficial or important to themselves. The complete list contains 7-items under this sub-scale with none of them reverse-coded.

6. **Perceived Choice**: Whether participants voluntarily choses the activity or not can be demonstrated using this sub-scale. The complete list contains 7-items under this sub-scale with five of them reverse-coded.

7. **Relatedness**: Based on the activity (if two or more persons are involved while performing the same task) this sub-scale can be added in to the IMI questionnaire. This questionnaire items focuses on how participants interact or feel connected to the other person or how trustworthy they were. The complete list contains 8-items under this sub-scale with four of them reverse-coded. (However, this is still not a validated sub-scale [96])

Based on the theoretical questions a researcher is addressing, he can choose whatever factors or measurements he would like to ask his participants about a particular task or activity and construct the whole IMI questionnaire based on these factors.

2.4.3 Player Experience of Need Satisfaction (PENS)

An antecedent of intrinsic motivation is need satisfaction, which is commonly measured using the Player Experience of Need Satisfaction (PENS) scale specifically for games [93, 88]. There is no argument that playing games is a fun activity but it is important to look closely and explore what leads to experienced enjoyment. This scale measures players' perceived competence (mastery of control), perceived autonomy (feeling of agency over the games outcome), perceived relatedness (connectedness other players), immersion (the extent to which players are transported to the game world), and intuitive controls (how simple the controls were to learn and use) [101].

2.4.4 Game Specific Attribution Questionnaire (GSAQ)

Attribution is an individual's desire to find and explain the causal effects of things happening around them [46, 115]. Previous research has investigated player's attribution in the context of games and when it comes to game-play, players tend to rationalize their game performance. So understanding how players attribute certain events during game-play provides deep insights about their play experience. To explain the perception of players' causal beliefs, Depping et al. developed the Game Specific Attribution Questionnaire (GSAQ) [26]. GSAQ assesses four dimensions of attribution theory [113, 114, 115] :

1. **Internality**: This questionnaire differentiates whether a particular achievement in game is attributed internally (the player himself is responsible) or that the reason lies external to the player (e.g., the player blames another agent, situation or the particular game system). This is also known as 'Locus of Control' (e.g., a player always fails in the boss fight because he thinks the system gives additional advantage to the boss by giving him extra health bars).

2. **Stability**: This specific measure determines how players attribute the consistency of in-game performance over time.

3. **Controllability**: The dimension of Controllability identifies how strongly a particular player believes in his or her ability to volitionally control or alter his or her performance (e.g., a player describes the cause of failing to achieve the game objective due to lack of effort in a particular game round).

4. **Globaility**: The last dimension relates to the player perception that success or failure in the game can be attributed globally to related tasks or specifically to a particular event in the game.

Chapter 3

Jumping Caveman: A Tool for Studying Player Experience

To be able to manipulate player experience, we created a game, 'Jumping Caveman' with a range of features including covert assistance and hindrance that can be varied as a priori, or in real-time, as needed, to change player perception. One of the biggest advantages of this research tool is that it can be adjusted according to the desired research questions to accommodate a researcher's hypothesis in order to manipulate player experience at a fine resolution.

3.1 Overview of the Game

To establish it as an effective research tool, we first performed a requirements analysis and identified what type of features, control sequences and environments were essential at the initial phases of development. The game environment was created to represent a simple casual platformer game since this would help future developers and researchers to be able to easily use the tool with minimal effort. A complex game environment might make it harder to allow them to modify parameters without proper guidance, and the interaction between different game objects might become a multi-level hierarchy problem (game objects are concurrently dependent to each other and modifying one object effects other objects as well). Hence, we chose single mouse interaction as our game control. Secondly, the game objective should be easily comprehensible, even with no previous experience in that domain. Since it is likely to be modified by researchers and will be available for studying player experience, we wanted a clean interface instead of building complicated graphics. Finally, while we were developing our tool, we planned to create a secondary interface that would require very minimal technical knowledge to modify. This would allow the researchers to modify the game objects according to their research objectives without diving into the code base.

3.1.1 Gameplay — Interface and Controls

The game starts with a caveman character standing on the first pole, awaiting player input. The objective of the game is to move forward by pressing and holding down the left mouse button to power up the caveman's jump. A power bar in the top left corner of the screen (see figure 3.1) indicates the current jump power and fills up in real-time to provide insight into the initiated jump level. To properly land on the poles with an adequate level of friction, we set the vertical jump velocity slightly different to each other (for our particular research purpose, we set a constant horizontal velocity and a variable vertical velocity- which can be modified in the code-base if needed). When the mouse button is released, the caveman jumps with the appropriate amount of power, as indicated by the power level. The power bar becomes empty, immediately after the jump initiation [1].

All poles have varying heights and distances, that can be changed to alter the difficulty of the game rounds. A higher variation in heights and distances makes it harder for the player to predict the next jump and this increases the chances of miscalculation. Alternatively, little or no difference in the pole heights and distances make accurate jump level prediction easier to do. Each consecutive jump adds one point to the player's score, and if they jump over multiple poles, extra points for each skipped pole are also added to the player's score. The scoring mechanism can be described using the following example: if the player jumps from 'Pole 1' to 'Pole 2', the current score will be '01' since he jumped only one consecutive pole. Now, if the caveman jumps from 'Pole 2' to 'Pole 5', he will be jumping over 'Pole 3 and 4' - which results in additional points on top of the regular '03' points. The current score would become '07' now (6 points awarded for this jump). The score is displayed at the top right corner of the screen, which initially started from zero and resets whenever the player starts

[1] Power Bar shown in figure 1.1 is just for demonstration purpose

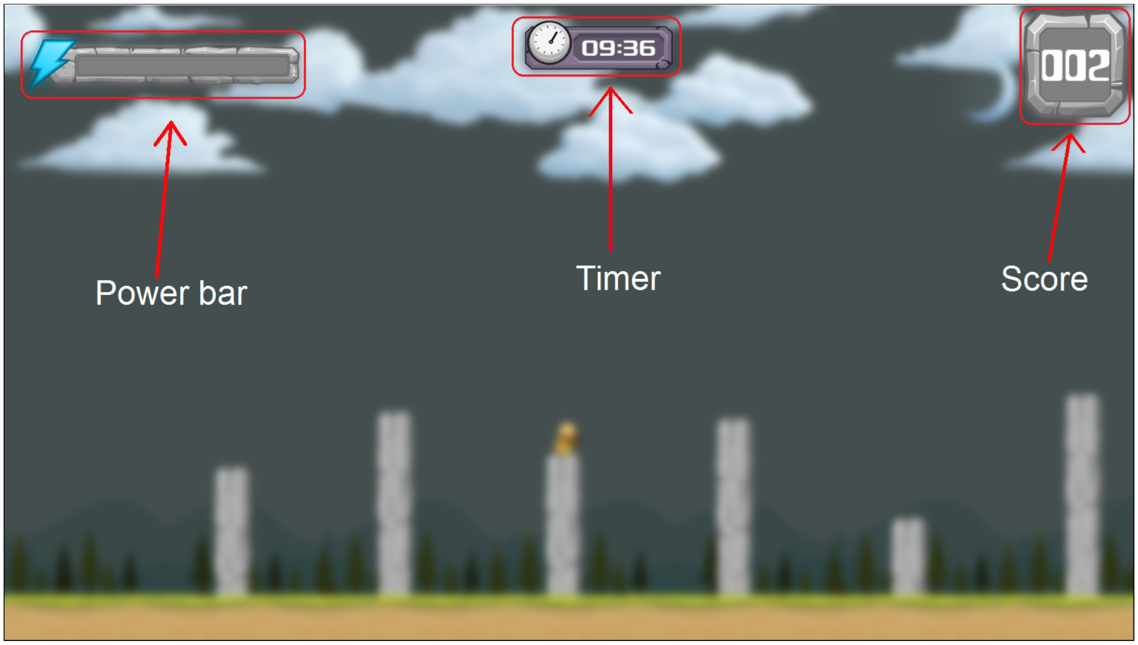


Figure 3.1: Game HUD.

from the first pole.

A timer in the upper middle section of the screen (see figure 3.1) indicates how much time is left in the current level. The players can play as many times as they want until the timer runs out, at which time they are directed to the next phase of the study. However, if a player finishes up a level before the timer runs out, he is automatically directed to the next phase as well. The timer can be modified based on the requirement of the study.

At first glance, the task of jumping from one pole to another might look like it's resembling Fitt's law [30]. This is a widely used law in the field of Human-Computer Interaction and it predicts that the time to move to a target area depends on the distance to it, yet relates inversely to the target size. And so fast movements and small target areas result in greater error rates. In our Jumping Caveman game, players had the freedom of choosing their jump frequency and their target poles were constant in size even though the pole height and pole-to-pole distance might vary. As such, this game does not resemble a Fitt's task.

3.1.2 Re-spawn Methods

Due to miscalculation and inaccurate prediction, the caveman might fall in the intermediate gap between two poles and hit the ground. At each death, the score box resets to zero, and a dialogue appears on the screen showing the player their most recent and best overall scores for the current game round, with a restart button to re-spawn the caveman either at the beginning of the level or at the last pole that was successfully landed on. We have implemented two different versions of re-spawn methods to accommodate different game scenarios:

At the beginning of the level: The caveman is re-spawned at the first pole (see figure 3.2) no matter where the caveman died. This method is important if the objective of the researcher is to force the players go through the whole game experience for a specific period of time and find a relation between game performance (number of attempts, failure or success or overall expertise during a complete game session) and player experience. This method is also effective when researchers would like to investigate the learning curve of novice players.

At the last completed pole: In this case the caveman is resurrected in its last completed pole (see figure 3.3) for which it was awarded a point.

3.2 Game Features

We implemented several features in the game that can easily be manipulated by a researcher. Some of these features are obvious to the player (overt features) and some are hidden from them (covert features).

3.2.1 Level Difficulty

The level difficulty in our game easily comprehensible to the players since anyone can assume whether its harder or easier to perform jump prediction based on the visual interpretation of the poles. We have created our level difficulty by using a combination of 'Pole Height' and 'Pole Distance'.

Figure 3.2: Re-spawn method at the beginning of the level (sequence of events in clockwise direction starting from upper-left

Pole Height determines the height of a pole from the ground. Based on the size of the sprite[2] that we used in our system, a block is 10 Unity units. That means a pole with 10 Unity units is completely above the ground and 0 unity unit refers to a pole levelled to the ground.

Pole Distance determines the spacing of one pole to the next. To keep similar configuration, we set 10 Unity units to be the maximum possible pole distance. That means a distance of 0 Unity units will place the poles in the exact same position overlapping each other.

Together, the pole height and distance can be used to create higher or lower variability,

[2] A 2D computer graphic that can be modified as a single entity or block, for more details - https://docs.unity3d.com/ScriptReference/Sprite.html

Figure 3.3: Re-spawn method at the last successive pole (sequence of events in clockwise direction starting from upper-left

which makes the game more or less difficult respectively. With further play-testing, we identified the maximum possible distance that can be travelled by the caveman is about 10 Unity units in the X-axis. And so, we have determined the distance would be about 5 Unity units on an average. So in summary, the minimum and maximum pole height and distances in our system is 1 Unity unit and 10 Unity units respectively. The pole height and distance are shown in figure 3.4 and figure 3.5.

Based on all the calculations and pilot tests that had been performed with several participants in multiple sessions, we came to the conclusion that the base level of variability in 'Pole Height' and 'Pole Distance' should be 5 Unity units - which means each pole with 5 Unity units height and positioned at 5 Unity units from each other would give us the average game difficulty. At figure 3.6, we demonstrate how a game level would look like with this type of 5 Unity units of level difficulty (we have not used this particular level difficulty at). Based on this base level difficulty calculations, we made three different type of level difficulty as follows:

High Level Difficulty: In the high level difficulty round, the maximum and minimum 'Pole Height' and 'Pole Distance' are 8.75 and 1.25 Unity units respectively. This means the 'Pole Height' and 'Pole Distance' can be anywhere from 8.75 to 1.25 Unity units which is 75% of variation than the base 5 Unity units (see figure 3.7). The more the variation is in the consecutive poles, the harder it becomes for the player to predict the caveman power before initiating a jump.

Medium Level Difficulty: In this level, the variation of pole is about 50% of the base 5 Unity units (see figure 3.7). This means the maximum and minimum values are now 7.5 to 2.5 Unity units.

Low Level Difficulty: This type of level difficulty is the lowest one that we have used in our system. The variation is now set to 25% of the base 5 Unity units - the maximum and minimum 'Pole Height' and 'Pole Distance' are 6.25 and 3.75 respectively (see figure 3.7).

All the distance calculations performed are entirely dependent on our system, the size of sprite that we had, and the results that we have obtained from play-testing. It can be further modified by any other researcher if he plans to alter the game avatars or any other game objects.

3.2.2 Covert Assistance

Jump Assistance/Hindrance is the main feature in our system. We provide assistance (see figure 3.8) by manipulating the caveman's trajectory to land closer to the pole nearest the destination of the unassisted trajectory (the destination pole), and hindrance (negative assistance) (see figure 3.8) by manipulating the caveman's trajectory away from the destination pole. When the player input is received, we take the input velocity and calculate the velocity to land on the destination ideal pole. We subtract these two numbers to get ΔV, which represents how far the player is from a successful landing. We calculate ΔV in the same manner for our assistance and hindrance condition. The complete procedure for performing

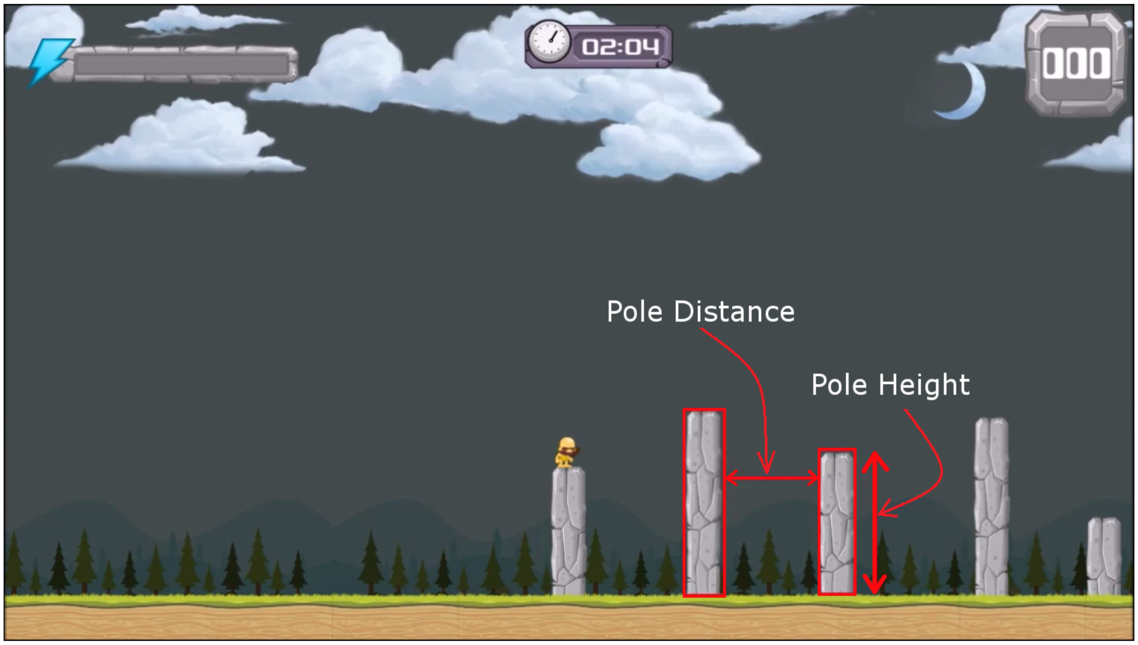


Figure 3.4: Level Difficulty Demonstration. Pole Distance – Distance from one pole to another, Pole Height – Height of a specific pole.

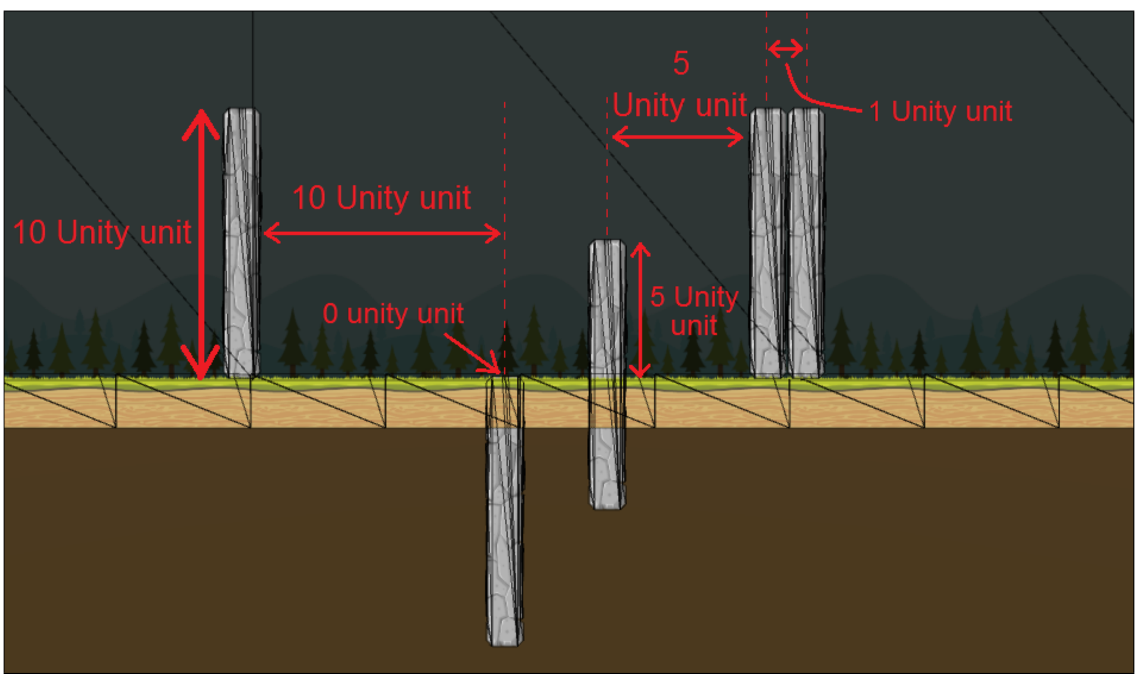


Figure 3.5: Maximum and Minimum Pole height and Pole Distance Demonstration in Unity unity in respect to our implemented system (Shaded Wireframe view).

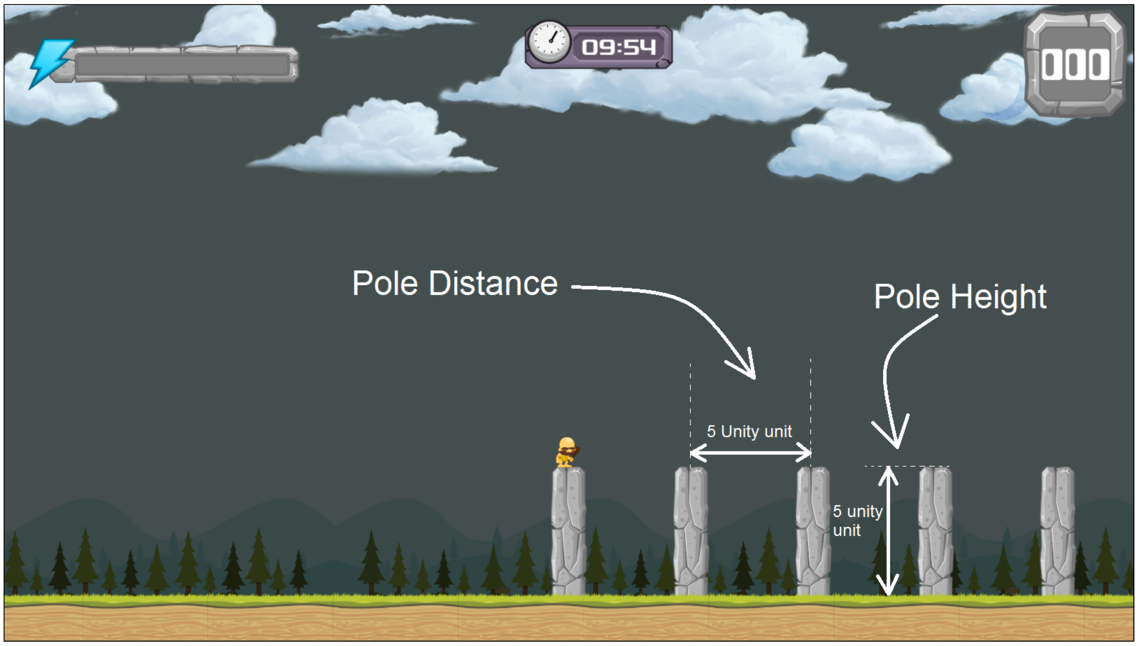


Figure 3.6: Average Level Difficulty with Pole Height and Pole Distances set to 5 unity unit.

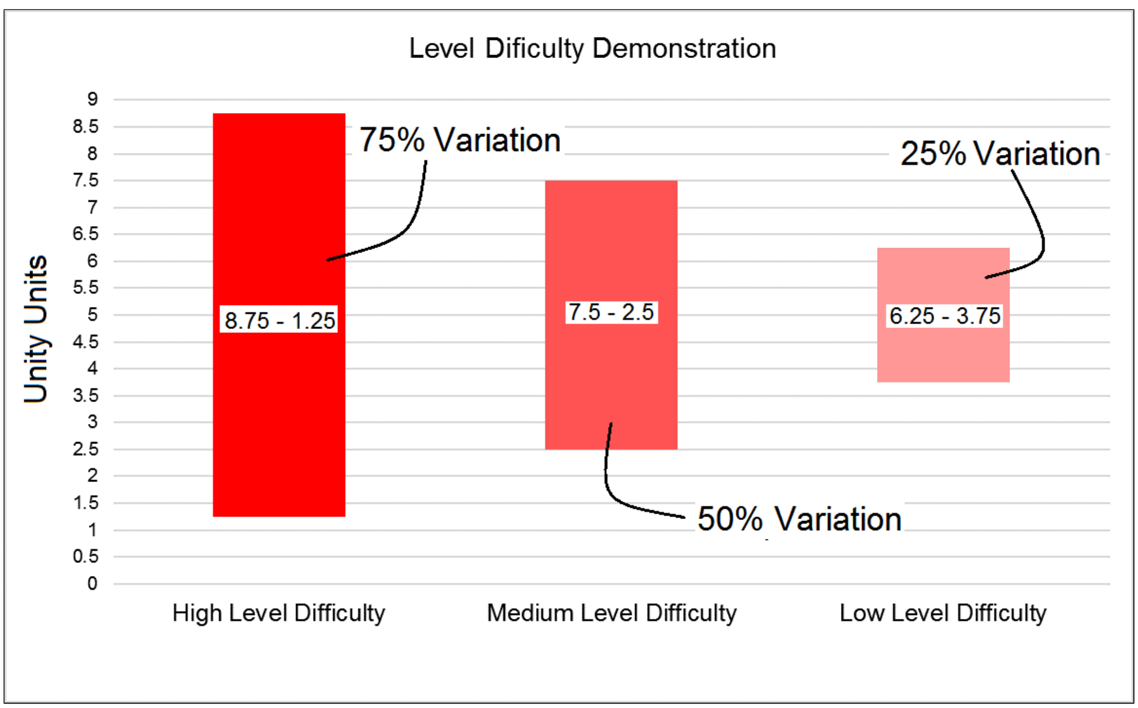


Figure 3.7: A demonstration chart of maximum and minimum 'Pole Height' and 'Pole Distance'.

the trajectory manipulation is described below:

Step 1: First, we calculated the range of the projectile based on player input using the equation 3.1

$$d = \frac{v^2 sin2\theta}{g} \tag{3.1}$$

Here,

v = Initial velocity of the player,

θ = Angle at which the player was thrown into the air

g = Default physics gravity in Unity (= $9.8ms^2$)

Step 2: Next, we identified the nearest pole position based on the range of projectile (d), by comparing it with all the pole positions in the associated configuration file.

Step 3: And then, we calculated the required velocity (v_0), to reach on top of the intended pole based on the equation 3.2

$$v_0 = \frac{1}{cos\theta}\sqrt{\frac{\frac{1}{2}gd^2}{dtan\theta + y_0}} \tag{3.2}$$

Here,

d = Range of the projectile, calculated with equation 3.1,

θ = Angle at which the player was thrown into the air,

g = Default physics gravity in Unity (= $9.8ms^2$),

y_0 = Launch height (difference of heights between player's current pole and its intended pole)

Step 4: Finally, we calculated the difference of a player's initial velocity (v) and required velocity (v_0), to find out ΔV and apply assistance/hindrance based on that.

Jump Assistance

While implementing Jump Assistance, we introduced two specific ranges in which we decide whether we will be applying assistance or not. The range or threshold values are determined

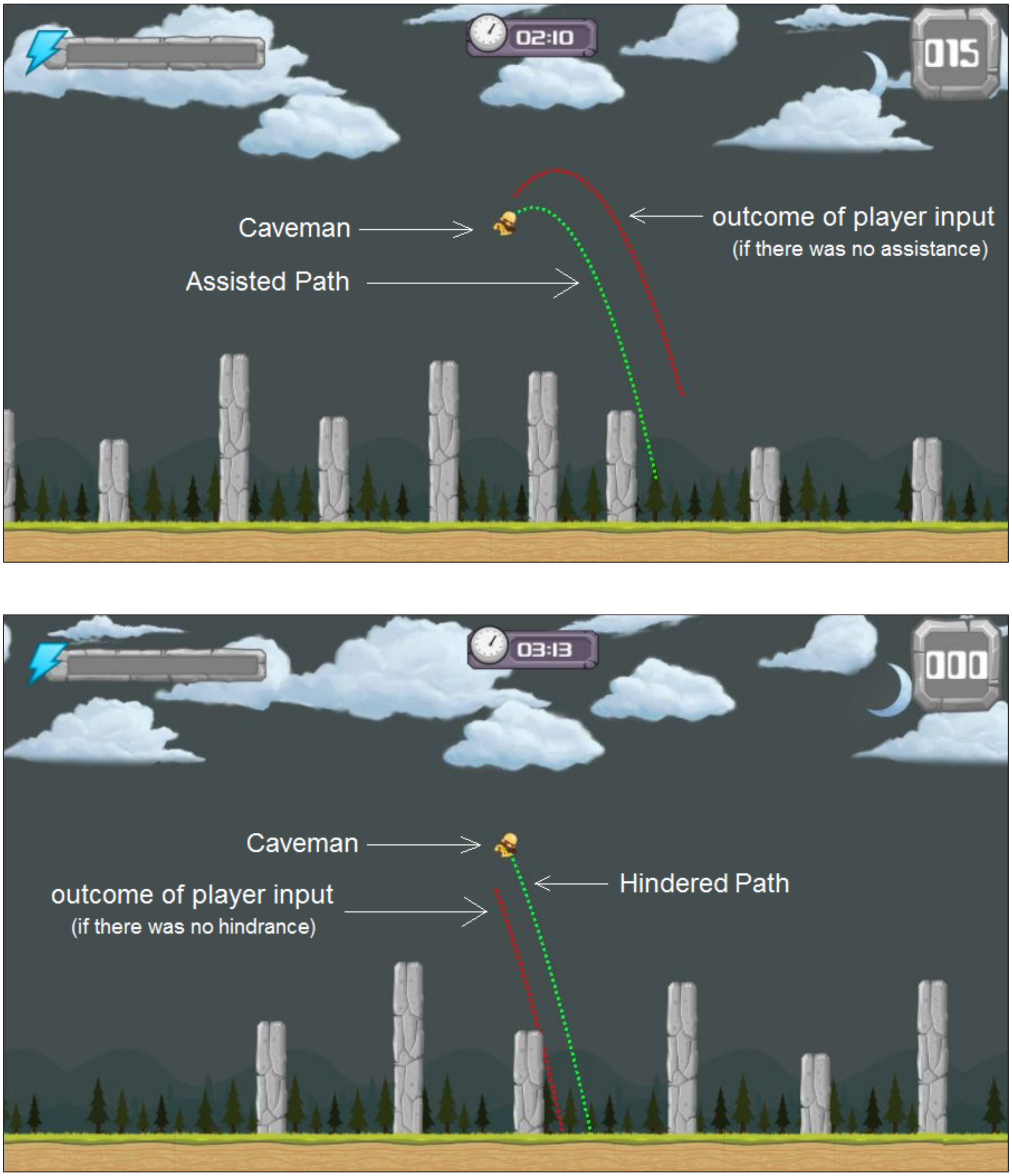


Figure 3.8: Jumping Caveman System: (Top) Assistance, (Bottom) Hindrance or Negative Assistance. Green line shows the assisted/hindered path and red line shows the outcome of only player input. In the real study none of the participants saw these trajectory paths – they are for demonstration and debugging purpose only.

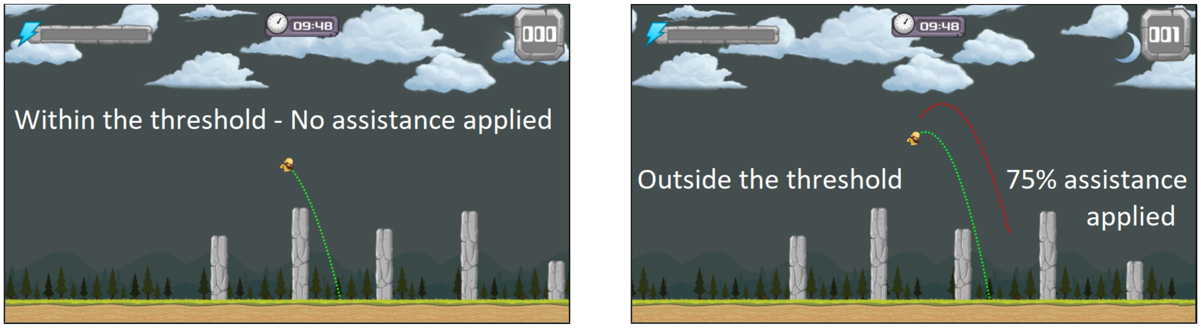


Figure 3.9: (Left image) - Visualization of Jump Assistance while the caveman is closer to the pole (within the threshold so no assistance is applied and no red trajectory line is shown) and (Right image) - assistance level is 75% where the red trajectory line shows the outcome of the player input only

by ΔV. If ΔV is below a very small threshold (0.25 Unity unit), no assistance is applied since the jump would already be successful. Applying any type of assistance would be redundant for this case. For all other cases, we adjust the player's input velocity by x% of ΔV, based on the degree of assistance set in the configuration file. For example, lets assume, in the configuration file the amount of assistance is set to 75% for a particular pole. This means we would adjust the velocity of the caveman either by increasing or decreasing 75% of current ΔV based on whether the caveman is close to the pole from left side or away from the pole on its right side (see figure 3.9 for a demonstration of 75% assistance).

Jump Hindrance

In the hindrance condition (see figure 3.8 – bottom), we used three difference thresholds, chosen through play-testing, and multiply the amount of correction by a factor, depending on the threshold. The largest threshold (1.00 Unity unit) is such that the ΔV is large enough for the caveman and it will miss the pole without assistance. Hence no hindrance is applied. This is intentional because we want to avoid any type of noticeability about our system manipulation since this might result in biased player experience [27]. In fact providing hindrance in such case might take the caveman closer to an unintended pole. The medium and small thresholds are where we hinder the caveman's velocity ($0.25 \sim 0.15$ and ≤ 0.15 Unity unit, respectively). To calculate how much the caveman's velocity is altered, we take the ΔV and adjust the caveman's velocity by an extra x% (as defined in the configuration file), in the direction away from the pole. To account for the variability of ΔV, the corrected velocity is multiplied by a factor, F, depending on which threshold it is in. In this way, if the player is very close to the pole, a larger F will be applied, causing the player to miss the pole. Within the medium threshold F=3, and within the small threshold F=6. The values for the thresholds and factors were chosen and tuned based on user feedback from play-testing.

3.2.3 Other Features

We also implemented several other features that could be used to manipulate the game level. Most of these have not been explored in the context of this thesis except pole magnetism. In the following sub-section we will be explaining them on a basis of priority:

Pole Magnetism

In an early implementation stage, we used pole magnetism as one of our primary features that was modifiable to use it as assistance or hindrance. In this mechanism, the pole surface acted like a magnet that can attract or repel the caveman while it's inside a certain range of distance and thus the assistance or hindrance mechanism worked. The properties of this feature is explained below:

Range of Magnetism/Magnetism Area - The range is calculated by a circular area around the surface of the pole. The centre of this circular area is at an offset to the centre of pole surface. The offset is measured based on the distance from the approximate centre of the caveman and the centre of the pole surface. We created three different ranges for this 'Magnetism Area'. The values are: 2.5, 5.0 and 7.5 Unity units respectively and they are represented as H (High), M (Medium), L (Low) in the config file (see figure 3.12). A visualization of 5 Unity units of 'Magnetism Range' can be viewed in figure 3.10.

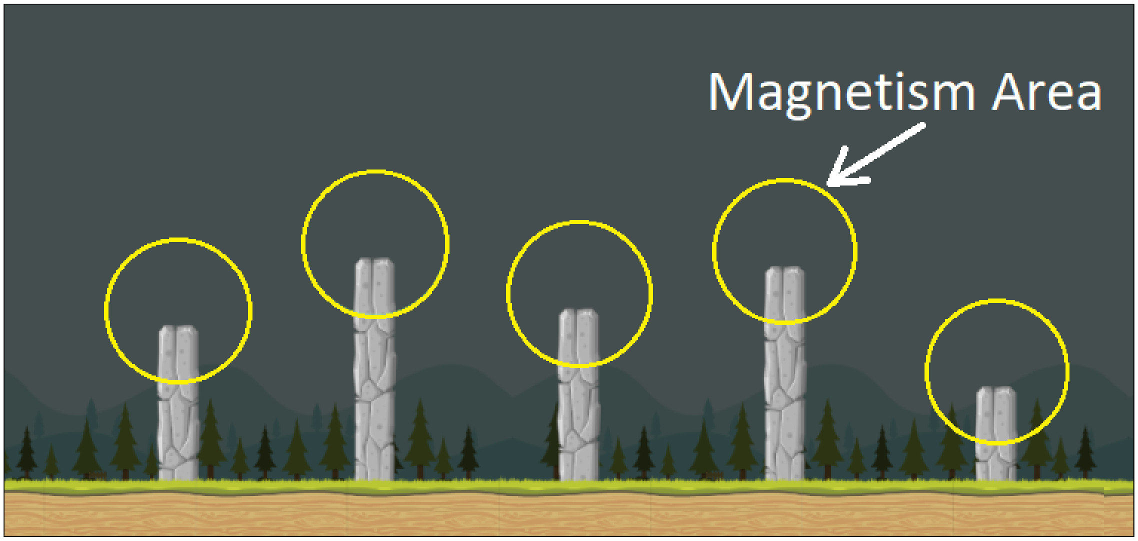


Figure 3.10: Circles around the poles indicates the Magnetism Area for 5 Unity units

Magnetism Level - The magnetism level is the amount of magnetism applied to our caveman. Since, the magnetized caveman's speed changes while it gets inside any pole's 'Magnetism Area', we introduced an air drag to visibly reduce the change in speed. This would help to reduce noticeability since experiencing an unusual trajectory path at certain locations (near the pole surface where the 'Magnetism Area' is defined) might make the player confused. The magnetism value is set to '30 Unity' units in our tool.

Other minor features

Trajectory Paths - Based on the player's input, we calculate the caveman's velocity and identified the trajectory paths that were demonstrated in figure 3.8. The trajectory paths that are shown are only available for debugging purpose. We showed the trajectory paths in order to demonstrate our scenario effectively. In an actual experiment session, we always turned off this visualization. However, with a small change in the inspector of unity, they can be activated. This can lead to answering different research questions with the help of our tool - like how players react while they realize that their performance is not completely dependent on their own input.

Altering the friction of the pole surface(Friction Value) - this will change the way the caveman behaves after landing on the pole. A high velocity jump might cause falling to the ground if the pole friction level is too low. The 'friction' value can be set from 0 to 1, while 0 is completely slippery and 1 refers to the highest level of 'friction'. In the current context of our thesis, we have set 'friction' level to 1 for all poles to remove it as a factor of interest.

Type of Pole/Visual material of the pole - We have introduced total five different types of poles which have distinctive category of friction for each of them. They have been titled based on their appearance: 01. Rock Pole, 02. Brick Pole, 03. Brick Pole with Grass, 04. Pipe and Ice Pole, 05. Ice Pole. Although each of these poles has a default friction value, it can be overridden with the configuration file's friction column. They are represented as numerical value 1-5 in the configuration file (see figure 3.12). Figure 3.11 shows the five different poles that we have implemented in our tool.

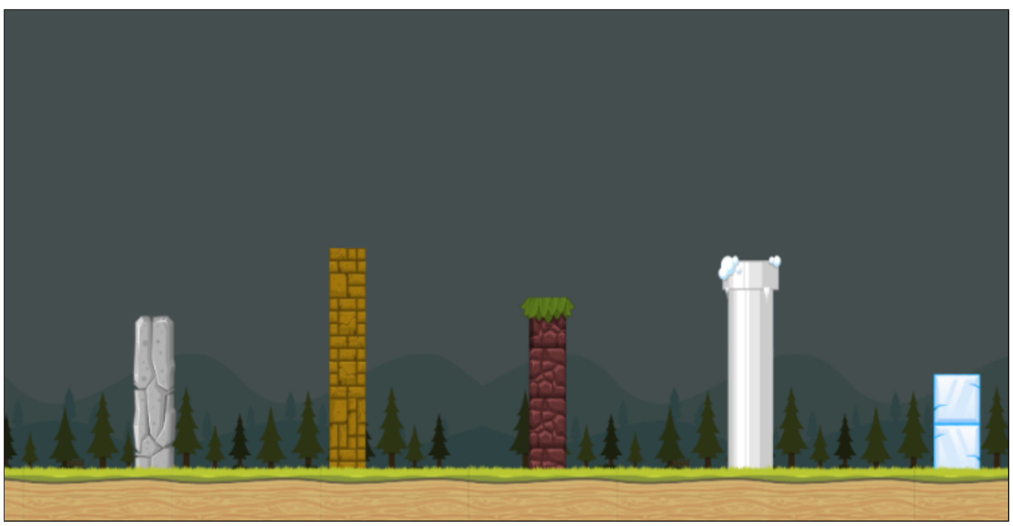


Figure 3.11: 5 different type of poles in our Jumping Caveman System

Since changing the graphical appearance of these poles was not the objective of any of our studies, we have used rock pole in general, and in this whole thesis when we use the term pole, we refer to the rock pole.

3.2.4 Configuration File

Although procedural generation of the game rounds is possible, we instead wanted to implement a method by which researchers could prescribe the game rounds to ensure similar experiences across participants. As such, we created a specific input configuration file that interfaces with the game system to define the game round parameters and set up game experiments in a controlled manner. The configuration file is used to define each pole and set its parameters, including height, positioning, and the degree of assistance/hindrance. The use of the configuration file, which is simply a standard comma-separated values text file, allows non-programmers to design game rounds and study their research question of interest.

Pole X	Pole Y	Magnetism Level	Type of Pole	Friction Value	Magnetism Area	Assistance Level
5.26	4.86	5	2	0.8	L	+ 75
10.98	3.75	7	5	1	H	+ 85
15.45	2.10	10	4	0.5	M	- 85
18.75	6.46	15	3	0.4	L	0
23.36	8.54	4	1	0.9	M	- 75
30.48	3.60	8	6	1	L	+80

Figure 3.12: Sample Configuration File (Note: This is just a demonstration. Not an actual screen-shot of our configuration file.)

The configuration file that we have created, simply encompasses the features (major and minor) that has been explained previously. A sample configuration file has the following columns sequentially (see figure 3.12):

- X position of Pole, Measures: Unity unit (e.g. 54.75 sets the corresponding pole at 54.75 in the X-axis)
- Y position of Pole, Measures: Unity unit (0-10) (e.g. 5.86 sets the corresponding pole at 5.86 in the Y-axis).
- Magnetism Level, Measures: Unity unit (0-30); higher magnetism value represents higher attraction.
- Type of Pole, Measures: Numerical value (1-5)

- Friction Value, Measures: Unity unit (0-1)
- Magnetism Area, Measures: single character (negative value for attraction and positive for repelling)
- Assistance Level, Measures: Numerical Value (range from -100 to +100)

Chapter 4

General Overview of the Studies

We conducted a series of studies to identify if we could manipulate player experience using the covert assistance technique of our tool. In our study 1, we used 'pole magnetism' (described in 3.2.3) as our initial covert assistance technique. However, since this assistance technique did not work according to the intended manner and failed to remain concealed, we discarded it from any of our future studies.

Later on, while designing our study 2, we developed 'Jump Assistance/Hindrance' that manipulates the caveman's original trajectory. This technique helped the caveman get closer to the intended pole or interrupted the direction of caveman jump. In this study, we controlled the hidden assistance to establish that our system works efficiently at manipulating experience in the hypothesized and intended way. From the very beginning of our study deployment, our expectation was that the covert assistance techniques, would alter the player's in-game performance. A secondary goal was to verify whether the players noticed the hidden assistance, since the pole variation was consistent throughout the whole game round for all conditions. In study 3, we explored whether these results hold regardless of the level difficulty, demonstrating that the strength of the assistance, and thereby the sense of success and failure can be scaled.

In study 3, we controlled the sequence of covert assistance – we created two levels with identical pole difficulty and assistance, varying only the placement of the assisted and hindered poles within the game round to demonstrate the resolution of the system.

In this chapter, we discuss the procedures of our experiments that were common to all of our studies. The first section gives an overview of our online deployment platform. Following this, the next sections go through all components of a complete study session from participant consent forms to two pre–game questionnaires, instructions to play our game, a short training round to get the players familiarized with the game environment and controls, a game session consisting of several game rounds (relevant to the research question), a follow-up post–game questionnaire after each game round, and a final debrief explaining the experiment purpose and the assistance mechanism.

4.1 Participant Recruitment Platform

We used Amazon Mechanical Turk (MTurk) to recruit our participants and navigate them to our online deployment. MTurk is an established and reliable research tool – also referred to as an online labour market [72] – for conducting behavioural research. This platform works as a mediator between parties who offers a variety of requests (known as 'Requesters') and qualified paid workers. The requesters post different tasks on MTurk, known as Human Intelligence Tasks (HITs) that cannot be easily solved by a web-based robot and requires active human participation (e.g., find the best photograph from the next set of photos that describe the front gate of a particular store located at your neighbourhood and answer the following surveys based on this task). We preferred MTurk over a controlled lab environment for the following reasons:

- One of the biggest advantages of using M-Turk is the statistical power that we would be able to achieve by reaching hundreds or thousands of diverse subjects with varying ages, task preferences, game expertise levels, and distinct personality traits. Access to a diverse range of participants helps with the ecological validity of our studies. Performing a study in a supervised lab environment might not provide us with the versatility to recruit such a group of participants due to geographical limitations. To manage and go through one complete experiment in lab settings might also take days or even months while MTurk allows is to complete such demanding tasks in only several hours.

- Less time investment in data collection means there is a possibility of more experimental iterations.

- MTurk gives researchers sufficient options to choose a subject pool according to the research requirements. MTurk lets the requesters create custom filters with age, location, HIT acceptance rate, and other criteria.

- Before accepting any HIT, requesters can also validate the tasks performed by a particular MTurk worker. This ensures high quality dataset.

Since we were working with human subjects, we obtained the ethical approval from University of Saskatchewan behavioural research ethics board and confirmed that our study strictly follows all the standard ethical guidelines that ensures privacy of any personal data. Additionally, we only invited participants located in the US who were 18 or older. For quality control, we opened our HIT to participants with a prior approval rate of 90% or more. Furthermore, we rechecked for data loss prevention and added qualifications for the group of participants who had completed the current study. The final qualification check ensures no data acquisition from same participant in our study series more than once.

4.2 Participation Consent

At the beginning of the study session, before navigating to the experiment, we always received a participant's informed consent [1], in which we presented details about our experiment procedure, time to complete the full study session, and other relevant details about our study. However, we did not receive any digital signature. Instead, the participants were required to agree to the consent form and move forward by clicking a button. The consent form had the following parts:

1. Title of the study,
2. Name and contact information of the researchers,
3. Complete study procedure description in brief,
4. Study Fund,
5. Information on the confidentiality of the study,

[1] The consent forms can be found in appendix A

Only game control is 'Left Mouse Button'

When the game starts, you will see a caveman standing on a rock pole.Your objective is to help him jump to the next pole. Holding down the left mouse button will charge his jump (the longer you hold the button down, the farther he will jump). You can see the charge in the power bar on the top left of the game window. If he falls, you will restart the level.

Figure 4.1: Instructions to play the game

6. Right to withdraw,

7. Follow-up information and questions or concerns.

4.3 Pre–Game Session Questionnaire

The pre–game questionnaire was divided into two two parts - the first part asked the participants about their demographics information (age, sex, education level, employment and marital status) and the second part gathered data about their previous gaming expertise (frequency of game-play, self-identification of gaming expertise, dominant hand, interested game genres, and devices they used to play with). Both of these questionnaires are shown in appendix B.1 and B.2.

4.4 Training Round

Following the pre–game questionnaire, participants were sent to the instructions page. The objective of this page was to give an overview of the type of game they were going to play, game controls, and the scoring method (see figure 4.1).

Before presenting our actual game session to the participants, we introduced a short practice session of one minute. This round had the following dissimilarities from the actual game session:

- First, it did not contain the similar configuration of level difficulty. It was relatively easier since the variance in pole height and pole distances were low. We used a separate configuration file for the training round. This file is different from the experimental conditions since participants were playing our game for the first time and some of them might not be experienced with game-playing at all.

- Second, we always used an average level of assistance in this training round. The average level was determined based on the degree of assistance we used for the respective study.

- And finally, although we logged data in this training round; none of these data were analyzed in the context of this thesis.

4.5 Post-Game Session Questionnaire

Subsequent to the training round, the participants were directed to our game-play conditions. A ‘Post–Game Session Questionnaire’ was presented after each round to reflect upon their gaming experience. We evaluated our manipulations by logging in–game behaviour as well as subjective measurements using questionnaires.

4.5.1 Player Experience(pX)

We used a 17-item questionnaire to investigate pX. The survey questions consists of the enjoyment and effort sub–scales from IMI scale [73, 96], competence from PENS [93, 88, 101], the internality construct from the game-specific attribution questionnaire (GSAQ) [26], and a one-item response on perceived difficulty. Each of these constructs were rated using a 7-point likert scale. A brief overview of the questions are given below while the complete question set is listed on appendix B.3.

Interest-Enjoyment and Effort

These sub–scales included questions such as: Enjoyment - ‘I enjoyed this game round very much’, Effort - ‘I put a lot of effort into this game round’.

Competence

Competence indicates a sense of mastery over challenges, e.g., My ability to play this game round is well matched with the game's challenges. We measured competence as it is a main predictor of game enjoyment [93] and also is the factor by which covert assistance manipulates player experience [27, 109].

Internality

This sub-scale measures how internally (vs. externally) players attribute their performance (e.g., 'How well I did in this game was completely due to me.'). We measured internality to identify whether the players attribute the cause of their performance to themselves or to the system as an important measure for the use of covert assistance manipulations. If internality dropped drastically, we could assume that players felt that their performance was due to the system.

Difficulty

This measurement was fairly straightforward, using a one-item question to reflect upon the difficulty - 'This game round was difficult'. We included this question item for the following reasons:

- To identify whether there is any interaction with the level difficulty and the covert assistance techniques
- To measure how well the player's ability or expertise was matched with the game's difficulty
- As a manipulation check for the implemented assistance/hindrance.

4.5.2 Game Performance Measures

While playing the game rounds, we logged several metrics related to the game and the player's performance for the sake of evaluation and data analyses. A complete list of the game logs are shown below: (see Table 4.1).

Table 4.1: Game Performance Measures (Left Column - Parameter Name, Right Column - Description of the log).

Parameter Name	**Description of the metrics**
roundNum	Which game round the participant is playing
poleNumber	At which pole the caveman is standing
poleDistance	The distance between current pole position and previous pole position of the caveman jump
poleHeight	Current pole height
distanceToTheNearestPole	After calculating the caveman's jump trajectory determine where it would land and distance to the nearest pole
magnetsimLevel	Level of 'Pole Magnetism' (while the feature is activated) – otherwise zero ('0')
magnetismArea	The area where 'Pole Magnetism' gets applied and the caveman gets attracted or repelled
trajectoryAssistance	The level of 'Jump Assistance'
trajectoryHindrance	The level of 'Jump Hindrance'
friction	Level of 'Pole Friction' (minor feature) – default 1 for the thesis context
power	The percentage of power bar that was filled up during the caveman jump
score	Current score of the game
success	Was the caveman jump successful?
failure	Did the caveman jump failed
successWithoutAnyAssist	Would the jump have been successful if there was no assistance?

However for the context of the studies in this thesis, we derived the following metrics for each round as performance indicators:

- **Number of Deaths**: calculated based on the total number of 'failure'(s) per 'roundNum'
- **Maximum score**: number of poles successfully jumped before death - calculated using the 'score' column.

The complete list of game performance logs gives an overview of the strength of our tool that can be leveraged for many other research ideas.

4.6 Debriefing Session

Since performance manipulation was different for each study, we will be presenting the debriefing texts in the following chapters 5 to 8. A debriefing is an essential part of the consent process for the following reasons:

- Debriefing is essential when an experiment involves any sort of deception to human participants. Our study involved trajectory manipulation for at least half of the conditions and players experienced a game-play that did not completely reflect their own game performance.

- The debriefing procedure gave full insight on how the hypotheses were tested for our research and the techniques used for the trajectory manipulation.

- The debriefing process also provided participants with instructions on how they could obtain further information about the full experiment procedure to clarify any scenario in respect to participation and collected data.

Before concluding the experiment, we asked the participants to give feedback about the purpose of the conducted experiment in their own words in a free-form text box. Following this feedback session, we performed a complete debrief about the purpose of our experiment, the type of manipulation, and the number of game rounds affected by this manipulation. Since the participants' game experience slight differed than it would have without performance manipulation, it was essential for us to debrief them with information on our procedures. For most of the studies half of the total game rounds were manipulated and they perceive a deviation in their game performance.

In addition to this, to ensure whether the participants were attentive in the debriefing session (since it could be skipped with a submit button), we presented a secondary survey page. We asked the following questions to answer either true or false: (a) My performance was completely dependent on my skill, (b) The trajectory of the caveman jump was manipulated.

4.7 Technical Overview and Further Testing

4.7.1 Brief Overview of the Online System

Each participant was presented a URL in the MTurk HIT that redirected them to our online server. The online system was built using the Flask app with the aid of Python scripts. Other functionalities included communication with a database - the programming interface was made using SQLAlchemy. This database was used to log all the game performance measures listed in 4.5.2.

4.7.2 Game Engine and Game Builds

The game was built using Unity[2] 5.0.1 initially and had been updated until the release of unity 5.6.0. We have used several assets from the Unity Asset Store[3] to create game backgrounds, several game objects (e.g., the caveman, clouds on the background, and all type of poles) and modeling the trajectory. Since we planned to use an online system to deploy our study, we exported our game in web player[4] initially for Study 1. However, building games and embedding into a web player got depreciated from unity and most of the modern browsers also stopped supporting NPAPI plugins[5]. As a result, we transferred our game builds to WebGL[6] as it is widely supported in almost all web browsers. For Study 1, before commencing the experiment, we had to confirm if the participants had installed web-player in their system. Changing the builds to WebGL also let us remove this extra precautionary step.

[2]A cross platform game engine, https://unity3d.com/

[3]https://www.assetstore.unity3d.com/en/

[4]https://unity3d.com/webplayer

[5]Netscape Plugin Application Programming Interface is an application programming interface(API) for web browsers; web player is one these NPAPI plugins

[6]https://developer.mozilla.org/en-US/docs/Web/API/WebGL_API

4.7.3 Performance Testing

Participants of our experiment used mostly the Chrome and Firefox browsers to view the session and play our game. However, due to different screen size of the monitors they could be using, there was a chance of unusual cut-off of the game screen. To minimize this issue, we exported our game in to a resolution of 900 x 500 to fit most screens and performed preliminary testing on most of the supported screen resolutions over different web browsers. The game data loaded on to the participant machine initially, while the relevant web-page loaded with the game into a data-frame. We tested our game at low-bandwidth (512 Kbps and 1 Mbps) to ensure uninterrupted experiment session.

Chapter 5

Study 1: Manipulating Player Experience with Pole Magnetism

The primary objective of this study was to explore the effectiveness of our system in terms of manipulating player experience and figure out how level difficulty would interact with the pole magnetism assistance technique. We also wanted to create an easily-configurable system to make sure researchers could change it according to their needs. The pole magnetism feature was established keeping in mind that it could be modified either to attract the caveman towards the pole (acting as assistance) or repel the caveman away from it (as a means to hindrance). A brief demonstration regarding how this pole magnetism works can also be found in 3.2.3.

5.1 Experimental Conditions

5.1.1 Game Rounds with Level Difficulty and Pole Magnetism

We used two types of level difficulty and two degrees of pole magnetism to build up the experimental conditions. Before explaining the four individual conditions, we will briefly explain the level difficulty and pole magnetism separately.

Easy Level

If we consider that difficulty is the only one factor that is controlling the game round, then this condition would actually mimic the 'Low Level Difficulty' round, explained in 3.2.1. It used the same 25% variation in pole height and distances. The avatar could easily jump over more than one pole because they were relatively close to each other. Since the poles are so close, this round was significantly easier than any other round even there was no assistance applied.

Hard Level

The pole height and inter-pole distance variations were set to a relatively wider range that imitated the 'High Level Difficulty' (see 3.2.1) that used about 75% of pole variation. It was considerably harder to jump over any single pole because they were placed relatively far away from each other and also varied quite a lot in their height. Based on the pole positions, this round was quite hard.

Two Variants of Pole Magnetism

We created two versions of pole magnetism, one with the assistance activated and another one with zero/no magnetism. The assistance technique was controlled by the 'Magnetism Area' (check 3.2.3 for reference) and the size of this area was kept same throughout the whole round. The magnitude of pole magnetism was set to 30 unity units (as stated in 3.2.3), selected based on several testing and debugging procedures:

1. This is the maximum possible value for the magnitude of this feature. Any value more than 30 unity units might have become noticeable to the players and they would have observed the manipulation,

2. While the caveman went through a particular 'Magnetism Area', he was dragged toward the respective pole surface due to the implemented magnetism on that particular pole. Based on its direction, it also might slightly bend toward the pole in a manner that does not resemble a parabolic motion. The magnitude of 30 unity units kept it at the minimum level while maintaining the magnetism level as powerful as possible.

Instead of using the level difficulty or pole magnetism factors individually, we crossed all of them and created four experimental conditions:

1. Magnetism - Easy,
2. Magnetism - Hard,
3. No Magnetism - Easy,
4. No Magnetism - Hard.

Participants re-spawned at the beginning of each level when they died (see 3.1.2 for details) and played each condition for 4 minutes.

5.1.2 Debriefing about the experimental conditions

As stated previously in 4.6, we always performed a debriefing session about our experiment. Although, the session came at the very end of the experiment, we describe it here:

Cover story: You were testing a game under development.

Explanation: We are interested in looking at the interplay between difficulty level and performance assistance. There were two difficulty levels - it is harder to play the game when the poles are spaced out more and vary more in height. We also had performance assistance. Sometimes the poles had 'magnetism' applied, which attracted the cavemen and made it easier to land on them.

Performance: Because we manipulated the trajectory of your avatar, your performance was not always completely dependent on your skill level. We helped you out a little with pole magnetism in half of the levels.

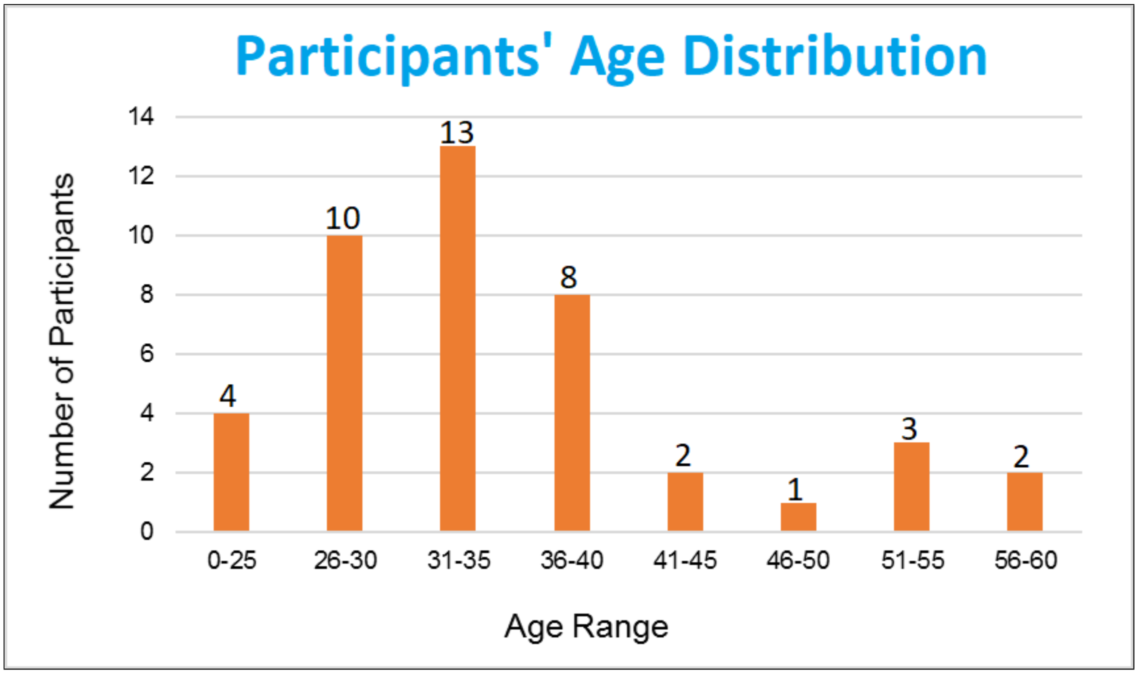
5.2 Procedure

At the very beginning of the experiment, participants were provided with a consent form, and they had to agree with it to advance into the experiment (see 4.2 for details). Next, we asked the participants about their demographics followed by another set of questionnaires asking them about their previous gaming experience. After this, they were presented with a training round of our 'Jumping Caveman' game with the pole magnetism set to 15 Unity units and a medium level difficulty (see 3.2.1 for details). We then presented our experimental conditions to the participants one after another followed by the post-game questionnaires subsequent to each round. This post-game session survey asked them about their enjoyment, effort, competence, internality, and difficulty of the relevant game round (as described in section 4.5). To avoid potential learning effects that could occur by presenting the rounds sequentially, we presented them using a balanced Latin Square design. In a Latin Square design, every start position is represented by recirculating the order of the conditions. In a fully counterbalanced design, the conditions are presented in all possible orders - which would require 24 possible conditions for our case. On the other hand, in a balanced Latin Square design, even though the number of combinations decreases, it protects and evens out the what-follows-what condition and protects against ordering effects, leaving us only four possible combinations. At the end of the experiment, participants were requested to provide any information they could imagine about the purpose of this experiment in a free text box form. Finally, before wrapping up everything, we debriefed them about the complete experimental procedure and how their performance was manipulated. The entire experiment took around 40 minutes to complete, and each of the participant was given $5 to compensate their time and effort.

5.3 Participants

All of the participants for this study were recruited through Amazon's Mechanical Turk (MTurk). In this online system, any active worker can accept the HIT if they are qualified to perform the task. Since we tracked all of our participants, a total of 63 participants initially got involved and completed our experiment either partially or entirely. However, if any participant wished to leave our system without completing the whole experiment or remained idle for 90 minutes, the HIT became invalid due to timeout, and some other participant became eligible to accept the HIT. Before performing any further analyses, it was indeed a necessity to carry out a filtering and cleansing procedure to remove any unwanted participants.

Initial Removal: 13 participants did not complete our experiment entirely and we only



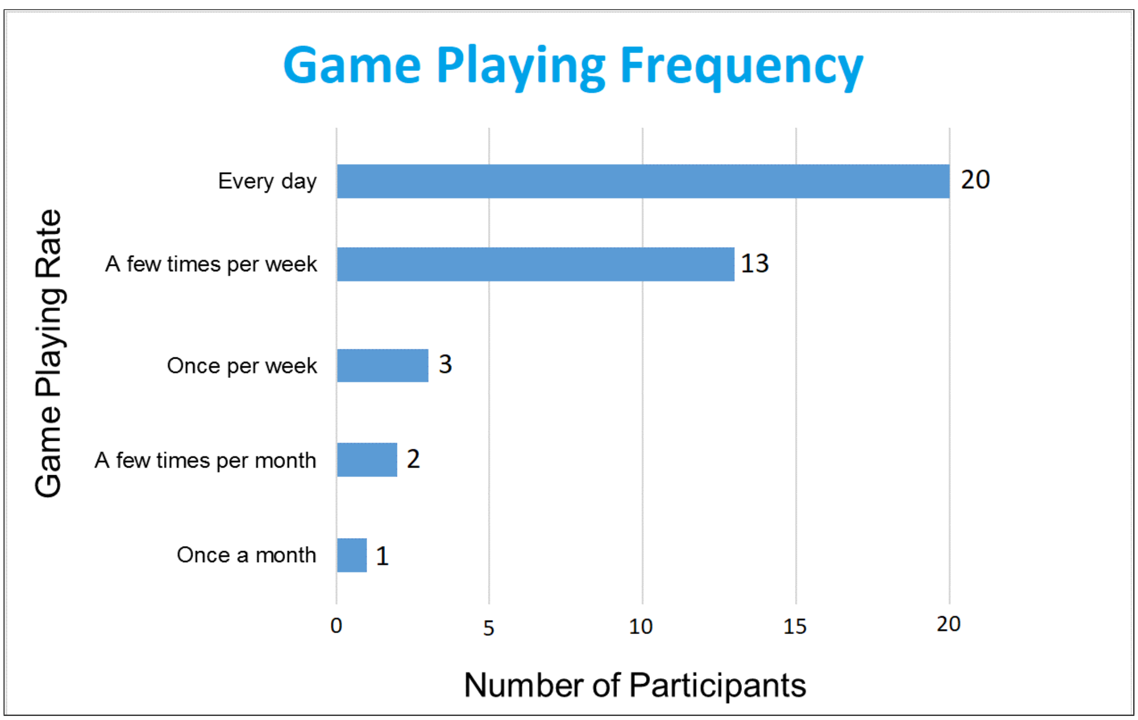
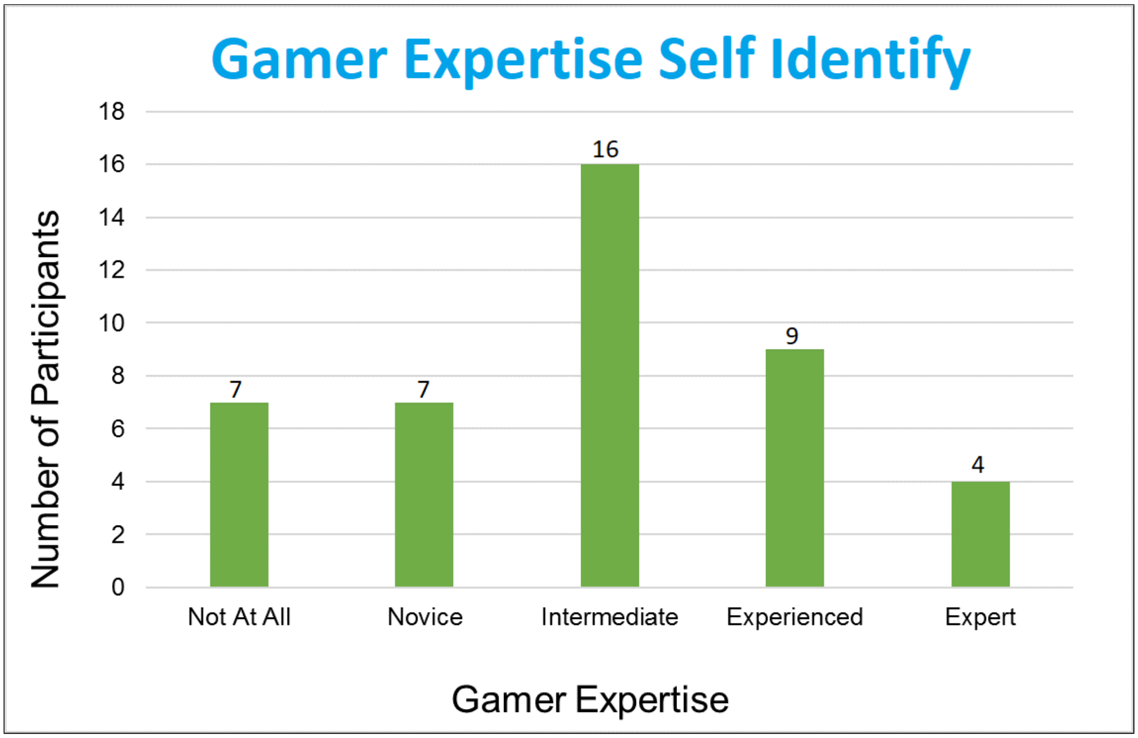


Figure 5.1: Brief Demographics (Part 1) overview of the Recruited Participants for Study 1



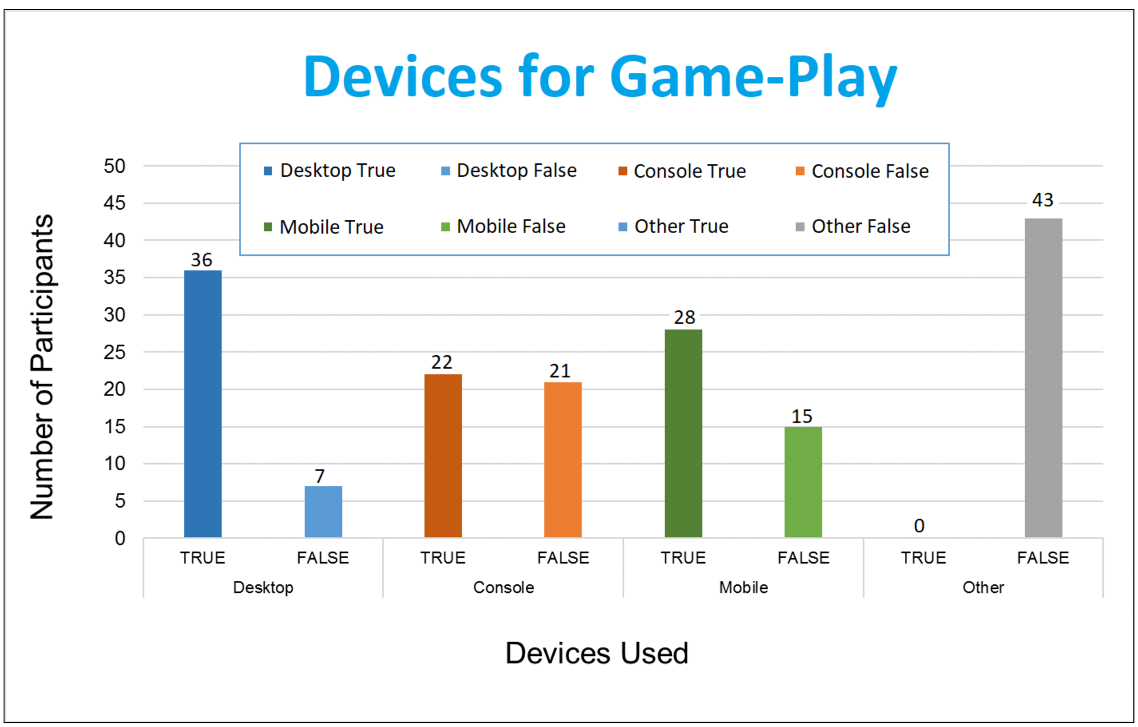


Figure 5.2: Brief Demographics (Part 2) overview of the Recruited Participants for Study 1

had a partial amount of data of these people, they were removed at the initial stage through MTurk. We were left with the remaining 50 participants (mean(m) = 34.38 years, Standard Deviation(SD) = 8.597 and 48% Female).

Participant Filtering: In the group of 50 participants, some tried to answer all/some part of the surveys randomly or very quickly. We identified these non-compliant data based on their response time and consistency metrics (as per suggestion of Meade and Craig [74]). We removed participants who spent less than 1.5 seconds/question on 2 or more of the combined pX questionnaires (n=2). Based on the variance of responses for each pX construct individually, we also excluded participants who had a variance of more than 3 SD above mean on any construct (n=5). Because of this removal of participants, we analyzed 43 of the participants (m = 35.14 years, SD = 8.983 and 51.2% Female). A brief overview of the participants is demonstrated in figure 5.1 and figure 5.2.

5.4 Hypotheses

Due to manipulating level difficulty with pole magnetism we expected difference in:
Difficulty: Each level of decrease in difficulty and increase in magnetism value will make the participants perceive a decrease in subjective game difficulty.
Competence: Players will perceive themselves as more competent in levels in which was assisted with pole magnetism and in levels with lower difficulty.
Enjoyment: Enjoyment will be increased with lower level difficulty and higher pole magnetism.
Effort: Player's perception of invested effort will be less while the magnitude of pole magnetism is higher and the level difficulty is lower.
Internality: Players will attribute their performance more internally when magnetism is provided. However, in levels with greater difficulty, the players would attribute more externally.
Performance: Players will die less frequently and obtain a higher maximum score with increased pole magnetism and decreased level difficulty.

5.5 Data Analyses

We have performed a repeated-measures MANOVA (performed in IBM SPSS[1]) with level difficulty (Easy and Hard) and pole magnetism (No Magnetism and Magnetism) as two within-subjects factors.

5.5.1 Dependent Measures

We had two types of dependent measures in our study:

Subjective

Perceived difficulty, enjoyment, effort, competence and internality - these measures were calculated based on the survey questionnaire that was presented after each round of game-play. Each of these feedback constructs were measured on a 7-point likert scale.

Objective

We also used game logs to calculate two objects dependent measures.
Number of Deaths: This is calculated based on total number of failed attempts for all poles in a particular round for each participant. We used the 'failure' column to calculate this value.
Maximum Score: This measure is calculated based on the highest score that one participant achieved in one round of game-play.

[1]https://www.ibm.com/analytics/data-science/predictive-analytics/spss-statistical-software

5.6 Results

5.6.1 Player Performance

Main Effect

There were main effects of level difficulty on max score ($F_{1,42} = 43.15$, $p<.001$) (see table 5.1), showing that players had higher maximum scores in the easy levels than hard levels. However we did not observe a similar effect of level difficulty on the death rate between the easy and hard conditions.

Furthermore, a main effect of pole magnetism showed that players had higher maximum scores ($F_{1,42} = 15.692$, $p<.001$) and died less frequently($F_{1,42} = 19.958$, $p<.001$) in the game rounds where pole magnetism assistance was applied.

Interaction Effect

There was significant interaction effect for Maximum Score ($F_{1,42} = 4.452$, p=.041) – demonstrating that players had higher maximum score in Easy Levels than Hard levels with pole magnetism activated, while compared to 'Easy-No Magnetism' and 'Hard-No Magnetism' rounds. A similar interaction effect was also found for the number of deaths ($F_{1,42} = 5.5634$, p=.022) showing that the death rate was higher in hard levels than easy levels while there was no magnetism, comparing it with the 'Easy-Magnetism' and 'Hard-Magnetism' rounds (see figure 5.3). Thus we can conclude that pole magnetism worked differently depending on the difficulty of the level.

Table 5.1: Repeated–measures MANOVA results: f–statistic, p–values and effect size for dependent measures for Study 1 (here, n.s = non significant)

	Level Difficulty			Covert Assistance			Interaction		
	$F_{1,42}$	p	η^2_p	$F_{1,42}$	p	η^2_p	$F_{1,42}$	p	η^2_p
Difficulty	12.935	**.001**	.235	4.744	**.035**	.101	.171	.682	.004
Enjoyment	.663	.420	.016	9.942	**.003**	.191	.230	.634	.005
Effort	.009	.926	n.s.	1.725	.196	.039	4.321	**.044**	.093
Competence	7.995	**.007**	.160	9.814	**.003**	.189	2.491	.122	.056
Internality	.333	.567	.008	.000	1.000	n.s.	.622	.435	.015
Number of Deaths	1.124	.295	.026	19.958	**<.001**	.322	5.634	**.022**	.118
Max Score	43.150	**<.001**	.507	15.692	**<.001**	.272	4.452	.041	.096

5.6.2 Player Experience

Main Effects

There were main effects of level difficulty on perceived difficulty ($F_{1,42}$ = 12.935, p<.001) and competence ($F_{1,42}$ = 7.995, p=.007) but not on enjoyment ($F_{1,42}$ = .663, p=.420), effort ($F_{1,42}$ = .009, p=.926) and internality ($F_{1,42}$ = .333, p=.567) (see table 5.1).

There were main effects of pole magnetism on perceived difficulty ($F_{1,42}$ = 4.744, p=.035), enjoyment ($F_{1,42}$ = 9.942, p=.003), and competence ($F_{1,42}$ = 9.814, p=.003), but not on effort ($F_{1,42}$ = 1.725, p=.196) and internality ($F_{1,42}$ = .000, p=.1.0) (see table 5.1).

Interaction Effects

There were no significant interactions between level difficulty and pole magnetism on perceived difficulty, enjoyment, competence or internality (see table 5.1). However the interaction effect on effort ($F_{1,42}$ = 4.321, p=.044) was significant. This showed that players invested more effort in the 'Hard - Magnetism' condition in comparison to 'Hard - No Magnetism' (see figure 5.4), but there was no effect of magnetism for the easy conditions.

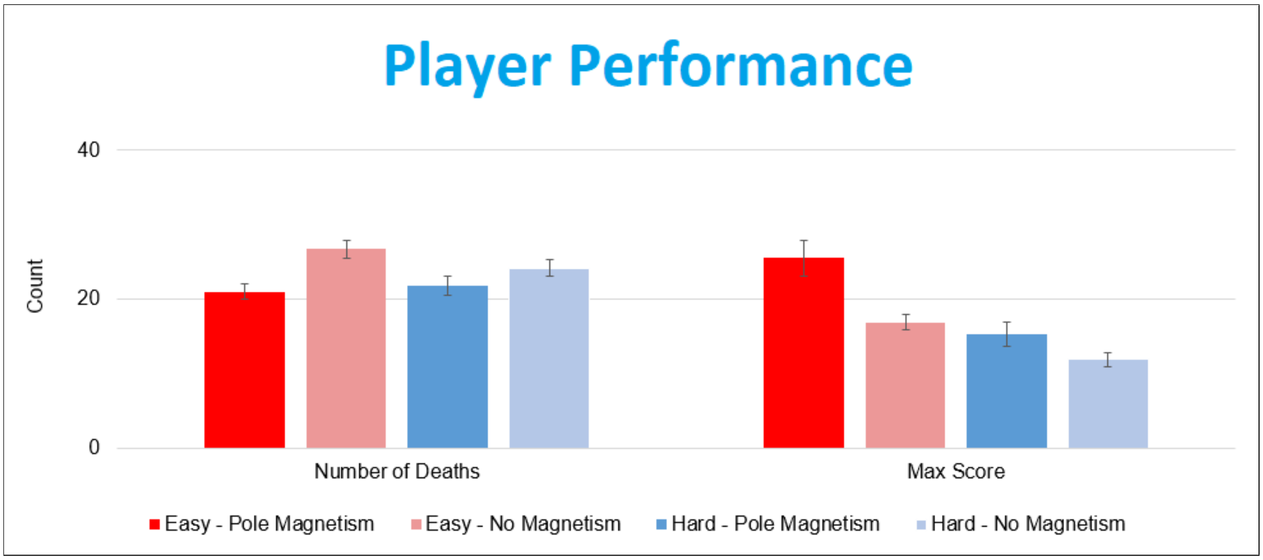


Figure 5.3: Mean ± Standard Error for representing Player Performance measures – Study 1.

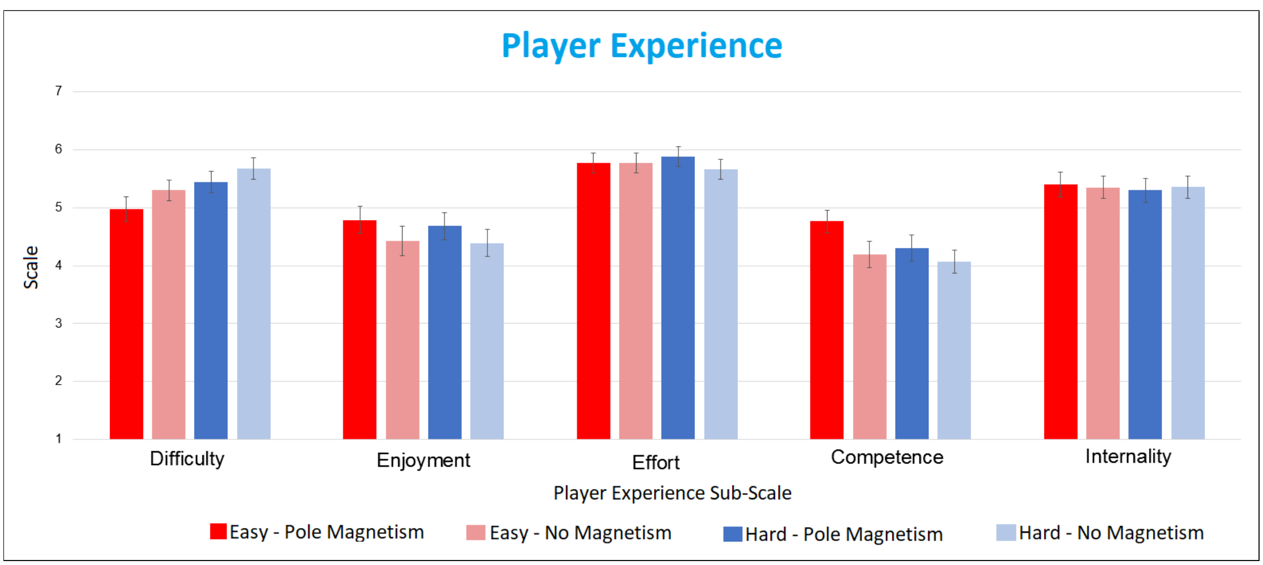


Figure 5.4: Mean ± Standard Error for representing Player Performance measures – Study 1.

5.7 Discussion of Study 1

The main effects of pole magnetism on most of the constructs (difficulty, enjoyment, competence, number of deaths and max score) proved that our implemented assistance mechanism worked quite well at affective experience. However the effect of perceived difficulty seems stronger in terms of overt manipulation (changing level difficulty) than covert manipulation (manipulating assistance) - this definitely suggests that these methods might not have been balanced properly.

If we closely look into the interaction effects, we also observe that only effort was marginally significant out of all the player experience measures. This showed that there was not much interaction going on between our assistance technique and level difficulty. This suggests that even though the main effects were not properly balanced, the effects of pole magnetism generally held when level difficulty was introduced.

However the main effect and interaction effect on internality for both level difficulty and pole magnetism is non-significant. This contradicts our expectations about the system since despite significant differences in player performance, the reflection on internality was similar. In fact, several participants informed that they were able to notice the effect of pole magnetism while the avatar was landing on the poles. They saw the avatar changing its direction while getting closer to the respected pole but did not rate the internality differently.

Based on the results obtained in this study, as a conclusion we can state that: even though some of the measured constructs were significant, the associated effect sizes were small. Even though our pole magnetism technique seem to be successfully manipulating some player experience, controlling the effect of this type of assistance is challenging. Based on these results and player feedback, we decided to implement a different type of assistance method – hence we introduced trajectory assistance (see 3.2.2 for details) that works solely on player input, and thus any alteration on player performance should remain hidden.

Chapter 6

Study 2: Manipulating Player Experience with Jump Assistance/Hindrance

Since the first study with pole magnetism was not effective enough to manipulate player experience while keeping the assistance hidden, we changed our covert assistance technique and developed jump assistance mechanism (described in 3.2.2). In contrast to our study with pole magnetism, we did not take level difficulty into consideration in order to implement a simpler study design and achieve significant manipulation results without players noticing the implemented assistance. The primary goal of this study was to identify if we could manipulate player experience by using our newest assistance method.

6.1 Experimental Conditions

Here, we created our experimental conditions based on different degrees of 'Jump Assistance/Hindrance': Positive Assistance (where trajectory manipulation helps the caveman on successful jump completion), Neutral State (neither assistance nor hindrance), Negative Assistance (pushing away the caveman from its intended pole). We maintained consistent level difficulty in all of these conditions in order to remove any potential interaction with the implemented assistance technique.

6.1.1 Three Types of Game Rounds

Level difficulty was kept at the medium level, regardless of condition, by using the same set of poles in all three conditions. We simply varied the starting pole so that participants would not experience each as the same game round, but would feel as if the rounds were unique.

High Assistance

We set the assistance as high as 80% for this condition. This means player's input velocity was adjusted by 80% of ΔV. (for reference this can be checked at 3.2.2 to know how the calculations were done with respect to ΔV).

Neutral

This is the most straightforward condition in which none of the player's performance was manipulated. In this condition, we did not perform any calculation of a player's input velocity, and we have not applied any assistance or hindrance. Hence, there was no manipulation of player trajectory. This also implied that, players' score will be entirely dependent on their performance (how accurately they could predict the jump).

High Hindrance/High Negative Assistance

We set the hindrance level to as high as 80% for this condition - meaning that a player's input velocity was hindered by 80% of ΔV. (for reference this can be checked at 3.2.2 to know how the calculations were done for the hindrance levels with respect to ΔV).

Participants re-spawned at the beginning of each level after they fall to the ground (re-spawning methods are described at 3.1.2) and played each condition for 4 minutes despite the number of times they died. The timer on the background was automatically paused until they hit the restart button on the score viewing screen, which appeared after each unsuccessful attempt (a screenshot of the game end screen can be found at the bottom right image of figure 3.2). There was no limitation on death rate or maximum possible score since the number of poles was too hard to complete within the time limit – this acted as an infinite game round from the perspective of the player.

6.1.2 Debriefing about the experimental conditions

The debriefing session was performed at the very end of the experiment as usual after receiving player feedback on our system. We used the debrief procedure to let our participants know about the objective of our research and what type of manipulations were performed:

Cover story: You were testing a game under development.

Explanation: We are interested in looking at the interplay between performance assistance and hindrance. There were three difficulty levels - it becomes easier to play when we provide assistance and harder while there is hindrance. We manipulated the jumps made by the caveman by adjusting their trajectory and initial velocity.

Performance: Because we manipulated the trajectory of your avatar, your performance was not always completely dependent on your skill level. We helped you out a little with trajectory assistance in one level and made it difficult to jump by providing hindrance in another one. The remaining level had no assistance or hindrance in it.

6.2 Procedure

After receiving informed consent, participants were asked to give their demographic information along with a survey questionnaire of their experience with games. After the pre–game session questionnaire, we presented a training round of our 'Jumping Caveman' game with a medium level of assistance (assistance level = 55%, which was set as +55% in the config file) and medium level difficulty (same as the original game rounds). Following this training round, participants played three consecutive game rounds - one for each experimental condition: High Assistance, Neutral and High Hindrance. To remove any learning effect, we fully counterbalanced the order in which conditions were presented. At the end of each game round, participants completed the post–game session questionnaire to reflect upon their current game round experience. At the end of the experiment session, we asked for their feedback about the purpose of the game session in a free text box similar to our study 1 and debriefed them. We also validated whether they had paid attention to our debriefing carefully by asking two short queries about it. The entire experiment took approximately 30 minutes to complete, and participants received a compensation of $5.

6.3 Participants

We recruited 54 participants (m = 36.13 years, SD = 11.31 and 38.9% Female) to participate in the study through MTurk. To exclude participants who did not take care in answering the survey questions, we identified noncompliance by response time and consistency metrics, as suggested by Meade & Craig [74]. We removed participants who spent less than 1.5 seconds per question on 2 or more of the combined pX questionnaires (n=1). We also detected the variance of responses for each pX construct individually and excluded participants who had a variance of more than 3 standard deviations above the mean on any construct (n=2) - this process excluded participants who consistently did not pay attention to reverse-coded items. This left 51 participants in our analyses (m = 35.98 years, SD = 10.67 and 40.8% Female, 86.3% right handed). In figure 6.1 and figure 6.2, we provide a brief overview of the demographics of our participants that were left after executing the filter.

6.4 Hypotheses

We expected that manipulating assistance and hindrance to create three levels would yield effects on:

Difficulty: Players will perceive a decrease in difficulty with increased assistance.

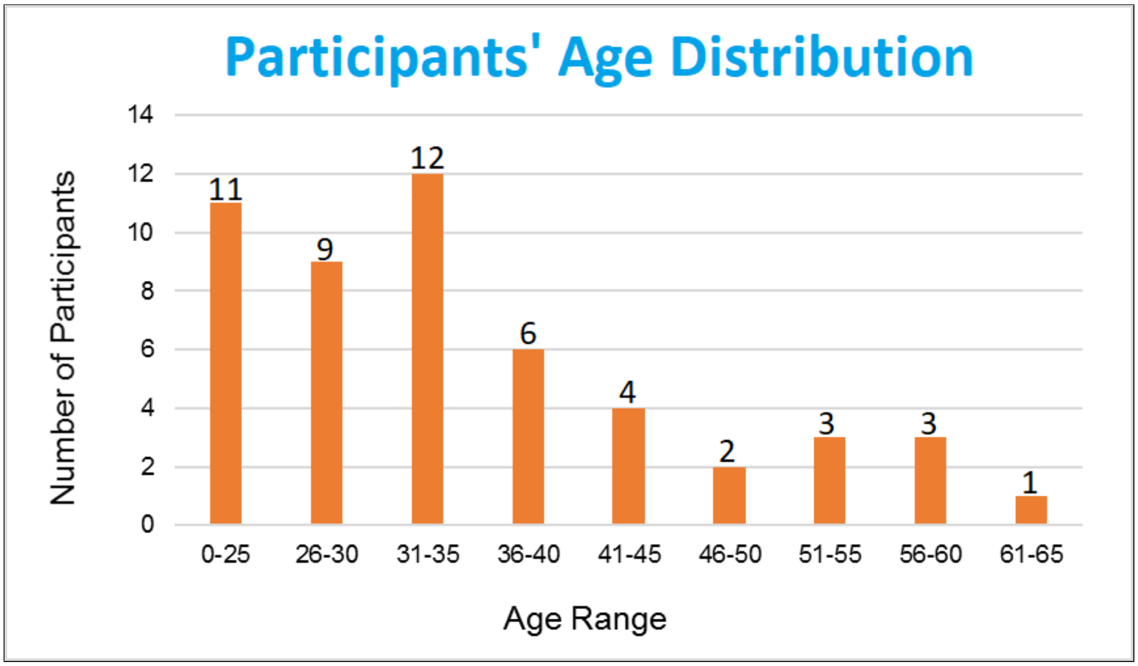
Competence: Players will perceive themselves as more competent when assisted more.

Enjoyment: Players will experience more enjoyment when assisted more.

Effort: Players will perceive that they invested less effort when assisted more.

Internality: Players will attribute their performance more internally when more assistance is provided.

Performance: Players will die less frequently and obtain a higher maximum score with increased assistance.



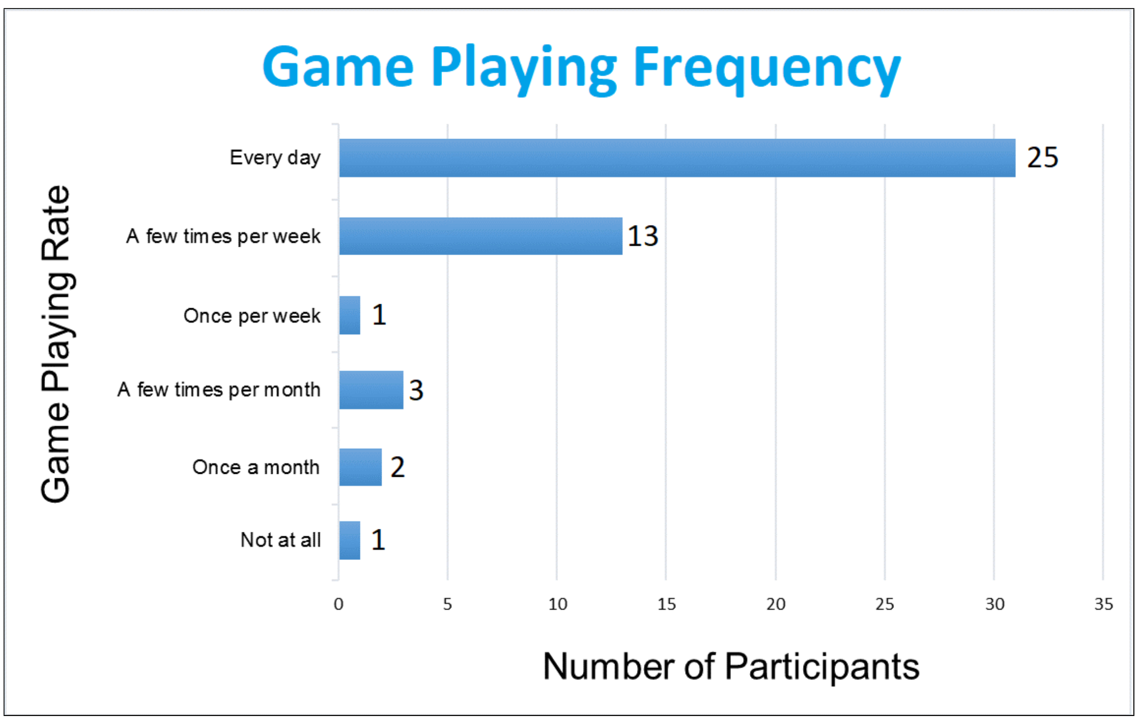
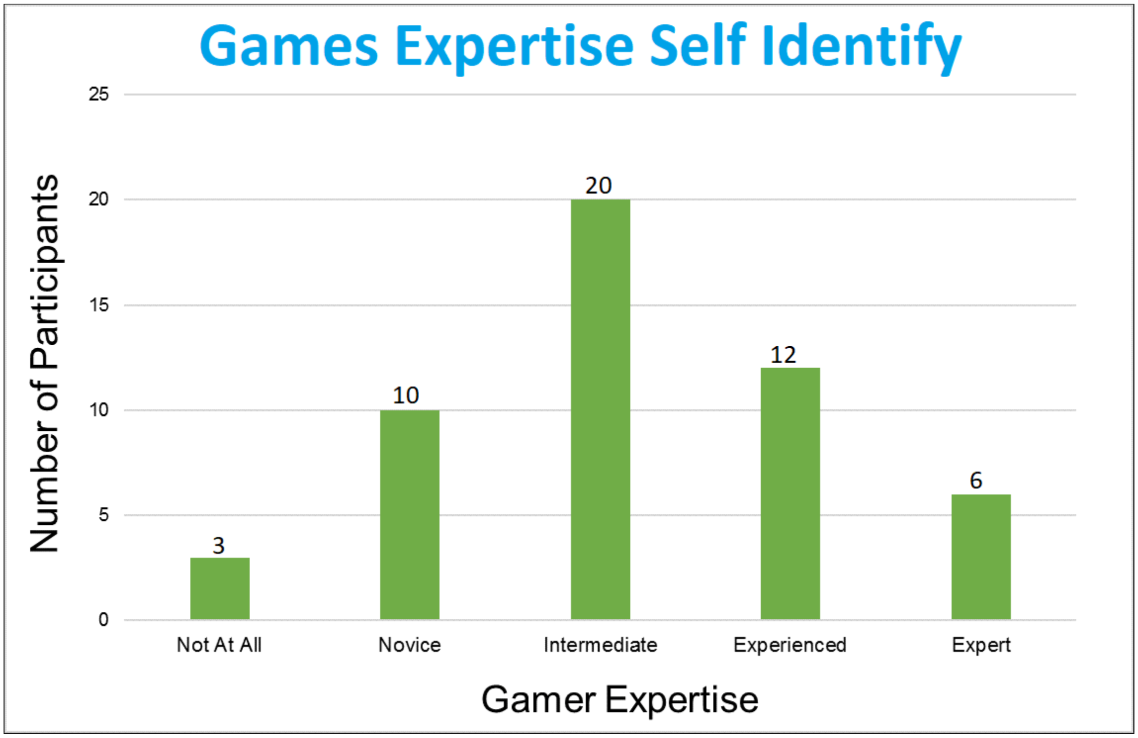


Figure 6.1: Brief Demographics (part 1) overview of the Recruited Participants for Study 2



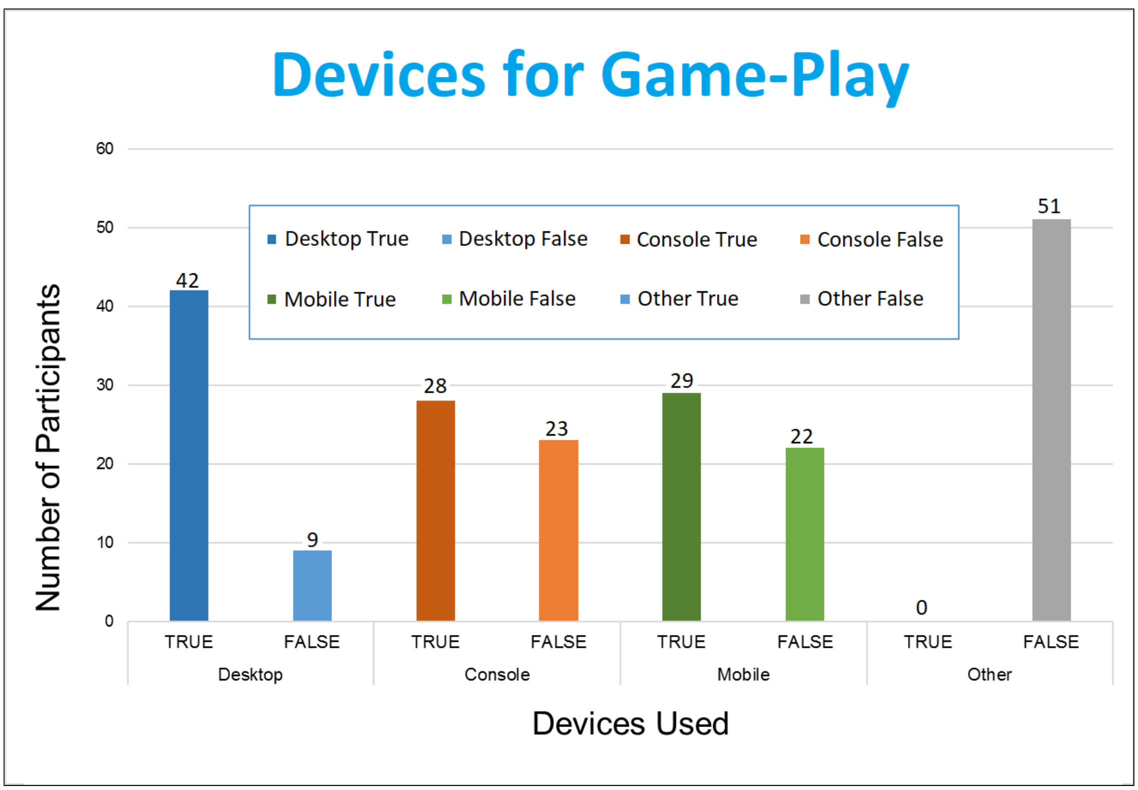


Figure 6.2: Brief Demographics overview (part 2) of the Recruited Participants for Study 2

Table 6.1: Repeated–measures MANOVA results: f–statistic, *p*–values and effect size for dependent measures for Study 2.

	$F_{2,100}$	p	η^2_p
Difficulty	110.0	**<.001**	.69
Enjoyment	20.9	**<.001**	.30
Effort	3.22	**.044**	.06
Competence	92.1	**<.001**	.65
Internality	11.2	**<.001**	.18
Number of Deaths	454.5	**<.001**	.90
Max Score	161.8	**<.001**	.76

6.5 Data Analyses

We conducted a repeated-measures MANOVA with the three levels of covert assistance (High Assistance, No Assistance, High Hindrance) as a within-subjects factor on the dependent measures of perceived difficulty, enjoyment, effort, competence, internality, number of deaths, and maximum score. Degrees of freedom for pairwise comparisons were adjusted using the least significant difference method. Alpha was set to 0.05.

6.6 Results

6.6.1 Player Performance

Assistance[1] significantly affected performance as there were main effects of assistance on both the total number of deaths and the maximum achieved score per game round (see table 6.1 and figure 6.3). Pairwise comparisons revealed that deaths decreased significantly with each level of assistance ($F_{2,100} = 454.5$, $p<.001$) and maximum score increased with each level ($F_{2,100} = 161.8$, $p<.001$).

[1] In this section while we use the term assistance, we refer to both Jump Assistance and Jump Hindrance where hindrance is considered as negative assistance.

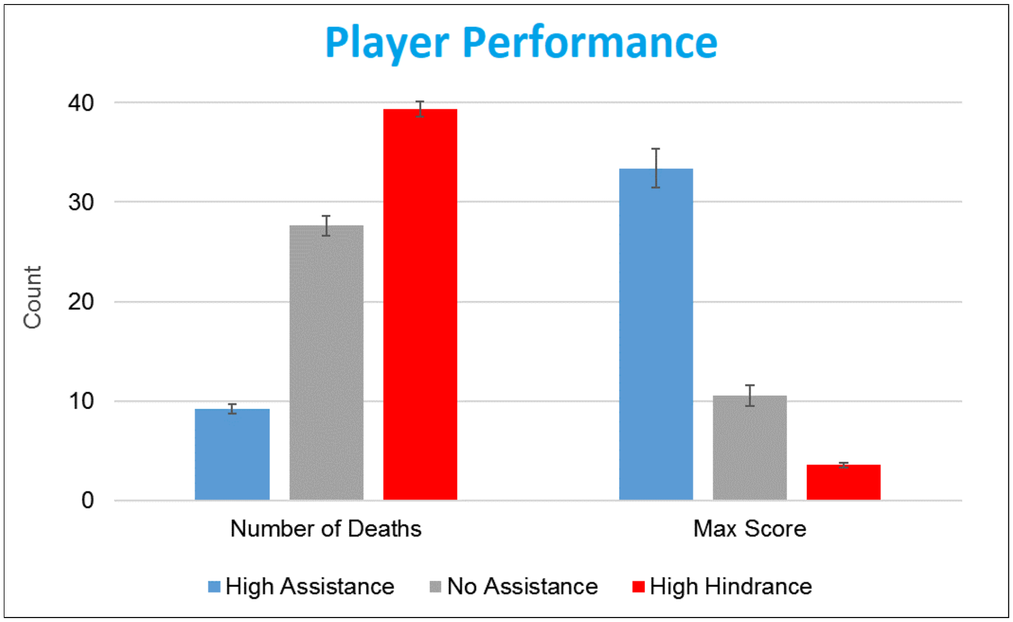


Figure 6.3: Means ± Standard Error of performance measures - Study 2.

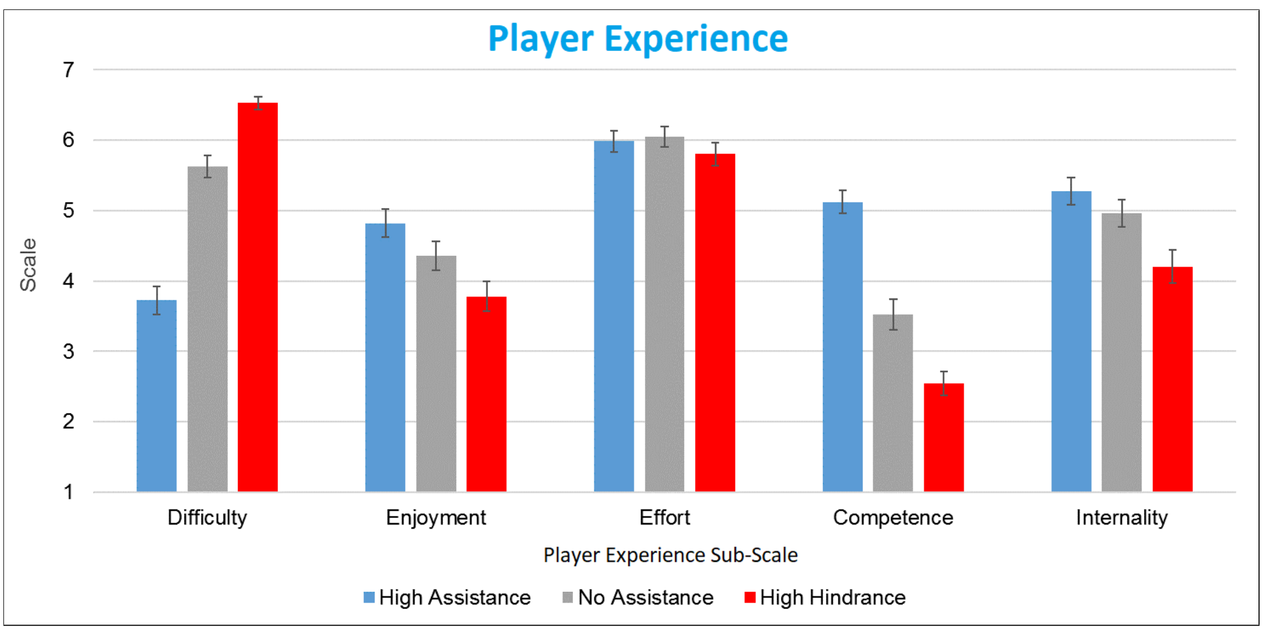


Figure 6.4: Means ± Standard Error for player experience measures – Study 2.

6.6.2 Player Experience

We observed significant main effect of assistance on each of the dependent measures (difficulty, enjoyment, effort, competence, and internality) (see table 6.1 and figure 6.4).

Difficulty: Pairwise comparisons revealed differences in perceived difficulty between each degree of assistance ($F_{2,100} = 110.0$, p<.001, ${\eta^2}_p = 0.69$), in which perceived difficulty decreased with each increase in assistance.

Enjoyment: There were also significant differences in enjoyment between each degree of assistance ($F_{2,100} = 20.9$, p<.001, ${\eta^2}_p = 0.30$), in which enjoyment increased with assistance. However, pairwise comparison revealed that perceived enjoyment was slightly lower in significance (p=.015) between the High Assistance and Neutral condition. For all other pairwise comparisons, p<.001.

Effort: There was a significant difference in invested effort between Neutral and High Hindrance ($F_{2,100} = 3.22$, p<.05, ${\eta^2}_p = 0.06$); however, no other pairwise comparisons were significant (p>.05).

Competence: There were also significant differences in perceived competence between each level of assistance ($F_{2,100} = 92.1$, p<.001, ${\eta^2}_p = 0.65$), in which perceived competence increased with each degree of assistance (all p<.001).

Internality: Finally, internality increased with assistance, and the difference between High Hindrance and Neutral was significant (p<.001) as was the difference between High Hindrance and High Assistance (p<.001) conditions; however, there was no significant difference between Neutral and High Assistance (p>.05). If both assistance techniques would show lower internality scores, we could have assumed that players noticed the manipulation. A more external attribution pattern associated with poor performance is coherent with our expectations of self-serving attribution biases [78].

6.7 Discussion of Study 2

The results of our second study indicated that we successfully manipulated player experience with Jump Assistance and Hindrance. As expected, players performed better with assistance and worse with hindrance, and this was reflected in their perceptions of the game's difficulty, their own competence, and their perceived enjoyment. Each of the main effects showed a proper trend of the constructs in the expected manner, showing how player performance increased gradually with higher degrees of assistance and how it was reflected with all other player experience measures. Furthermore, the effects on internality (significant difference only between High Hindrance and the other two conditions) are consistent with self-serving attribution biases [78], in that players attributed their failure to the system and their success to themselves. The large effects on performance and perceived difficulty also showed that the different degree of assistance and hindrance were very strong and were perceived by the players. However, the strength of assistance seemed stronger in High Assistance rounds - which yielded greater performance results (low death rates and high maximum scores - based on figure 6.3).

If we look back and compare with the results of our study 1, we would find that our current outcome had improved significantly. This assistance technique is more controllable and easily configurable than the previous one. With pole magnetism, it was not easy to establish the fact that if the strength of magnetism was doubled, the probability of success rate would increase. However with Jump Assistance/Hindrance, by examining the results reported at figure 6.3, we can formulate that if the assistance is doubled, the probability of failure would decrease to half. None of the participants reported noticing any covert assistance technique or abnormality in caveman's jump trajectory. In the study with pole magnetism, the self-serving attribution bias could not be established - which was shown as a significant difference in this study.

Chapter 7

Study 2B: The Interaction of Covert Assistance and Level Difficulty

In our previous chapter, we presented Study 2, in which the level difficulty was set to medium, and we looked into the strength of our assistance techniques. We implemented relatively strong amounts of assistance and hindrance. Other researchers might use games that are inherently more difficult or a lot easier or may wish to match the level difficulty to the expertise of the player irrespective of his previous game-play experience. Other researchers might also be limited in how strong the assistance can be while keeping it unobtrusive to the players. If we consider this relatively unnoticeable covert assistance to be an effective method of manipulating player experience, we need to identify and understand if the effects hold over different types of level difficulty. As such, we conducted a follow-up to our previous Study 2, in which we looked into at two different degrees of assistance in an easy and hard game difficulty level.

7.1 Experimental Conditions

We implemented two difficulty levels (easy and hard) and two degree of assistance (low and high).

7.1.1 Game Rounds with Level Difficulty and Assistance

These two game rounds are exactly similar to 'Low Level Difficulty' and 'High Level Difficulty' as explained in 3.2.1. Other properties of these levels are also explained in 5.1.

Easy Level

We varied the pole height and inter-pole distances at a very small range (25% of the base 5 Unity units, so the maximum and minimum 'Pole Height' and 'Pole Distance' are 8.75 and 1.25 respectively) throughout the whole level. The avatar could easily jump over more than one pole because the poles were relatively close to each other.

Hard Level

The pole height and inter-pole distance variations were set to a relatively wider range (75% of the base 5 Unity units, so the maximum and minimum 'Pole Height' and 'Pole Distance' are 6.25 and 3.75 respectively). It was harder to jump over any single pole because they were comparatively far from each other.

Low Assistance

A low degree of assistance was applied throughout the whole game round. We applied a random amount of assistance between 30% and 40% to every pole.

High Assistance

A high degree of assistance was provided throughout the whole game round. We applied a random amount of assistance between 70% and 80% to every pole.

In the 'Low Assistance' and 'High Assistance' rounds, the degree of assistance is always randomly assigned (within a particular range) to reduce noticeability by the players, and provide some variation.

We crossed these two factors (level difficulty and degree of assistance) to create four experimental conditions, i.e., Easy-Low, Easy-High, Hard-Low, Hard-High (see figure 7.1). In all these four conditions - players re-spawned at the beginning of the round and played each round for 4 minutes irrespective of the score achieved or the number of times the player died. As with the other experiments, the game was paused in the event of failure until the reset button was clicked.

7.1.2 Debriefing about the experimental conditions

Similar to Study 2, the debriefing session was performed at the very end of the experiment. We used the following debrief to let our participants know about the objective of our research and what type of manipulations were performed in Study 2B:

Explanation: We are interested in looking at the interplay between difficulty level and performance assistance. There were two difficulty levels - it is harder to play the game when the poles are spaced out more and vary more in height. We also had performance assistance. Sometimes we manipulated the jumps made by the caveman by adjusting their trajectory and initial velocity, which made it easier to land on the poles.

Performance: Because we manipulated the trajectory of your avatar, your performance was not always completely dependent on your skill level. We helped you out a little with trajectory assistance in all levels.

7.2 Procedure

The procedure was identical to Study 2. We received player consent before initiating the study session. Then we asked them to complete the demographics questionnaire as well as a questionnaire about their previous gaming experience. Then we navigated them into the relevant training session that had medium level difficulty, and the 'Degree of Assistance' was +55 throughout the whole game round. The next four rounds were the experimental conditions: Easy-Low, Easy-High, Hard-Low, Hard-High. To remove learning effect from our results, we counterbalanced the order of presentation using the factors: half of the participants started on the easy level (half of those with low assistance), completing both before moving on to the hard level (see table 7.1, easy game rounds are coloured in grey to help interpreting the table). At the end of each level, participants completed the player experience questionnaire (which is similar to the previous studies). In addition to this, we performed the debriefing session at the very end of the experiment session to summarize the performance manipulation techniques. An average of 30-35 minutes was required to complete the entire experiment

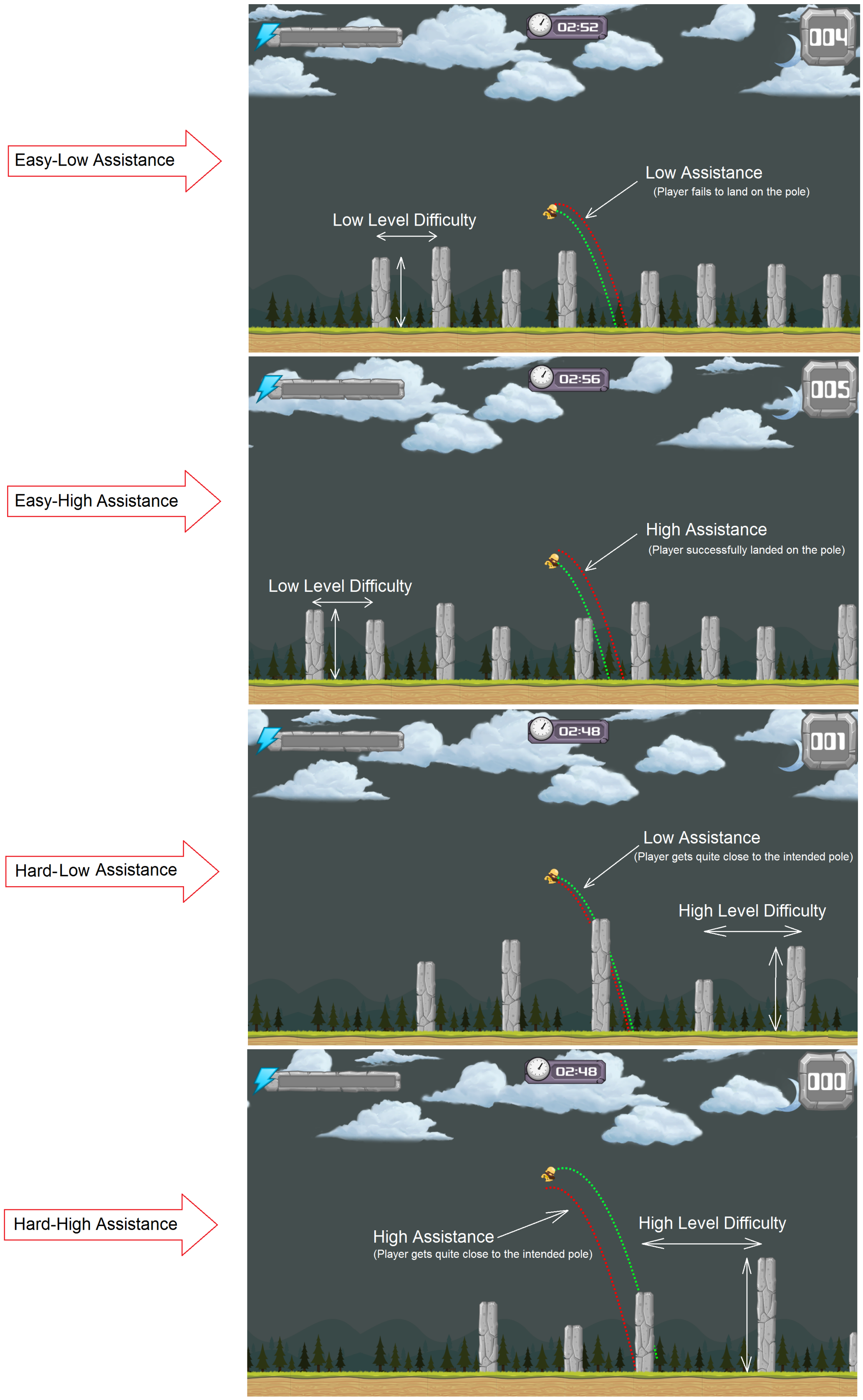


Figure 7.1: Experimental Conditions for Study 2B.

Table 7.1: Sequence of Experimental Conditions for Study 2B.

Condition No.	Game Rounds
1	Easy – High Assistance
	Hard – High Assistance
	Easy – Low Assistance
	Hard – Low Assistance
2	Easy – Low Assistance
	Hard – Low Assistance
	Easy – High Assistance
	Hard – High Assistance
3	Hard – High Assistance
	Easy – High Assistance
	Hard – Low Assistance
	Easy – Low Assistance
4	Hard – Low Assistance
	Easy – Low Assistance
	Hard – High Assistance
	Easy – High Assistance

session and participants were paid $5 each.

7.3 Participants

Through MTurk, we recruited 49 participants by setting a filter for those who had previously participated in Study 1 and Study 2. Similarly to Study 2, we excluded participants for spending less than 1.5 seconds per question on 2 or more of the combined player experience questionnaires (n=5) and for having a variance of more than 3 standard deviations above the mean on any construct (n=3), leaving 41 participants in our analyses. Unfortunately, while running this study, we lost a some demographics data due to a technical error. However, all player experience and performance data were retained, and so we did not remove these participants from our data analyses.

7.4 Hypotheses

We expected that manipulating level difficulty and different 'Degrees of Assistance' to create four game rounds would yield effects on:

Difficulty: Players will perceive a decrease in difficulty with increased assistance and decrease in level difficulty.

Competence: Players will perceive themselves as more competent when assisted more and with decreases in level difficulty.

Enjoyment: While the degree of assistance is higher and level difficulty is lower, players will experience more enjoyment.

Effort: Players' perception of invested effort will be less while the degree of assistance is higher and the level difficulty is lower.

Internality: Players will attribute their performance more internally when more assistance is provided. However, with the increase in level difficulty, players will attribute more externally.

Performance: Players will die less frequently and obtain a higher maximum score with increased assistance and decreased level difficulty.

7.5 Data Analyses

We conducted a repeated-measures MANOVA with level difficulty (Low, High) and Assistance (Low, High) as two within-subjects factors on the dependent measures of perceived difficulty, enjoyment, effort, competence, internality, number of deaths, and maximum score.

Table 7.2: Repeated–measures MANOVA results: f–statistic, p–values and effect size for dependent measures for Study 2B.

	Level Difficulty			Covert Assistance			Interaction		
	$F_{1,40}$	p	η^2_p	$F_{1,40}$	p	η^2_p	$F_{1,40}$	p	η^2_p
Difficulty	30.3	**<.001**	.43	25.4	**<.001**	.39	12.8	**<.001**	.24
Enjoyment	13.6	**<.001**	.25	5.8	**.021**	.13	0.14	.711	n.s.
Effort	1.2	.135	n.s.	.001	.981	n.s.	0.97	.332	n.s.
Competence	20.5	**<.001**	.34	24.8	**<.001**	.38	0.02	.880	n.s.
Internality	1.0	.317	n.s.	3.8	.057	.09	1.6	.213	n.s.
Number of Deaths	27.4	**<.001**	.41	38.7	**<.001**	.49	2.5	.121	n.s.
Max Score	132.4	**<.001**	.77	104.5	**<.001**	.72	36.9	**<.001**	.48

7.6 Results

7.6.1 Player Performance

Main Effects

There were main effects of level difficulty on the Number of Deaths ($F_{1,40} = 27.4$, $p<.001$) and Max Score ($F_{1,40} = 132.4$, $p<.001$) (see table 7.2) showing that players died more and had lower maximum scores in the hard conditions than easy conditions (see figure 7.2). Similarly, a main effect of Assistance showed that players died less ($F_{1,40} = 38.7$, $p<.001$) and had higher maximum scores ($F_{1,40} = 104.5$, $p<.001$) when assisted more (see figure 7.2).

Interaction Effect

The significant interaction on Max Score ($F_{1,40} = 36.9$, $p<.001$) showed that the effect of assistance was greater in the easy conditions than the hard ones (see figure 7.2) which resulted into a significantly higher Max Score in the 'Easy High Assistance' (the easiest round) comparing it to other conditions. There was no significant difference in Max Score while comparing 'Easy Low Assistance' and 'Hard High Assistance' - this shows the balanced trade-off between the level difficulty and the 'Degree of Assistance'.

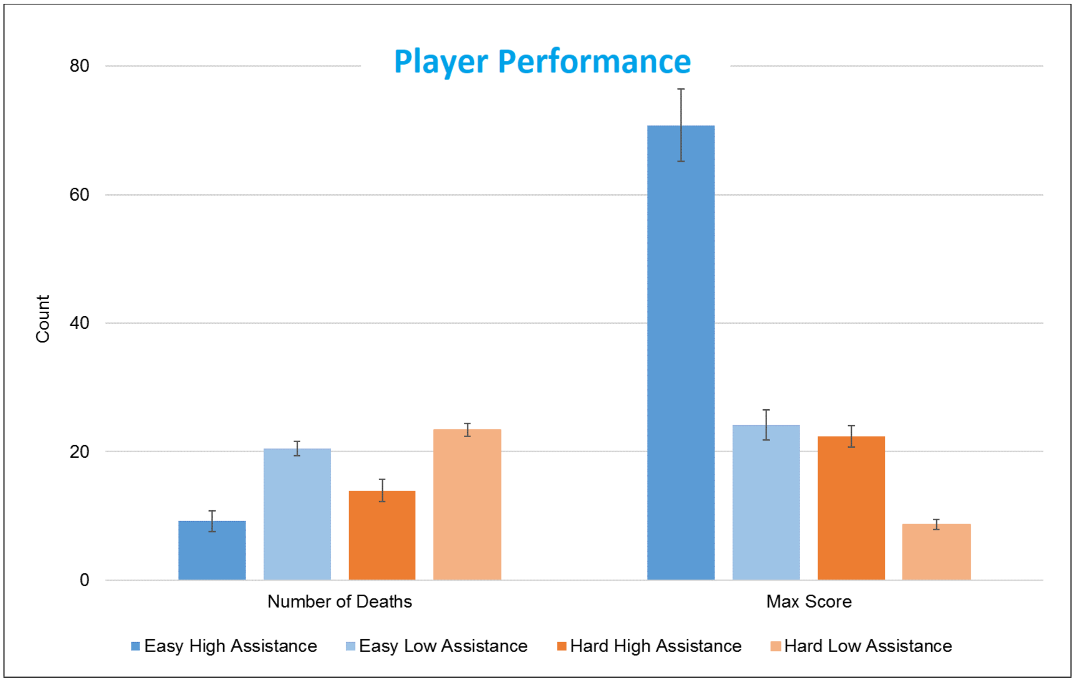


Figure 7.2: Mean ± Standard Error for Player Performance measures – Study 2B.

7.6.2 Player Experience

Main Effects

1. There were main effects of level difficulty on perceived difficulty ($F_{1,40}$ = 30.3, p<.001), enjoyment ($F_{1,40}$ = 13.6, p<.001), and competence ($F_{1,40}$ = 20.5, p<.001) but not on effort ($F_{1,40}$ = 1.2, p=.135) or internality ($F_{1,40}$ = 1.0, p=.317) (see table 7.2 and figure 7.3).

2. There were main effects of assistance on perceived difficulty ($F_{1,40}$ = 25.4, p<.001), enjoyment ($F_{1,40}$ = 5.8, p=.021), competence ($F_{1,40}$ = 24.8, p<.001), but not on effort ($F_{1,40}$ = .001, p=.981) or internality ($F_{1,40}$ = 3.8, p=.057) (see table 7.2 and figure 7.3).

Interaction Effects

As figure 7.3 shows, decreasing the level difficulty or increasing assistance resulted in more experienced competence and enjoyment. There were no significant interactions (p>.05) between level difficulty and assistance on enjoyment, effort, competence or internality (see table 7.2); however, a significant interaction on perceived difficulty ($F_{1,40}$ = 12.8, p<.001)

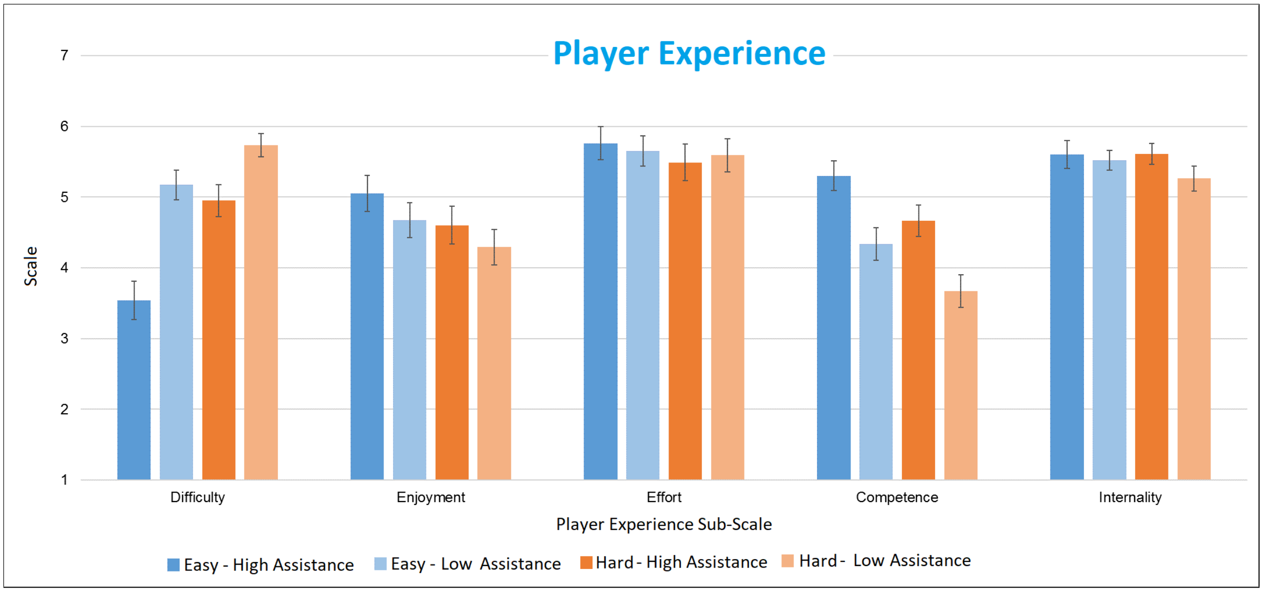


Figure 7.3: Mean ± Standard Error for Player Experience measures – Study 2B.

simply reflected that for maximum deaths, the effect of assistance was greater in the easy conditions than the hard ones (figure 7.3).

7.7 Discussion of Study 2B

The goal of Study 2B was to determine whether the results of Study 2 held, even with different levels of game difficulty. Although the effect of assistance yielded greater performance advantages in the easy condition, we saw no significant interactions between level difficulty and assistance on any of the experience measures (except for perceived difficulty). As such, we can conclude that the approach and the system works, regardless of level difficulty. As such, we can use covert assistance much in the same way as leaderboards were used in [11]. Our effect sizes, however, were much larger than in Study 2, explaining between 18-69% of the variance in the experience measures in Study 2 and 25-43% in Study 2B.

However, there are some small differences in the main effects of enjoyment. Although, for both cases the results were significant; level difficulty yielded into slightly larger significant result than covert assistance. This reflects back to the results of interaction effects that showed the effect of assistance seemed to be greater in the easy conditions than the hard ones.

If we compare these results with Study 1 (compare table 7.2 and table 5.1), we would see that the interaction effects were slightly different. In Study 1, we did not observe any significant interaction for perceived difficulty ($p > .05$) while effort was marginally significant ($p < .05$). This showed that participants needed to invest more effort while our implemented assistance technique was pole magnetism. The situation changed for Study 2B - in which we observed that participants invested effort remained same all conditions.

Chapter 8

Study 3: Manipulating Experience at a High Resolution

Study 2 showed that we could use different degrees of assistance (and hindrance) to manipulate player experience of success and failure regardless of the game's overt level of difficulty and we managed to successfully achieve this outcome. However, previous studies manipulated the player experience at the resolution of an entire game round. We wanted to identify if our system could be used to assess momentary and more complex player experience effects, since we know that a player's retroactive judgment of gaming experience can be affected by the moments of climax and the ending events of that particular experience [19]. However, we also know that there is a trade-off between the skills of the player and the level of difficulty when designing a game, and that optimal balance is obtained whenever these two are well–matched [1]. More momentary control would allow researchers to ask interesting research questions, such as those asked by Gutwin et al. in their paper on peak and end experiences in games [38]. In Gutwin et al.'s work [38], they used a series of games to manipulate the position of the difficulty or challenge in a game level, showing effects on pX. Instead of looking just into the overall performance, they also investigated other aspects of player experience like perception of performance, challenge, tension, effort, and enjoyment. They conducted two studies by implementing three custom games ('Match-3', 'Whac-a-Germ', 'Shootout'), operationalizing difficulty and challenge uniquely in each.

In this study, we used the Jumping Caveman system to manipulate the sequence of different degrees of assistance in a single game round. Our objective was to keep the overall difficulty of the game round the same throughout, but look at whether having the assistance at the end or the hindrance at the end would affect player perceptions. The design of the experiment was influenced by Gutwin et al.'s work [38], in which they used a sequence of game rounds from 'Match-3' and 'Whac-a-Germ' to represent a complete full experience where each game round represented one type of experience (either success or failure).

8.1 Experimental Conditions

We created two rounds that were identical in overall difficulty and assistance by using round sections. There were three sections in each round.

Base Section: In the base sections, players were assisted with a constant low amount (30%) - which means the degree of assistance was 30% for each of the poles located in this section. There were two base sections of 3 poles each.

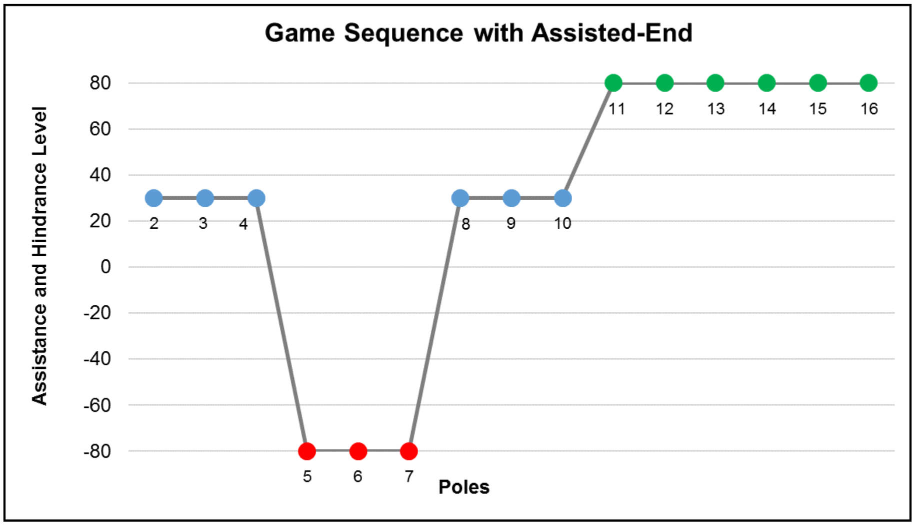
Hindrance Section: In the hindrance section, players were hindered with a high level of hindrance (80%) for 3 poles.

Assistance Section: In the assistance section, players were assisted with high assistance (80%) for 6 poles. We used 6 poles in this section because the high level of assistance meant that players were successful most of the time and would finish this section much more quickly than the hindered section. Our intention was to make the assisted and hindered sections comparable in terms of time spent.

The pole height and spacing of the 3 or 6 poles were kept consistent in each section to ensure that difficulty of the level did not vary from one participant to another. We ordered these sections to create two levels (see figure 8.1) that were identical overall, but differed only in the ordering of the assisted and hindered sections. To players, the sections were imperceptible and it played like a single game round.

8.1.1 Assisted-End

In this round, players completed a base section (comprised of 3 poles), then the hindered section (comprised of 3 poles), another base section (comprised of 3 poles), and finished with the assisted section (comprised of 6 poles) (see figure 8.1).



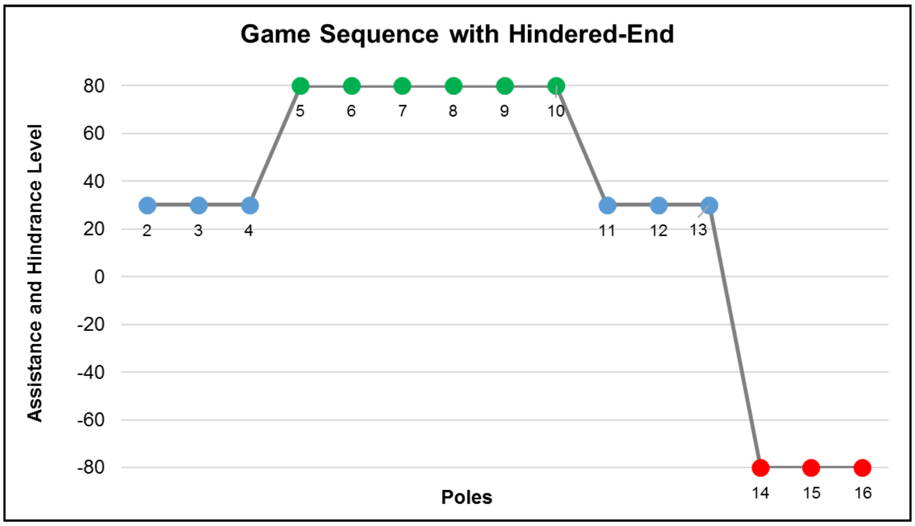


Figure 8.1: Experimental conditions for Study 3 showing the pole sequence and associated assistance and hindrance (Top image: Assisted-End, Bottom image: Hindered-End.)

8.1.2 Hindered-End

In this round, players completed a base section (comprised of 3 poles), then the assisted section (comprised of 6 poles), another base section (comprised of 3 poles), and finished with the hindered section (comprised of 3 poles) (see figure 8.1).

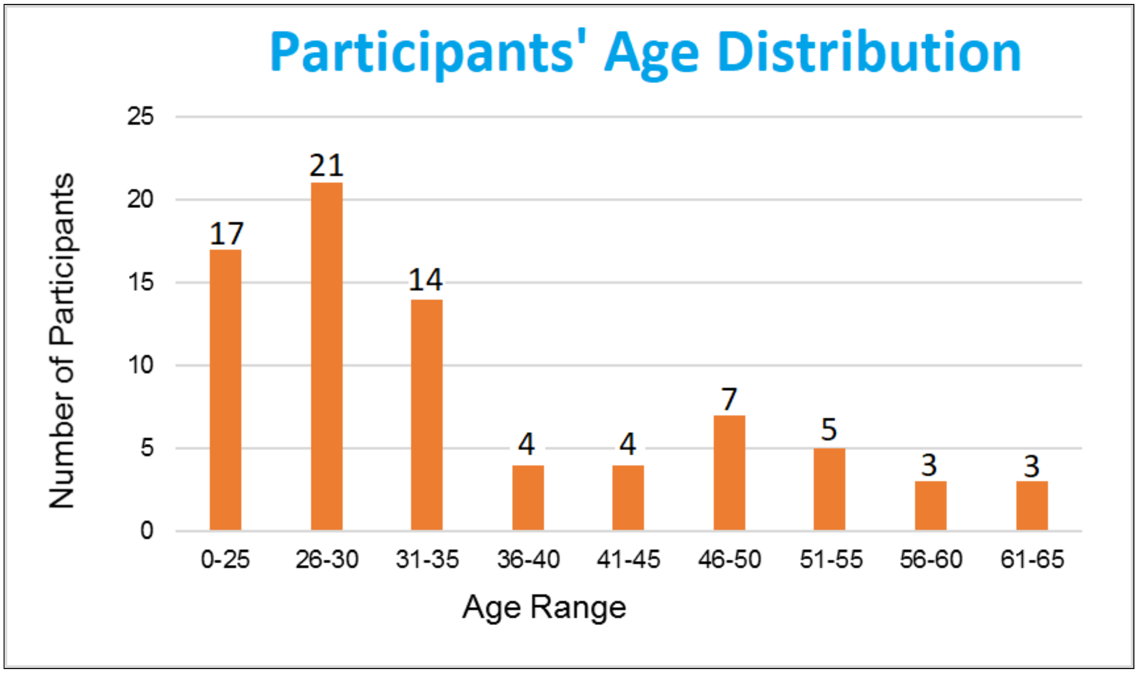
In contrast to Study 2, we had players re-spawn at the last-completed pole and complete the entire level, rather than play for a prescribed duration. This was necessary so that every participant would finish each level and experience the Assisted-End or Hindered-End. However, to ensure that we were paying participants ethically, we had a timeout of 10 minutes for each game round; however, no player encountered a timeout in our experiment. Furthermore, we spaced the poles so that jumping over a pole was no longer possible. This was to ensure that participants had comparable experiences in each level.

8.2 Procedure

After receiving participant's informed consent and demographics information, they were presented with a training round. In this training round, participants completed the same number of poles as in the experimental conditions, but with Low Assistance throughout (i.e., 30% for every pole). Followed by the training round, our Assisted-End and Hindered-End levels were presented and the order of presentation of each was counterbalanced. Similar to all other studies, we presented player experience questionnaire (enjoyment, effort, competence, internality, and difficulty) after each round in which the participants could reflect upon their momentary play-experience. Before proceeding in to debriefing session, they were asked to provide feedback about the game rounds. Similar to all other studies, we debriefed our manipulation methods. As they have to complete the whole game round to proceed in to the next page; even though the number of poles were limited and there were only two game rounds in this session; the completion time for this experiment varied between 15 minutes to 35 minutes. To compensate their time and effort, each of the participants were paid $5 through MTurk.

8.3 Participants

We recruited 78 participants, who had not participated in Study 2, to participate in the study through MTurk. As in Study 2, we excluded participants for spending less than 1.5 seconds per question on 2 or more of the combined pX questionnaires (n=5) and for having a variance of more than 3 standard deviations above the mean on any construct (n=1). We additionally



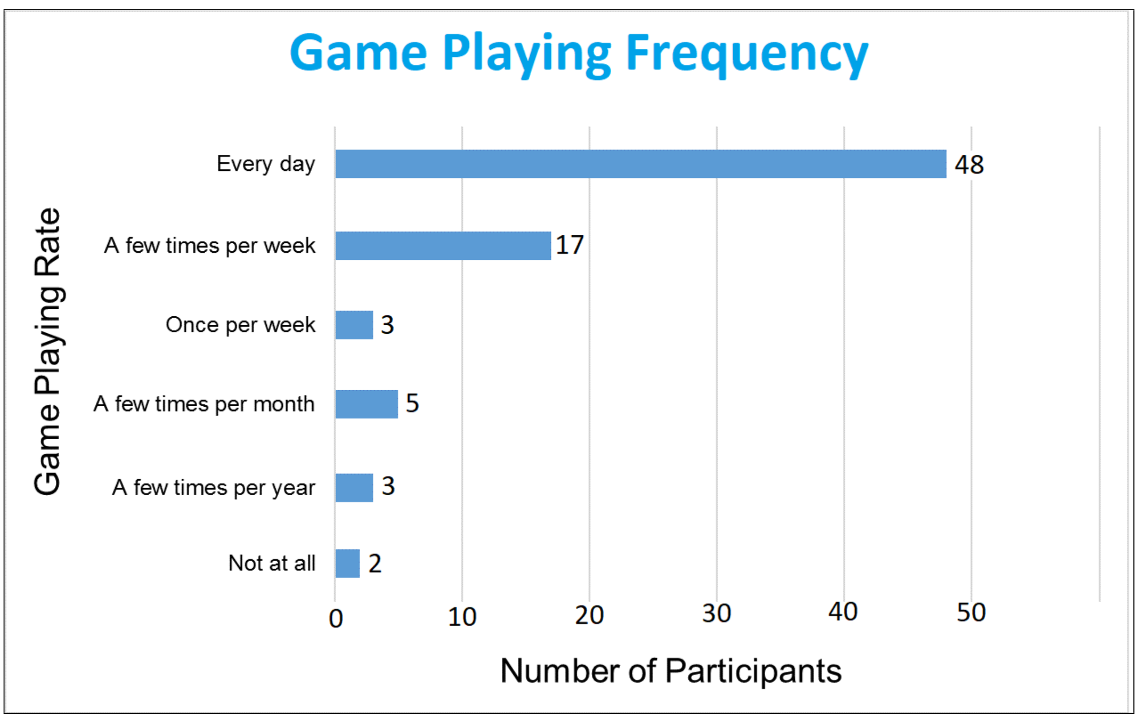
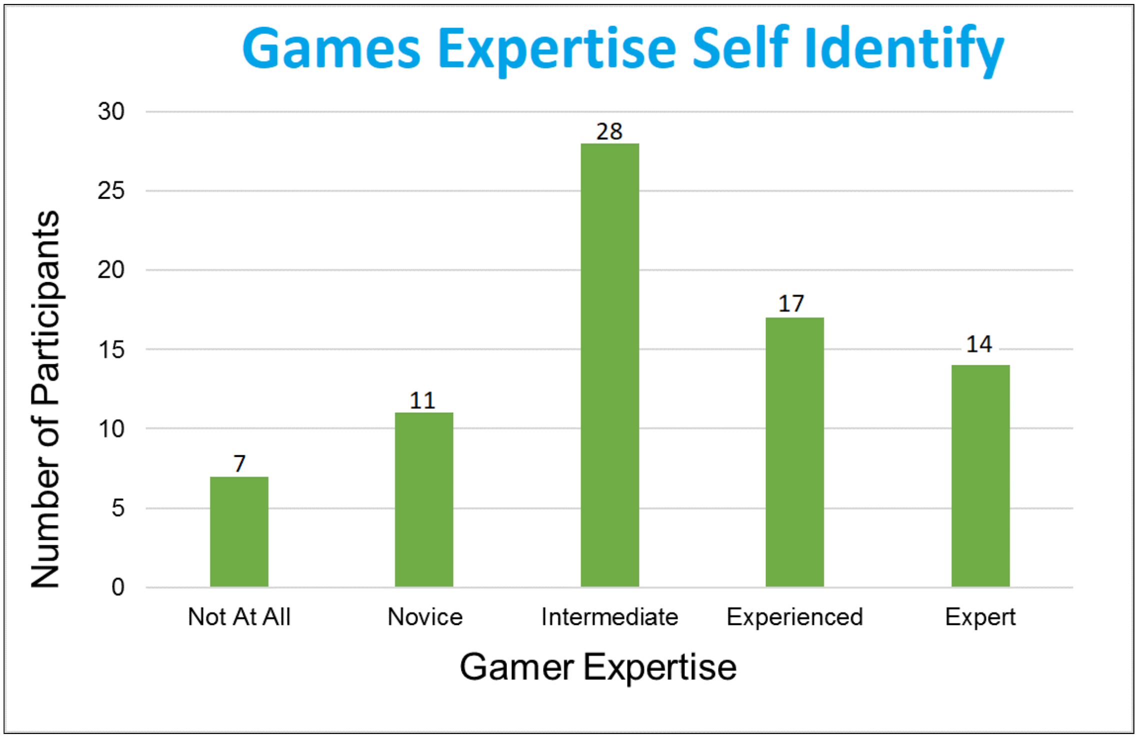


Figure 8.2: Brief Demographics overview (part 1) of the recruited participants for Study 3



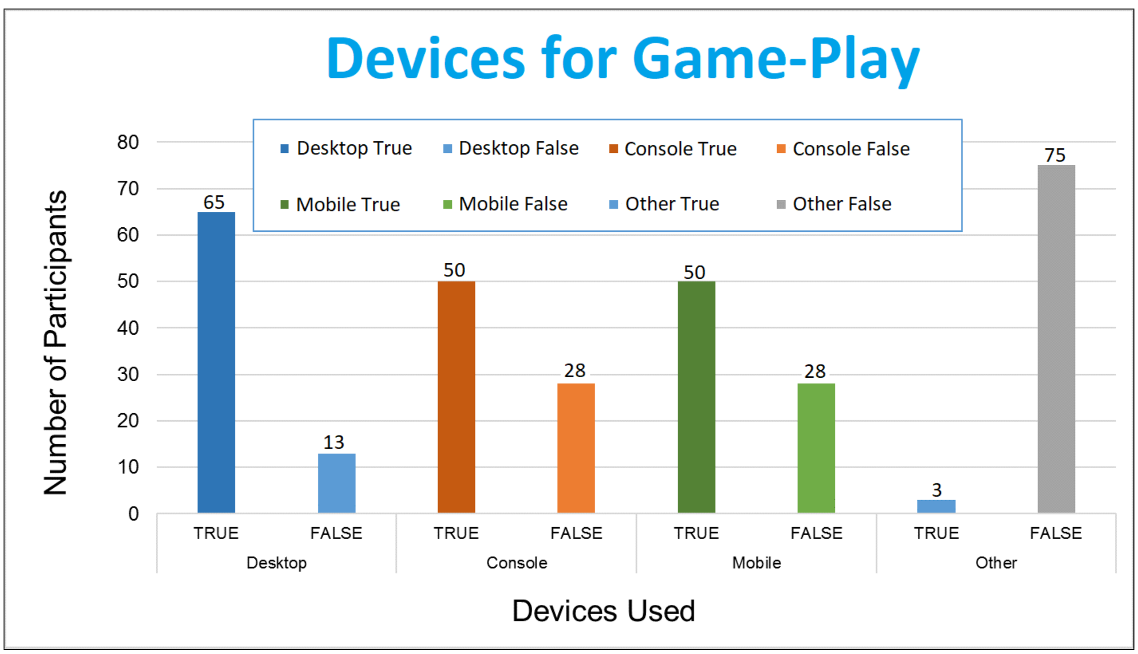


Figure 8.3: Brief Demographics overview (part 2) of the recruited participants for Study 3

removed 4 outlier participants who were greater than 3 standard deviations below the mean on enjoyment in all sequences as their experience was very negative throughout, which skewed the distribution; there were no outliers on the positive end of enjoyment. We retained 68 participants (Mean = 35.13 years, Standard Deviation = 12.68 and 48.5% Female, 91% right-handed) in our analyses. A brief overview of our recruited participants are shown in figure 8.2 and figure 8.3.

8.4 Hypotheses

Unlike other studies, we did not have a particular hypothesis for this one. Since Gutwin et al.'s work [38] found significant differences for the peak and end experience of the players while using a series of games, we expected that we might find similar effects while manipulating experience in particular portions of one game round. However, there was no theory to base the direction of the differences on.

8.5 Data Analyses

As it always appeared first, we excluded the training round from our analyses. We conducted a paired-samples t-test of Sequence (Assisted-End, Hindered-End) on the dependent measures of Difficulty, Enjoyment, Effort, Competence, Internality, and Number of Deaths. There was no Maximum Score recorded, because participants finished every round.

8.6 Results

8.6.1 Player Performance

As expected, there was no significant difference in number of deaths (Assisted End: mean = 12.8088, SD=7.3632, Hindered-End: mean=12.3824, SD = 12.15574) based on sequence (t_{67}= 0.303, p=.763). Because players completed all the poles and these were identical in both sequences (but ordered differently), we did not expect performance differences.

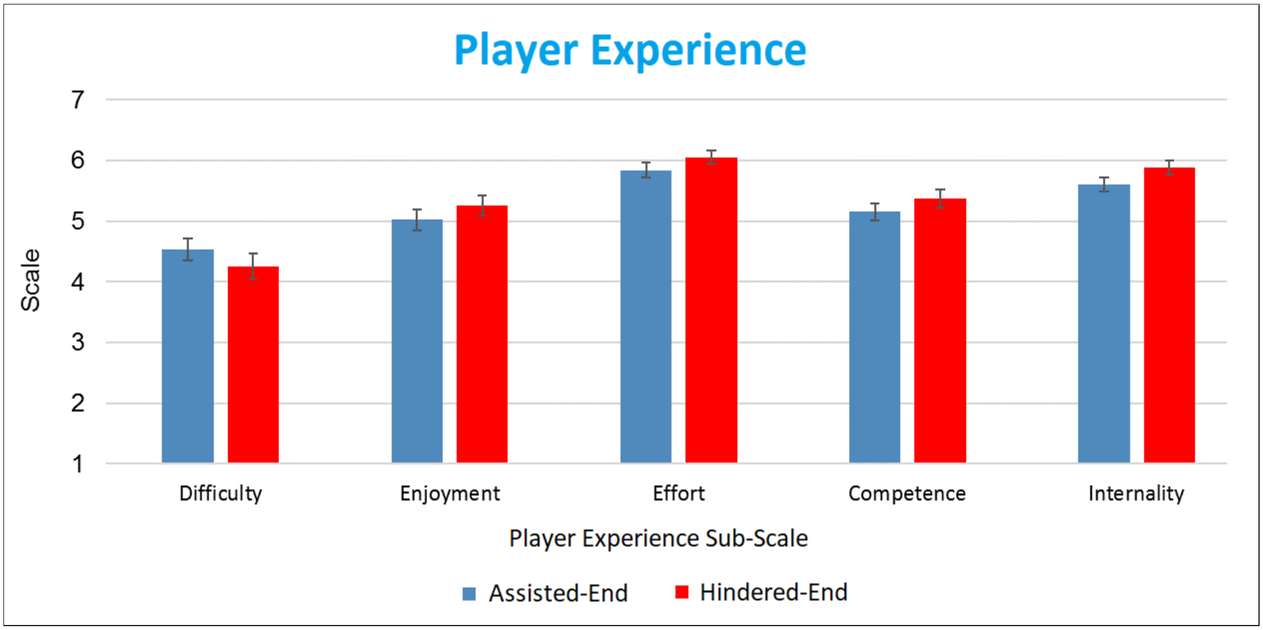


Figure 8.4: Mean ± Standard Error dependent Measures of Study 3.

8.6.2 Player Experience

There was no difference in perceived difficulty based on sequence (t_{67}=1.41, p=.164). There was a difference in enjoyment (t_{67}=2.31, p=.024) with participants preferring the sequence with hindrance at the end over the sequence with assistance at the end (see figure 8.4). There was a similar significant effect on effort (t_{67}=2.25, p=.028), with participants feeling like they invested greater effort in the sequence with hindrance at the end (see figure 8.4). There was no significant difference in perceived competence (t_{67}=1.67, p=.100). There was a significant difference in internality (t_{67}=2.36, p=.021), with participants experiencing greater internality in the sequence with hindrance at the end than assistance at the end (see figure 8.4).

8.7 Discussion of Study 3

Our study 3 revealed that the hindered-end was perceived as more enjoyable, as players having invested more effort, and also attributed their performance more internally. Figure 8.4 shows that the effects are small, but the manipulation is also very subtle; the conditions were identical, but sequenced differently. In our study, we had a assisted peak - hindered end or hindered peak - assisted end combination only, and Gutwin et al. [38] used both positive or negative peak-end (assisted or hindered peak-end), which is different. In addition to this, at first glance, our results seem to contradict those found in [38]; however, there are a few differences that explain the findings. In Gutwin's work, players were not required to the finish the game round and so easier game rounds were completed with higher success rates. In that particular study [38], participants chose easier game rounds as more fun, easier, and more interesting. In addition to this, there were significantly more responses to repeat the easier game round than the harder one. In contrary to this, in our experiments, all participants always finished the game round, so the hindered-end mirrors a boss battle that was successfully won; where in [38], the harder game round always ended with participants failing. Since participants enjoyed the sequence with the hindered-end, like a final battle, it is not surprising for them to enjoy the hindered-end more and is also the guiding principle behind games that feature boss battles.

Chapter 9

Discussion

Our goal was to create an open-source tool for games user research that can experimentally manipulate the experience of success and failure while still maintaining the mundane realism of play. This tool effectively balances the trade-off between experimental realism of play and mundane realism. It allows the researchers a lot of experimental control without sacrificing the ecological validity of the 'gameplay experience'. We made the tool readily available and configurable by games user researchers who might possess little to no technical expertise on game development. This allows them to modify the game objects according to their proposed hypothesis without investing too much time and effort. Previous literature [18, 108, 109, 110] suggested that hidden assistance techniques are a powerful tool to manipulate the player perception of the game and remain quite un-noticeable to the players. Based on this, we built and evaluated the jumping caveman tool that uses hidden assistance and hindrance to manipulate player experience. Our findings suggested that in our system, we can effectively influence player performance by inducing states of success and failure. Furthermore, the manipulations of game performance translated into the expected changes in player experience - as we can see in our Study 2. In a follow-up study (Study 2B), we demonstrated that these effects can be scaled, allowing us to implement assistance in different degrees of strength. The effects of our manipulations are also stable with varying levels of difficulty. In our final study, we established that we could manipulate player experience at a high resolution, by changing assistance within a particular game round.

9.1 Implications for Games User Research

Our tool allows researchers to ask and answer a variety of research questions related to game difficulty with added assistance or hindrance and perceptions of success and failure. In this section, we discuss three potential research topics to illustrate the usefulness of the tool developed.

9.1.1 Player Response and Resilience to Failure

An interesting question in game design is how players respond to failure and how do they react to overcome the situation. What type of experience a player goes through in the face of repeated failure while confronting the same enemy in the same spot of a game? Does he gets disheartened and stop playing or keep continuing expecting that this would be his last chance of succeeding or he might discover a new way of defeating the enemy? While conventional models of player motivation suggest that games are enjoyable because they make players feel competent and in control, some games are (in)famous for being challenging and impossible to beat. Games like Dark Souls [32], for example, are very successful despite their impossible difficulty and their disregard for catering to the player's need for success. Using our tool, we could investigate why some players feel disheartened while other feel motivated by failure. Some other research questions also revolve around this particular scenario of failure, which can be answered by our tool if the configuration file is modified accordingly:

- What is the relation between player performance and failure in a game? Does a player's previous gaming performance have any connection on how he experiences failure or death while playing a particular game? Does it differ when he plays in different difficulty settings?

- How does a particular type of player responds to failure or death in a game? What kind of relation does it have with his individual traits? For example, some people respond to the event failure by becoming motivated to accomplish their goals, while others get paralysed and start making more mistakes. The Action Control Scale (ACS) [65] can be used in the context of different games to figure out how someone would responds to losing in-game objectives.

- In particular game matches, players might undergo the feeling of wrath and frustration but the drive to achieve success in a particular scenario may keep him motivated to let go of the feeling of failure and go through the game round again. What motivates players to keep trying harder and harder in the face of repeated failure? Do they figure out a new technique to achieve success? Since, our tool logs every action performed by the player, it would give detailed information on how players overcome difficult pole positions.

- Researchers can experimentally determine game performance regardless of players' actual skill level and investigate the effect of performance on motivation and how this relationship is influenced by personality, age, experience or any other player characteristics.

9.1.2 Interplay between Covert and Overt Difficulty

The Jumping Caveman tool also allows researchers to investigate the interplay between objective and subjective difficulty. There are many research questions that could be asked:

- What determines how difficult a player thought a game was? Does the overt game difficulty matches with it's covert difficulty level? We can use the in-game performance metrics and compare it with player experience measures to identify it.

- What would be the best way to alternate objective or subjective difficulty if we want to manipulate the feeling of success or failure for in-game events? To verify this, we can create separate game conditions where in one condition we would manipulate their subjective difficulty using hidden Jump Assistance/Hindrance and in other case we would alter the overt difficulties by adjusting pole positions.

- What type of interaction can be found between objective game difficulty and the subjective one? Do players experience subjective difficulty the same way they experience objective difficulty? Comparing the game logs of overt vs. covert manipulation can answer these questions.

- To improve player performance, between subjective and objective difficulty, controlling which one is more important? Does game balancing changes if we control subjective difficulty and objective difficulty separately? In our Study 2, we manipulated both overt and covert difficulty simultaneously. If we change the study design and separately manipulate them, we would be able to identify what type of manipulations improves player performance.
- Is subjective difficulty determined by how well a player performed? If so, are there primacy or recency[1] effects? Perhaps subjective difficulty is an interplay between how difficult a challenge looks and how well a player performed.

The effect of performance on subjective difficulty might be different in easy games compared to difficult games. Studies that use objective difficulty to influence how successful players are conflate the concepts of objective difficulty and performance. Our system allows researchers to manipulate both of them separately.

9.1.3 Learning Curves of Players

A growing body of research has investigated the effect of scaffolding techniques on the learning curves of players [56]. Games often use assistance techniques to help novice players ease into the challenges of the game. First-person shooter games, for example, can use aiming assistance and open world games often help their players navigate through terrain by showing which direction they should be going. Psychological frameworks on learning, such as the guidance hypothesis [95], would suggest that assisting someone in a task prevents them from becoming better. The trade-off between assistance and learning is difficult to investigate as learning curves are not always easy to track and in many research settings actual skill is impossible to disentangle from performance under assistance. Using our tool, researchers will be able to investigate long-term effects of assistance on actual skill as assistance is easily implemented and scaled, performance is logged for each pole allowing high-resolution analysis, and the game logging allows researchers to determine how far off a pole the player actually would have landed without assistance.

[1] How well a player remembers the sequence of events

9.1.4 Player's Choice of Game

Player preference plays a major role in Games User Research, and it can depend on what type of experience a player went through while playing the game. Imagine there is a game series under the same title and the second version of the game is disliked by most of the audience because of its extremely lengthy and boring missions. Players might eventually opt-out themselves even buying the third sequel of the same series without knowing whether the issue has been resolved or not. Alternatively, preference can be entirely dependent on player enjoyment and previous research has not identified the role of difficulty in these choices. Our study 2 showed that players enjoyed and felt more competent with the easy levels. But Study 3 showed that players enjoyed a round with a challenging end point. Player Experience differs in game rounds depending on what players achieve at the ending moments of the game. Players feel more enjoyment and competence while they successfully end the game, even if the difficulty was higher.

9.2 Limitations

This research has some limitations that should be acknowledged:

- **Not a tool to advance our knowledge on game design:** This research was not intended to advance our knowledge of game design; our intent was to design, develop and evaluate a tool that can be used by research groups to ask questions that relates to player behaviour while accomplishing in-game objectives.
- **Configuration file issues:** While developing the research tool, our primary focus was to make it easily configurable even without accessing or editing the source code - which would make it easier for games researchers to create different game conditions without spending too much time on learning different game programming aspects. However, even though, the configuration file can be edited with a regular text file viewer software, to confirm the listed pole positions and their associated properties are working properly, the game still needs to be opened in Unity for play-testing. This might still require limited knowledge of using the Unity game engine in order to use our research tool. In

addition to this, the configuration file is written in a specific file format (.csv) and it requires to be recreated in a particular format (following the guidelines of creating a csv file and specifications reported at 3.2.4) to ensure that it can be recognized by the source code. The length of each file, proper use of line indentations and proper format of csv file is required to adjust the pole positions and its associated properties. If any researcher wants to create their own file format, he/she would have to edit the source code to support the new configuration file.

- **Limitation of game genre:** The 'Jumping Cavemen' system is built in the genre of a casual game. While assistance as a general mechanic has shown consistent results over many genres of games (e.g., driving [5, 18], FPS [110, 108, 109, 27]), results found with our system can only ever be generalized with caution beyond casual games. However, the results can still be generalized in respect to other genres of games as well if we consider the fact that we showed how the player experience would look like when we make them successful or failed in achieving particular game objective.

- **Other aspects of game-play:** We manipulate experience through faking success and failure; researchers who wish to study other aspects of player experience (e.g., social play, eudemonia)[2] cannot use our open-source system.

9.3 Future Work

We have verified the efficacy of the tool based on two types of assistance techniques and different difficulty levels. In addition to this, the tool is ready to be tested on its other features that have been described in 3.2.3. However, based on the results we have obtained about the efficacy of the implemented assistance techniques, we would consider the following scenarios listed as our future work:

[2]derived from a Greek work - "Eudaimonia", meaning happiness or welfare.

9.3.1 Dynamic Difficulty Adjustment

To achieve more control over the experiment, we set our objective difficulty (pole positions and inter-pole distance) and assistance prior to the game initialization by using a configuration file. With minimal modification, this could be changed to be adjusted dynamically, with both the subjective and objective difficulty being altered based on player performance in real-time. A hypothesized step-by-step procedure is demonstrated here on how we can implement this technique:

- **Game Initialization** - Players would start a game round with an average level of difficulty with no assistance or hindrance.

- **Covert manipulation** - assistance or hindrance level would be calculated and deployed based on how many times the player died or successfully landed on the pole. If the caveman continually landed on several subsequent poles, we would introduce hindrance, and it would be gradually increased until he died. After the resurrection of the caveman, it would have no hidden assistance or hindrance like the beginning.

- **Overt manipulation** - the poles would be placed at a wider distance with larger variation while the player kept landing successfully. Poles would be brought closer with smaller variation whenever he died.

9.3.2 Multi-player Game Environment

Due to the robustness of the tool, it can also be extended to a multi-player environment. Two players could simultaneously play a specific game round and compete with each other to finish the round. Any type of level difficulty or assistance could be introduced to adhere with particular player expertise. Multi-player techniques would also help the researchers to ask questions about player balancing for casual games.

9.3.3 Other Possible Research Ideas

Here are some potential research ideas that other researchers might want to explore:

- One of the potential research ideas is to identify how players respond differently depending on their personality traits, using time pressure. Since we already have introduced timer settings in our tool and the game automatically ends while the timer ends, it is easily configurable to a 'Time Trial' - based game round. Researchers can explore how a diverged group of players respond to this time pressure or how different play experiences like enjoyment, competence, perceived difficulty relates to experience under pressure.

- Other researchers might want to introduce a separate type of hindrance technique based on our tool. 'Pole Friction' is one of the possible attributes to use - which can be used to implement hindrance. Increasing Pole Friction will make it easier to land on the pole. A zero friction would technically make landing completely impossible - hence inducing highest level of hindrance. Researchers could use the visual material of the pole to provide a visual representation of what type friction or assistance or hindrance technique is used in a particular pole, perhaps investigate whether congruence is visual indications of friction and pole behaviour affect play experience.

- In all of our deployed experiments, we kept the trajectory paths hidden. However, this tool can reveal whether assistance or hindrance has been used or not and researchers might explore how a player responds to this disclosure, or how this might benefit skill development.

Chapter 10

Conclusion

10.1 Summary

In this thesis, we present a research tool that is capable of assessing player experience by modifying overt game objects like level difficulty (visible to player perception) and covert game objects like hidden assistance or hindrance (visually unnoticeable) and evaluated the tool with approximately 160 participants. We showed how player experience would look like when their in-game performance is manipulated both overtly and covertly. The tool has been made open source in an online repository. Any games user researcher can use this tool to ask research questions that are revolving around the experience of in-game success or failure, or extend on it and conduct experiments with new tool features.

The game avatar in this tool can be controlled with only one button, which makes it similar to a 2D casual platformer game. This genre of games require very little or no previous gaming expertise. This also removes the limitation of finding the right type of participants to evaluate a particular research idea. In the experiments reported in this thesis, participants played two or four rounds of game in a particular experimental session and reported on their game-play experience in a questionnaire. They were asked to rate their enjoyment, competence, effort, internality and overall perceived difficulty for the most recent game round. To evaluate player performance, the tool logged several game events that occurred during the game-play; e.g., pole positions and relevant assistance or hindrance attributes, game score, other caveman or pole attributes like caveman's jump power or pole friction level, and success or failure if there had been no assistance or hindrance applied. In the context of this thesis, we did not analyze all log information; other researchers might need to analyze these data to understand characteristics about players' psychology.

To investigate the efficacy of this research tool, we performed four studies on two different covert assistance techniques and explored the interactions with overt level difficulty. The efficacy of pole magnetism (that we found in Study 1) was reasonable, and it successfully manipulated most of the constructs. However, we found it to be less powerful than the overt manipulation and it also did not support the self-serving attribution bias about player performance. And due to the nature of this game feature (pole magnetism), many participants were able to notice it.

Based on the results and participant feedback, from our second study and onwards, we changed the assistance technique to trajectory manipulation, and this showed a greater effect of player experience manipulation. A follow-up study (Study 2B) with overt manipulation also established that the efficacy of this feature holds, regardless of how visually difficult the level was. In the final study (Study 3), we manipulated player experience by controlling assistance or hindrance rate at different parts of the game round and evaluated how player experience responds in terms of momentary success or failure in the ending moments of the game. We found that our tool also resembles the end-game boss battle, with players preferring the game round that ends with higher difficulty or hindrance.

10.2 Closing Thoughts

Well-established and publicly accessible research tools available to games researchers for manipulating player experience are limited in number. Games user researchers, who wish to investigate in-game success or failure and design experiments balancing mundane realism and external validity will benefit from using our research tool. Our tool manipulates the experience of success and failure in a game at a high resolution and allows researchers with or without technical expertise to ask and answer interesting questions in games research through controlled experiments.

References

[1] Sami Abuhamdeh and Mihaly Csikszentmihalyi. The Importance of Challenge for the Enjoyment of Intrinsically Motivated, Goal-Directed Activities. *Personality and Social Psychology Bulletin*, 38(3):317–330, mar 2012.

[2] Lorraine G. Allan. The perception of time. *Perception & Psychophysics*, 26(5):340–354, sep 1979.

[3] Aronson, E. R. and J. M Carlsmith. Experimentation in social psychology. In *Handbook of social psychology*. 1968.

[4] Alexander Baldwin, Daniel Johnson, Peta A. Wyeth, Alexander Baldwin, Daniel Johnson, and Peta A. Wyeth. The effect of multiplayer dynamic difficulty adjustment on the player experience of video games. In *Proceedings of the extended abstracts of the 32nd annual ACM conference on Human factors in computing systems - CHI EA '14*, pages 1489–1494, New York, New York, USA, 2014. ACM Press.

[5] Scott Bateman, Regan L. Mandryk, Tadeusz Stach, and Carl Gutwin. Target assistance for subtly balancing competitive play. In *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11*, page 2355, New York, NY, USA, 2011. ACM.

[6] Max V. Birk, Cheralyn Atkins, Jason T. Bowey, and Regan L. Mandryk. Fostering Intrinsic Motivation through Avatar Identification in Digital Games. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*, pages 2982–2995, New York, NY, USA, 2016. ACM.

[7] Max V. Birk, Regan L. Mandryk, and Cheralyn Atkins. The Motivational Push of Games: The Interplay of Intrinsic Motivation and External Rewards in Games for Training. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '16*, pages 291–303, New York, New York, USA, 2016. ACM Press.

[8] Max V. Birk, Dereck Toker, Regan L. Mandryk, and Cristina Conati. Modeling Motivation in a Social Network Game Using Player-Centric Traits and Personality Traits. In *User Modeling and User-Adapted Interaction (UMAP 2015)*, pages 18–30. Springer, Cham, 2015.

[9] Jim Blascovich, Jack Loomis, Andrew C. Beall, Kimberly R. Swinth, Crystal L. Hoyt, and Jeremy N. Bailenson. TARGET ARTICLE: Immersive Virtual Environment

Technology as a Methodological Tool for Social Psychology. *Psychological Inquiry*, 13(2):103–124, apr 2002.

[10] Blizzard Entertainment. Star Craft II: Wings of Liberty. Game[Micorsoft Windows, OSX]. July 27, 2010. Blizzard Entertainment, California, US., 2010.

[11] Jason T Bowey, Max V Birk, and Regan L Mandryk. Manipulating Leaderboards to Induce Player Experience. In *The ACM SIGCHI Annual Symposium on Computer-Human Interaction in Play (CHI PLAY'15)*, pages 115–120, London, United Kingdom, 2015. ACM Press.

[12] Jason T. Bowey, Ansgar E. Depping, and Regan L. Mandryk. Don't Talk Dirty to Me: How Sexist Beliefs Affect Experience in Sexist Games. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2017)*, New York, NY, USA, 2017. ACM.

[13] Nicholas D. Bowman and Ron Tamborini. Task demand and mood repair: The intervention potential of computer games. *New Media & Society*, 14(8):1339–1357, dec 2012.

[14] Evren Bozgeyikli, Andrew Raij, Srinivas Katkoori, and Rajiv Dubey. Point & Teleport Locomotion Technique for Virtual Reality. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '16*, pages 205–216, New York, NY, USA, 2016. ACM.

[15] Robert S. Brewer, Nervo Verdezoto, Thomas Holst, and Mia Kruse Rasmussen. Tough Shift: Exploring the Complexities of Shifting Residential Electricity Use Through a Casual Mobile Game. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '15*, pages 307–317, New York, NY, USA, 2015. ACM.

[16] Capcom Productions. Resident Evil 4. Game [Microsoft Windows]. (January 11 2005). Capcom, Chuo-ku, Osaka, Japan., 1983.

[17] CD Projeckt RED. The Witcher 3: Wild Hunt. Game [Microsoft Windows]. (May 19 2015). CD Projeckt, Warsaw, Poland., 1994.

[18] Jared E. Cechanowicz, Carl Gutwin, Scott Bateman, Regan L. Mandryk, and Ian Stavness. Improving player balancing in racing games. In *Proceedings of the first ACM SIGCHI annual symposium on Computer-human interaction in play - CHI PLAY '14*, pages 47–56, New York, NY, USA, 2014. ACM.

[19] Andy Cockburn, Philip Quinn, and Carl Gutwin. Examining the Peak-End Effects of Subjective Experience. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*, pages 357–366, New York, NY, USA, 2015. ACM.

[20] Kate Compton and Michael Mateas. Procedural level design for platform games, 2006.

[21] João P. Costa, Rina R. Wehbe, James Robb, and Lennart E. Nacke. Time's up. In *Proceedings of the First International Conference on Gameful Design, Research, and Applications - Gamification '13*, pages 26–33, New York, NY, USA, 2013. ACM.

[22] Creative Assembly. Halo. Game [Microsoft Windows]. (February 21 2017). Microsoft Studios, Redmond, Washington, US., 2017.

[23] Steve Dahlskog and Julian Togelius. Patterns and procedural content generation: revisiting Mario in world 1 level 1. In *Proceedings of the First Workshop on Design Patterns in Games - DPG '12*, pages 1–8, New York, New York, USA, 2012. ACM Press.

[24] Frederik De Grove, Johannes Breuer, Vivian Hsueh Hua Chen, Thorsten Quandt, Rabindra Ratan, and Jan Van Looy. Validating the Digital Games Motivation Scale for Comparative Research Between Countries. *Communication Research Reports*, 34(1):37–47, jan 2017.

[25] E L Deci, H Eghrari, B C Patrick, and D R Leone. Facilitating internalization: the self-determination theory perspective. *Journal of personality*, 62(1):119–42, mar 1994.

[26] Ansgar E. Depping and Regan L. Mandryk. Why is This Happening to Me? How Player Attribution can Broaden our Understanding of Player Experience. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2017)*, New York, NY, USA, 2017. ACM.

[27] Ansgar E. Depping, Regan L. Mandryk, Chengzhao Li, Carl Gutwin, and Rodrigo Vicencio-Moreira. How Disclosing Skill Assistance Affects Play Experience in a Multiplayer First-Person Shooter Game. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*, pages 3462–3472, New York, NY, USA, 2016. ACM.

[28] Amy M Do, Alexander V Rupert, and George Wolford. Evaluations of pleasurable experiences: the peak-end rule. *Psychonomic bulletin & review*, 15(1):96–8, feb 2008.

[29] Christopher Dring. More money is spent on games than movies and music combined, says ihs. *MCV*, June 2015. Retrieved June 2, 2016, from: http://www.mcvuk.com/news/read/more-money-is-spent-on-games-than-movies-and-music-combined-says-ihs/0151059.

[30] Paul M. Fitts. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, 47(6):381–391, 1954.

[31] Julian Frommel, Katja Rogers, Thomas Dreja, Julian Winterfeldt, Christian Hunger, Maximilian Bär, and Michael Weber. 2084 – Safe New World: Designing Ubiquitous Interactions. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '16*, pages 53–64, New York, NY, USA, 2016. ACM.

[32] FromSoftware. Dark Souls [Microsoft Windows]. (August 24 2012). Namco Bandai Games, Shinagawa, Tokyo, Japan., 2014.

[33] Gamasutra. *The Designer's Notebook: Difficulty Modes and Dynamic Difficulty Adjustment*. 2008.

[34] GameSoundCon. Video Games Bigger than the Movies? Don't be so certain.... — Conference on Composing Video Game Music and Sound Design, 2015.

[35] Kathrin M. Gerling, Regan L. Mandryk, and Michael R. Kalyn. Wheelchair-based game design for older adults. In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '13)*, pages Article 27, 8 pages, New York, NY, USA, 2013. ACM.

[36] Kathrin M. Gerling, Matthew Miller, Regan L. Mandryk, Max V. Birk, and Jan D. Smeddinck. Effects of balancing for physical abilities on player performance, experience and self-esteem in exergames. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14*, pages 2201–2210, New York, NY, USA, 2014. ACM.

[37] Janet Go, Rafael Ballagas, and Mirjana Spasojevic. Brothers and sisters at play: exploring game play with siblings. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work - CSCW '12*, page 739, New York, NY, USA, 2012. ACM.

[38] Carl Gutwin, Christianne Rooke, Andy Cockburn, Regan L. Mandryk, and Benjamin Lafreniere. Peak-End Effects on Player Experience in Casual Games. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*, pages 5608–5619, New York, NY, USA, 2016. ACM.

[39] Carl Gutwin, Rodrigo Vicencio-Moreira, and Regan L. Mandryk. Does Helping Hurt?: Aiming Assistance and Skill Development in a First-Person Shooter Game. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '16*, pages 338–349, New York, NY, USA, 2016. ACM.

[40] John Harris, Mark Hancock, and Stacey D. Scott. Leveraging Asymmetries in Multiplayer Games. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '16*, pages 350–361, New York, NY, USA, 2016. ACM.

[41] Brent Harrison and David L. Roberts. Analytics-driven dynamic game adaption for player retention in Scrabble. In *2013 IEEE Conference on Computational Inteligence in Games (CIG)*, pages 1–8. IEEE, aug 2013.

[42] Chris Harrison, Brian Amento, Stacey Kuznetsov, and Robert Bell. Rethinking the progress bar. In *Proceedings of the 20th annual ACM symposium on User interface software and technology - UIST '07*, page 115, New York, NY, USA, 2007. ACM.

[43] Chris Harrison, Zhiquan Yeo, and Scott E. Hudson. Faster progress bars. In *Proceedings of the 28th international conference on Human factors in computing systems - CHI '10*, page 1545, New York, NY, USA, 2010. ACM.

[44] Susan Harter. Pleasure Derived from Challenge and the Effects of Receiving Grades on Children's Difficulty Level Choices. *Child Development*, 49(3):788, sep 1978.

[45] Heinz Heckhausen. Achievement motivation and its constructs: A cognitive model. *Motivation and Emotion*, 1(4):283–329, dec 1977.

[46] Robert M. Hessling, Craig A. Anderson, and Daniel W. Russell. Attributional styles. In *Encyclopedia of psychological assessment*, pages 116–120. 2002.

[47] H W Hogan. A theoretical reconciliation of competing views of time perception. *The American journal of psychology*, 91(3):417–28, sep 1978.

[48] Robin Hunicke and Robin. The case for dynamic difficulty adjustment in games. In *Proceedings of the 2005 ACM SIGCHI International Conference on Advances in computer entertainment technology - ACE '05*, pages 429–433, New York, New York, USA, 2005. ACM.

[49] Ioanna Iacovides and Anna L. Cox. Moving Beyond Fun: Evaluating Serious Experience in Digital Games. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*, pages 2245–2254, New York, NY, USA, 2015. ACM.

[50] id Software. Doom. Game [Microsoft Windows]. (May 13 2016). Bethesda Softworks, Mesquite, Texas, U.S., 1991.

[51] IO Interactive. Hitman. Game [Microsoft Windows]. (March 11 2016). Square Enix, Shinjuku, Tokyo, Japan., 1998.

[52] Aaron Isaksen, Dan Gopstein, and Andy Nealen. Exploring Game Space Using Survival Analysis. In *Foundations of Digital Games*, 2015.

[53] Aaron Isaksen, Dan Gopstein, Julian Togelius, and Andy Nealen. Discovering Unique Game Variants. In *The Sixth International Conference on Computational Creativity, ICCC*, 2015.

[54] Charlene Jennett, Anna L. Cox, Paul Cairns, Samira Dhoparee, Andrew Epps, Tim Tijs, and Alison Walton. Measuring and defining the experience of immersion in games. *International Journal of Human-Computer Studies*, 66(9):641–661, 2008.

[55] Martin Jennings-Teats, Gillian Smith, and Noah Wardrip-Fruin. Polymorph: A Model for Dynamic Level Generation. In *Proceedings of the 2010 Workshop on Procedural Content Generation in Games - PCGames '10*, pages 1–4, New York, New York, USA, 2010. ACM Press.

[56] Colby Johanson and Regan L. Mandryk. Scaffolding Player Location Awareness through Audio Cues in First-Person Shooters. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*, pages 3450–3461, New York, NY, USA, 2016. ACM.

[57] Daniel Johnson and John Gardner. Personality, motivation and video games. In *Proceedings of the 22nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction - OZCHI '10*, pages 276–279, New York, NY, USA, 2010. ACM.

[58] Daniel Kahneman, Barbara L. Fredrickson, Charles A. Schreiber, and Donald A. Redelmeier. When More Pain Is Preferred to Less: Adding a Better End. *Psychological Science*, 4:401–405.

[59] Steve L. Kent. *The ultimate history of video games : from Pong to Pokemon and beyond : the story behind the craze that touched our lives and changed the world*. Three Rivers Press, 2001.

[60] Ketchapp Games. Spring Ninja. Game [Android]. (March 2015). Ketchapp, Paris, France., 2014.

[61] Mohammad M. Khajah, Brett D. Roads, Robert V. Lindsey, Yun-En Liu, and Michael C. Mozer. Designing Engaging Games Using Bayesian Optimization. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*, pages 5571–5582, New York, NY, USA, 2016. ACM.

[62] Paul. Kline. *The new psychometrics : science, psychology, and measurement*. Routledge, 1998.

[63] Raph. Koster. *A theory of fun for game design*. Paraglyph Press, Scottsdale, AZ :, 2005.

[64] Yubo Kou and Xinning Gui. Playing with strangers: Understanding Temporary Teams in League of Legends. In *Proceedings of the first ACM SIGCHI annual symposium on Computer-human interaction in play - CHI PLAY '14*, pages 161–169, New York, NY, USA, 2014. ACM.

[65] Julius Kuhl and Jurgen Beckmann. *Volition and personality : action versus state orientation*. Hogrefe & Huber Publishers, 1994.

[66] Petri Kuittinen. Rogue - Exploring the Dungeons of Doom (1980).

[67] Ichiro Lambe. Procedural Content Generation: Thinking With Modules, 2012.

[68] Thomas Langer, Rakesh Sarin, and Martin Weber. The retrospective evaluation of payment sequences: duration neglect and peak-and-end effects. *Journal of Economic Behavior & Organization*, 58(1):157–175, sep 2005.

[69] Derek Lomas, Kishan Patel, Jodi L. Forlizzi, and Kenneth R. Koedinger. Optimizing challenge in an educational game using large-scale design experiments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13*, page 89, New York, NY, USA, 2013. ACM Press.

[70] J. Derek Lomas, Kenneth Koedinger, Nirmal Patel, Sharan Shodhan, Nikhil Poonwala, and Jodi L. Forlizzi. Is Difficulty Overrated? In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*, pages 1028–1039, New York, NY, USA, 2017. ACM Press.

[71] Thomas W. Malone. Toward a Theory of Intrinsically Motivating Instruction*. *Cognitive Science*, 5(4):333–369, oct 1981.

[72] Winter Mason and Siddharth Suri. Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods*, 44(1):1–23, mar 2012.

[73] Edward McAuley, Terry Duncan, and Vance V. Tammen. Psychometric Properties of the Intrinsic Motivation Inventory in a Competitive Sport Setting: A Confirmatory Factor Analysis. *Research Quarterly for Exercise and Sport*, 60(1):48–58, mar 1989.

[74] Adam W. Meade and S. Bartholomew Craig. Identifying careless responses in survey data. *Psychological Methods*, 17(3):437–455, 2012.

[75] Elisa D. Mekler, Florian Brühlmann, Alexandre N. Tuch, and Klaus Opwis. Towards understanding the effects of individual gamification elements on intrinsic motivation and performance. *Computers in Human Behavior*, 71:525–534, 2017.

[76] Matthew K. Miller and Regan L. Mandryk. Differentiating in-Game Frustration from at-Game Frustration using Touch Pressure. In *Proceedings of the 2016 ACM on Interactive Surfaces and Spaces - ISS '16*, pages 225–234, New York, NY, USA, 2016. ACM.

[77] Mossmouth. No Spelunky. Game[Microsoft Windows]. (December 21, 2008). Mossmouth, LLC. 2008.

[78] David G. Myers. *Exploring social psychology*. McGraw-Hill Education, 2015.

[79] Lennart E. Nacke, Gregor G. McEwan, Carl Gutwin, and Regan L. Mandryk. "I'm just here to play games": social dynamics and sociality in an online game site. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work - CSCW '12*, page 549, New York, NY, USA, 2012. ACM Press.

[80] Bonnie A. Nardi. *My Life as a Night Elf Priest: An Anthropological Account of World of Warcraft*. University of Michigan Press, 2009.

[81] Javad Nasiry and Ioana Popescu. Dynamic Pricing with Loss-Averse Consumers and Peak-End Anchoring. *Operations Research*, 59(6):1361–1368, dec 2011.

[82] Nielsen. Video Games Score 5% of U.S. Household Entertainment Budget, 2010.

[83] Nielsen. U.S. Games 360 Report: 2017, 2017.

[84] Nintendo. Super Mario Bros. Game[Nintendo]. (September 13, 1985). Nintendo R&D4, Kyoto, Japan., 1985.

[85] Rita Orji. *Design For Behaviour Change: A Model-Driven Approach For Tailoring Persuasive Technologies. Ph.D. Dissertation.* PhD thesis, University of Saskatchewan, 2014.

[86] Randy J. Pagulayan, Kevin Keeker, Dennis Wixon, Ramon L. Romero, and Thomas Fuller. User-centered design in games. In *The human-computer interaction handbook*, pages 883–906. Lawrence Erlbaum Associates, 2003.

[87] Serge Petralito, Florian Brühlmann, Glena Iten, Elisa D Mekler, and Klaus Opwis. A Good Reason to Die: How Avatar Death and High Challenges Enable Positive Experiences. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2017)*, New York, NY, USA, 2017. ACM.

[88] Andrew K. Przybylski, C. Scott Rigby, and Richard M. Ryan. A motivational model of video game engagement. *Review of General Psychology*, 14(2):154–166, 2010.

[89] D A Redelmeier and D Kahneman. Patients' memories of painful medical treatments: real-time and retrospective evaluations of two minimally invasive procedures. *Pain*, 66(1):3–8, jul 1996.

[90] Donald A Redelmeier, Joel Katz, and Daniel Kahneman. Memories of colonoscopy: a randomized trial. *Pain*, 104(1-2):187–94, jul 2003.

[91] R. Ryan and E. Deci. Self-determination theory and the facilitation of intrinsic motivation. *American Psychologist*, 55(1):68–78, 2000.

[92] Richard M. Ryan and Edward L. Deci. Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. *Contemporary Educational Psychology*, 25(1):54–67, jan 2000.

[93] Richard M. Ryan, C. Scott Rigby, and Andrew Przybylski. The Motivational Pull of Video Games: A Self-Determination Theory Approach. *Motivation and Emotion*, 30(4):344–360, dec 2006.

[94] Jesse. Schell and Jesse. *The art of game design : a book of lenses*. Elsevier/Morgan Kaufmann, 2008.

[95] Richard A. Schmidt, Douglas E. Young, Stephan Swinnen, and Diane C. Shapiro. Summary knowledge of results for skill acquisition: support for the guidance hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(2:352), 1989.

[96] Selfdeterminationtheory.org. selfdeterminationtheory.org Intrinsic Motivation Inventory (IMI).

[97] Noor Shaker, Georgios N. Yannakakis, and Julian Togelius. Towards player-driven procedural content generation. In *Proceedings of the 9th conference on Computing Frontiers - CF '12*, page 237, New York, New York, USA, 2012. ACM Press.

[98] Mike Sheinin and Carl Gutwin. Quantifying Individual Differences, Skill Development, and Fatigue Effects in Small-Scale Exertion Interfaces. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '15*, pages 57–66, New York, NY, USA, 2015. ACM.

[99] Katharina Spiel, Sven Bertel, and Fares Kayali. "Not another Z piece!" Adaptive Difficulty in TETRIS. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2017)*, New York, NY, USA, 2017. ACM.

[100] Penelope Sweetser and Peta Wyeth. GameFlow. *Computers in Entertainment*, 3(3):3, jul 2005.

[101] Ron Tamborini, Nicholas David Bowman, Allison Eden, Matthew Grizzard, and Ashley Organ. Defining Media Enjoyment as the Satisfaction of Intrinsic Needs. *Journal of Communication*, 60(4):758–777, dec 2010.

[102] The Entertainment Software Association. Essential facts about the computer and video game industry. Technical report, 2016.

[103] Julian Togelius, Noor Shaker, and Mark J. Nelson. Introduction. pages 1–15. 2016.

[104] Sabine Trepte, Leonard Reinecke, and Katharina-Maria Behr. Avatar Creation and Video Game Enjoyment: Effects of Life-Satisfaction, Game Competitiveness, and Identification With the Avatar. *Journal of Media Psychology: Theories, Methods, and Applications*, 22(4):171–184, 2010.

[105] Wouter van den Hoogen, Karolien Poels, Wijnand IJsselsteijn, and Yvonne de Kort. Between Challenge and Defeat: Repeated Player-Death and Game Enjoyment. *Media Psychology*, 15(4):443–459, oct 2012.

[106] Kellie Vella, Daniel Johnson, and Leanne Hides. Playing Alone, Playing With Others: Differences in Player Experience and Indicators of Wellbeing. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '15*, pages 3–12, New York, NY, USA, 2015. ACM.

[107] Eduardo Velloso, Carl Oechsner, Katharina Sachmann, Markus Wirth, and Hans Gellersen. Arcade+: A Platform for Public Deployment and Evaluation of Multi-Modal Games. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '15*, pages 271–275, New York, NY, USA, 2015. ACM.

[108] Rodrigo Vicencio-Moreira, Regan L. Mandryk, and Carl Gutwin. Balancing multiplayer first-person shooter games using aiming assistance. In *2014 IEEE Games Media Entertainment*, pages 77–84. IEEE, oct 2014.

[109] Rodrigo Vicencio-Moreira, Regan L. Mandryk, and Carl Gutwin. Now You Can Compete With Anyone: Balancing Players of Different Skill Levels in a First-Person Shooter Game. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*, pages 2255–2264, New York, NY, USA, 2015. ACM.

[110] Rodrigo Vicencio-Moreira, Regan L. Mandryk, Carl Gutwin, Scott Bateman, Rodrigo Vicencio-Moreira, Regan L. Mandryk, Carl Gutwin, and Scott Bateman. The effectiveness (or lack thereof) of aim-assist techniques in first-person shooter games. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14*, pages 937–946, New York, NY, USA, 2014. ACM.

[111] David Watson and Lee Anna Clark. The PANAS-X: Manual for the Positive and Negative Affect Schedule -Expanded Form THE PANAS-X Manual for the Positive and Negative Affect Schedule -Expanded Form. *Ames: The University of Iowa.*, 1999.

[112] Rina R. Wehbe, Elisa D. Mekler, Mike Schaekermann, Edward Lank, and Lennart E. Nacke. Testing Incremental Difficulty Design in Platformer Games. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2017)*, New York, NY, USA, 2017. ACM.

[113] Bernard Weiner. A theory of motivation for some classroom experiences. *Journal of educational psychology*, 71(1):3–25, feb 1979.

[114] Bernard Weiner. Intrapersonal and Interpersonal Theories of Motivation from an Attributional Perspective. *Educational Psychology Review*, 12(1):1–14, 2000.

[115] Bernard Weiner and Bernard. An attributional theory of achievement motivation and emotion. *Psychological Review*, 92(4):548–573, 1985.

[116] Glenn R. Wichman. A Brief History of Rogue.

[117] Procedural Content Generation Wiki. What PCG is - Procedural Content Generation Wiki.

[118] G. N. Yannakakis and J. Togelius. Experience-Driven Procedural Content Generation. *IEEE Transactions on Affective Computing*, 2(3):147–161, jul 2011.

[119] Georgios N. Yannakakis and John Hallam. TOWARDS OPTIMIZING ENTERTAINMENT IN COMPUTER GAMES. *Applied Artificial Intelligence*, 21(10):933–971, nov 2007.

Appendix A

Forms

A.1 Study 1 Consent Form

Before proceeding, please read the following. You must give your consent to continue.

Title: Jumping Caveman: Understanding Player Experience.

Researcher(s):

- Rasam Bin Hossain, Graduate Student, *email:* rasam.bin.hossain@usask.ca
- Dr. Regan Mandryk, Associate Professor, *email:* regan@cs.usask.ca
  Department of Computer Science, University of Saskatchewan, 306-966-2327

Purpose(s) and Objective(s) of the Research: The purpose of this project is to better understand and model player experience.

Procedures:

- In this study you will be asked some questions based on your demographics, previous gaming history and a short survey that includes psychological questionnaire about your decisions and your personality. Following these questionnaire, we will present 4 short game rounds where each round lasts 3 minutes. After each round of gameplay you will be requested to answer a brief survey questionnaire to reflect on your game experience regarding the current round.
- This study will take approximately 30 minutes to complete.

Funded by: The Natural Sciences and Engineering Research Council of Canada (NSERC).

Potential Risks and Benefits: There are no known or anticipated risks to you by participating in this research. Your participation will help us to design games which aid novices in the learning of fundamental skills.

Confidentiality:

- Confidentiality will be maintained throughout the study. The entire process and data will be anonymized. Data will only be presented in the aggregate and any individual user comments will be anonymized prior to presentation in academic venues.
- Only the principal researcher and their research assistants will have access to the data to ensure that your confidentiality is protected.
- Storage of Data
  - Data (including survey and interview responses, logs of computer use, and videos of interaction) will be stored on a secure password-protected server for 7 years after data collection.
  - After 7 years, the data will be destroyed. Paper data will be shredded and digital data will be wiped from hard disks beyond any possibility for data recovery.

Right to Withdraw:

- Your participation is voluntary. You may withdraw from the research project for any reason, at any time without explanation.
- Should you wish to withdraw, you may do so at any point, and we will not use your data; we will destroy all records of your data.
- Your right to withdraw data from the study will apply until the data have been aggregated (one week after study completion). After this date, it is possible that some form of research dissemination will have already occurred and it may not be possible to withdraw your data

Follow up: To obtain results from the study, please contact Rasam Bin Hossain (rasam.bin.hossain@usask.ca).

Questions or Concerns:

- Contact the researcher(s) using the information at the top.
- This research project has been approved on ethical grounds by the University of Saskatchewan Research Ethics Board. Any questions regarding your rights as a participant may be addressed to that committee through the Research Ethics Office ethics.office@usask.ca (306) 966-2975. Out of town participants may call toll free (888) 966-2975.

A.2 Study 2 Consent Form

Before proceeding, please read the following. You must give your consent to continue.

Title: Jumping Caveman: Understanding Player Experience.

Researcher(s):

- Rasam Bin Hossain, Graduate Student, *email:* rasam.bin.hossain@usask.ca
- Dr. Regan Mandryk, Associate Professor, *email:* regan@cs.usask.ca
  Department of Computer Science, University of Saskatchewan, 306-966-2327

Purpose(s) and Objective(s) of the Research: The purpose of this project is to better understand and model player experience.

Procedures:

- In this study you will be asked some questions based on your demographics, previous gaming history and a short survey that includes psychological questionnaire about your decisions and your personality. Following these questionnaire, we will present 3 short game rounds where each round lasts 4 minutes. After each round of gameplay you will be requested to answer a brief survey questionnaire to reflect on your game experience regarding the current round.
- This study will take approximately 30 minutes to complete.

Funded by: The Natural Sciences and Engineering Research Council of Canada (NSERC).

Potential Risks and Benefits: There are no known or anticipated risks to you by participating in this research. Your participation will help us to design games which aid novices in the learning of fundamental skills.

Confidentiality:

- Confidentiality will be maintained throughout the study. The entire process and data will be anonymized. Data will only be presented in the aggregate and any individual user comments will be anonymized prior to presentation in academic venues.
- Only the principal researcher and their research assistants will have access to the data to ensure that your confidentiality is protected.
- Storage of Data
  - Data (including survey and interview responses, logs of computer use, and videos of interaction) will be stored on a secure password-protected server for 7 years after data collection.
  - After 7 years, the data will be destroyed. Paper data will be shredded and digital data will be wiped from hard disks beyond any possibility for data recovery.

Right to Withdraw:

- Your participation is voluntary. You may withdraw from the research project for any reason, at any time without explanation.
- Should you wish to withdraw, you may do so at any point, and we will not use your data; we will destroy all records of your data.
- Your right to withdraw data from the study will apply until the data have been aggregated (one week after study completion). After this date, it is possible that some form of research dissemination will have already occurred and it may not be possible to withdraw your data

Follow up: To obtain results from the study, please contact Rasam Bin Hossain (rasam.bin.hossain@usask.ca).

Questions or Concerns:

- Contact the researcher(s) using the information at the top.
- This research project has been approved on ethical grounds by the University of Saskatchewan Research Ethics Board. Any questions regarding your rights as a participant may be addressed to that committee through the Research Ethics Office ethics.office@usask.ca (306) 966-2975. Out of town participants may call toll free (888) 966-2975.

A.3 Study 2B Consent Form

Before proceeding, please read the following. You must give your consent to continue.

Title: Jumping Caveman: Understanding Player Experience.

Researcher(s):

- Rasam Bin Hossain, Graduate Student, *email:* rasam.bin.hossain@usask.ca
- Dr. Regan Mandryk, Associate Professor, *email:* regan@cs.usask.ca
  Department of Computer Science, University of Saskatchewan, 306-966-2327

Purpose(s) and Objective(s) of the Research: The purpose of this project is to better understand and model player experience.

Procedures:

- In this study you will be asked some questions based on your demographics, previous gaming history and a short survey that includes psychological questionnaire about your decisions and your personality. Following these questionnaire, we will present 4 short game rounds where each round lasts 3 minutes. After each round of gameplay you will be requested to answer a brief survey questionnaire to reflect on your game experience regarding the current round.
- This study will take approximately 30 minutes to complete.

Funded by: The Natural Sciences and Engineering Research Council of Canada (NSERC).

Potential Risks and Benefits: There are no known or anticipated risks to you by participating in this research. Your participation will help us to design games which aid novices in the learning of fundamental skills.

Confidentiality:

- Confidentiality will be maintained throughout the study. The entire process and data will be anonymized. Data will only be presented in the aggregate and any individual user comments will be anonymized prior to presentation in academic venues.
- Only the principal researcher and their research assistants will have access to the data to ensure that your confidentiality is protected.
- Storage of Data
  - Data (including survey and interview responses, logs of computer use, and videos of interaction) will be stored on a secure password-protected server for 7 years after data collection.
  - After 7 years, the data will be destroyed. Paper data will be shredded and digital data will be wiped from hard disks beyond any possibility for data recovery.

Right to Withdraw:

- Your participation is voluntary. You may withdraw from the research project for any reason, at any time without explanation.
- Should you wish to withdraw, you may do so at any point, and we will not use your data; we will destroy all records of your data.
- Your right to withdraw data from the study will apply until the data have been aggregated (one week after study completion). After this date, it is possible that some form of research dissemination will have already occurred and it may not be possible to withdraw your data

Follow up: To obtain results from the study, please contact Rasam Bin Hossain (rasam.bin.hossain@usask.ca).

Questions or Concerns:

- Contact the researcher(s) using the information at the top.
- This research project has been approved on ethical grounds by the University of Saskatchewan Research Ethics Board. Any questions regarding your rights as a participant may be addressed to that committee through the Research Ethics Office ethics.office@usask.ca (306) 966-2975. Out of town participants may call toll free (888) 966-2975.

A.4 Study 3 Consent Form

Before proceeding, please read the following. You must give your consent to continue.

Title: Jumping Caveman: Understanding Player Experience.

Researcher(s):

- Rasam Bin Hossain, Graduate Student, *email:* rasam.bin.hossain@usask.ca
- Dr. Regan Mandryk, Associate Professor, *email:* regan@cs.usask.ca
  Department of Computer Science, University of Saskatchewan, 306-966-2327

Purpose(s) and Objective(s) of the Research: The purpose of this project is to better understand and model player experience.

Procedures:

- In this study you will be asked to fill out questionnaires about your demographics and previous gaming history. Following these questionnaires, we will present 2 short game rounds which are expected to be finished within 5 to 10 minutes depending on your skill. After each round of gameplay you will be asked to answer a brief survey questionnaire to reflect on your game experience regarding the recent game round.
- This study will take no more than 30 minutes to complete.

Funded by: The Natural Sciences and Engineering Research Council of Canada (NSERC).

Potential Risks and Benefits: There are no known or anticipated risks to you by participating in this research. Your participation will help us to design games which aid novices in the learning of fundamental skills.

Confidentiality:

- Confidentiality will be maintained throughout the study. The entire process and data will be anonymized. Data will only be presented in the aggregate and any individual user comments will be anonymized prior to presentation in academic venues.
- Only the principal researcher and their research assistants will have access to the data to ensure that your confidentiality is protected.
- Storage of Data
  - Data (including survey and interview responses, logs of computer use, and videos of interaction) will be stored on a secure password-protected server for 7 years after data collection.
  - After 7 years, the data will be destroyed. Paper data will be shredded and digital data will be wiped from hard disks beyond any possibility for data recovery.

Right to Withdraw:

- Your participation is voluntary. You may withdraw from the research project for any reason, at any time without explanation.
- Should you wish to withdraw, you may do so at any point, and we will not use your data; we will destroy all records of your data.
- Your right to withdraw data from the study will apply until the data have been aggregated (one week after study completion). After this date, it is possible that some form of research dissemination will have already occurred and it may not be possible to withdraw your data

Follow up: To obtain results from the study, please contact Rasam Bin Hossain (rasam.bin.hossain@usask.ca).

Questions or Concerns:

- Contact the researcher(s) using the information at the top.
- This research project has been approved on ethical grounds by the University of Saskatchewan Research Ethics Board. Any questions regarding your rights as a participant may be addressed to that committee through the Research Ethics Office ethics.office@usask.ca (306) 966-2975. Out of town participants may call toll free (888) 966-2975.

Appendix B

Survey Questionnaire

B.1 Demographics Questionnaire

Please answer the following questions.

Whats your age?

25

Indicate your gender:

Male

What is the highest degree or level of school you have completed? If currently enrolled, mark the previous grade or highest degree received:

Master's degree

If you are a student, please indicate your subject:

Computer Science

Please indicate your employment status:

A student

Please indicate your marital status:

Single - never married

B.2 Previous Gaming Experience related Questionnaire

Please answer the following questions.

Please indicate how often (on average) you play games:

A few times per week

If you have played games in the past, please indicate how often you have played at peak times:

Every day

How much do you self-identify as a gamer on the following scale?:

Not at all — Gamer

How do you self-identify your expertise with games?:

Novice — Expert

Which is your dominant hand?

	Left	Right
	○	◉

Please indicate the genres that you enjoy playing:

- [x] Action
- [x] Platform games
- [x] First Person Shooter
- [ ] Beat 'em up
- [ ] Adventure
- [x] Role Playing Games
- [ ] Mass Multiplayer Role Playing Games (MMORPG)
- [ ] Simulation
- [ ] Vehicle simulation
- [ ] Strategy
- [ ] Music games
- [ ] Puzzle games
- [ ] Sport games
- [ ] Multiplayer Online Battle Arena (MOBA)
- [x] Casual games
- [ ] Different genre(s)

Please indicate on which devices you play:

- [x] Desktop (e.g. Windows, Linux, OS X, etc.)
- [x] Console (e.g. X-Box, Play Station, etc.)
- [ ] Mobile device (e.g. phone, tablet, PS Portable, etc.)
- [ ] Different device(s)

B.3 Post-Game Questionnaire

Reflect on your play experiences and rate your agreement with the following statements.

	Strongly disagree	Disagree	Slightly disagree	Neutral	Slightly agree	Agree	Strongly agree
I enjoyed this game round very much	○	○	○	○	○	◉	○
How well I did in this game round was completely due to me	○	○	○	○	○	◉	○
I would describe this game as very interesting	○	○	○	○	◉	○	○
While playing this round, I was thinking about how much I enjoyed it	○	○	○	○	○	◉	○
It was important to me to do well at this game	○	○	○	○	○	◉	○
My ability to play this game round is well matched with the game's challenges	○	○	○	○	◉	○	○
I put a lot of effort into this game round	○	○	○	○	○	◉	○
In this game round, my performance was determined by my abilities	○	○	○	○	◉	○	○
This game round was difficult	○	○	○	○	○	◉	○
The reasons underlying my performance in this game lie within me	○	○	○	○	○	◉	○
I didn't try very hard at playing this round	○	◉	○	○	○	○	○
I feel very capable and effective when playing	○	○	○	○	○	◉	○
Playing the game round was fun	○	○	○	○	○	○	◉
This game round did not hold my attention	◉	○	○	○	○	○	○
My effort determined how well I did in this game round	○	○	○	○	○	◉	○
I feel competent at this game round	○	○	○	○	◉	○	○
I tried very hard while playing the game	○	○	○	○	○	◉	○