Balancing Randomness and Predictability in Games

Manuel Mendes Parracho Correia

Thesis to obtain the Master of Science Degree in

Information Systems and Computer Engineering

Supervisor: Prof. Carlos António Roque Martinho

Examination Committee

Chairperson: Prof. Alberto Manuel Rodrigues da Silva
Supervisor: Prof. Carlos António Roque Martinho
Member of the Committee: Prof. Rui Filipe Fernandes Prada

November 2021

This work was created using LaTeX typesetting language
in the Overleaf environment (www.overleaf.com).

Acknowledgments

To start, I want to acknowledge Prof. Carlos Martinho, my advisor, for all the valuable guidance and support. At every juncture, he was available to assist me by sharing his experience and perspective. He was invaluable to this work.

On a personal note, I would like to state my gratitude to my family and friends, all of whom helped in all the ways available to them. From the camaraderie of friends to the care and kindnesses on the day-to-day from family. Just as all the academic support was needed, this moral support too was essential.

Lastly, I would like to thank all those who took their time to participate in the experiments of this thesis. Both friends and members of online Gungeon communities alike provided the data, which made this work possible.

Thank you to one and all.

Abstract

Randomness is a useful tool in game design as a source of excitement and novelty. Nevertheless, it is often troublesome to balance. Overdo it, and the player will feel no control. Else, if it is minimal, it may be redundant. This work hypothesizes that a dynamic adjustment of randomness can positively impact the player experience (PX). First, we analyze previous works on the human perception of randomness and how it shapes game design. Several use cases are reviewed to understand how it informs the PX, and how to balance that. There is also an investigation on how to evaluate the potential changes to these aspects. Second, we formulate a model for this adjustment based on the player and intervening entities' resources throughout a playthrough. The selected case study was boss battles, challenges often difficult to balance given player disparities. With Enter the Gungeon as a testbed game, the decision-making behavior and respective action functionality involving random number generation (RNG) of the adversary were modified. Participants played against three versions: one unpredictable, one predictable, and one dynamic; then reported on their experiences. On the reported aspects, it was shown that players felt more competent with predictable behavior, significantly more than with an unpredictable one. Similarly, the dynamic behavior made participants feel more capable of growth. Moreover, there were indications that players who perceived themselves in control, according to a Locus of Control assessment, preferred predictable experiences. On the other hand, this mindset seemed connected to more effort put into their performance.

Keywords

Randomness, Dynamic Adjustment, Balancing, Digital Games

Resumo

Aleatoriedade é uma ferramenta útil em design de jogos, dado que fumenta divertimento e espontaneidade. Mas, é frequentemente problemática de balancear. Em excesso, o jogador não se sente em controlo. Quando é mínima, poderá ser redundante. Este trabalho propõem que o ajuste dinâmico de aleatoriedade terá um impacto positivo na experiência do jogador (PX). Primeiro, analiso trabalhos prévios sobre a percepção humana de aleatoriedade e como informa o design de jogos. Múltiplos exemplos são examinados para entender como influência a experiência dos jogadores, e como balanceá-la. Ainda, investigo métodos para avaliar o potencial impacto. Segundo, formulo um modelo deste ajuste dinâmico baseado nos recursos do jogador e entidades relevantes durante a sessão de jogo. Como caso de teste foram selecionados confrontos com bosses, estes são geralmente difíceis de balancear dadas as diferenças entre jogadores. Com Enter the Gungeon como jogo de teste, o processo de decisão do adversário e respectivas ações envolvendo aleatoriedade foram modificadas. Os participantes jogaram contra esta versão, e duas outras: uma imprevisível e outra previsível; reportarando depois as suas experiências. Dos dados recolhidos, demonstrou-se que jogadores se sentiram mais competentes com comportamento previsível, significantemente mais do que com imprevisível. Similarmente, o comportamento dinâmico fez os participantes sentir uma maior capacidade para melhorar. Mais, houve indicações de que jogadores que acreditam estar em controlo, com base no Locus de Controlo, preferiram previsibilidade. Por outro lado, estes pareceram também esforçar-se mais.

Palavras Chave

Aleatoriedade; Ajuste Dinâmico; Balanceamento; Jogos;

Contents

List of Figures

List of Tables

Acronyms

AI	Artificial Inteligence
DDA	Dynamic Difficulty Adjustment
DLL	Dynamic-Link Library
DnD	Dungeons and Dragons
EEG	Electroencephalography
EtG	Enter the Gungeon
FFXIV	Final Fantasy XIV
GEQ	Game Engagement Questionnaire
LoC	Locus of Control
IMI	Intrinsic Motivation Inventory
MMO	Massive Multiplayer Online
MtG	Mod the Gungeon
NPC	Non-Player Character
PX	Player Experience
QA	Quality Assurance Testing
RNG	Random Number Generation
RPG	Role-Playing Game
TCG	Trading Card Game
UCD	User-Centered Design

1

Introduction

Contents

1.1 Motivation

Throughout video games, chance plays a pivotal role in defining the user's experience. Unpredictability can both create replayability and excitement when appropriately employed. Likewise, removing chance in games that hinge on deterministic behavior is essential. However, some scenarios are not clear cut, for instance, when randomness plays a part in the level of the challenge the player faces. Unpredictability can be a source of difficulty beyond the player's control and, thus, a point of frustration. Entirely predictable behavior, on the flip side, may feel repetitive. And, different players, based on their expertise, have different definitions of what constitutes an appropriate hurdle. For these reasons, reaching a static implementation that suits a broad player base is not simple or, perhaps, even possible.

1.2 Problem

As video games have become an ever more ubiquitous form of entertainment, balancing the Player Experience (PX) has, too, grown in complexity.

As a game mechanic, randomness can impact several aspects of the PX, such as excitement and difficulty. With large or growing player bases, a static solution will struggle by not suiting all players and, in trying to, may please none. Beyond this, randomness can create highly variable gameplay experiences that incentivize replayability. As a result of this design, players grow to play significantly differently throughout their time with the game, invalidating existing balance. Here, the consequence of static solutions is repetitiveness, which is antithetical to the overall design.

As such, we will address the problem of effectively balancing randomness to create a PX uniformly satisfactory for a growing audience, with increasingly diverse players.

1.3 Hypothesis

The player resources can be indicative of performance, which can inform the dynamic adjustment of randomness. This process can take the PX from pure chaos to rigid patterns, potentially improving it.

When facing a more powerful adversarial entity in the game, a player bridges the power gap through strategy and mastery of the controls. These challenges, referred to as 'boss fights,' are used as a test to conclude a level. Throughout these confrontations, either the player or enemy will get the upper hand, and a comparison of the player resources and the enemy resources can determine which (e.g.: health percentage). That can then be the basis for adjusting the behavior of the player's adversary. If the player has substantially more resources than the enemy, the adversary's behavior can become unpredictable, limiting the player's ability to strategize. Or, were the enemy to be the one with the most resources, they would adjust to follow clearer patterns, giving the player more breathing room. This responsiveness should result in changes both to the player's reported perception of the gameplay and the logged metrics during the playthrough. In turn, this data should allow for an evaluation of the impact of the manipulation of randomness.

With this, we hypothesize that by dynamically adjusting the predictability of the adversary's behavior based on the resources of both entities in real-time, the PX can be improved by better suiting each player.

1.4 Contributions

This thesis makes varied contributions, from theoretical to practical. To begin, it includes a review of the literature on randomness as a concept, its use within games, its impact on the player experience, and how that impact can be measured. Then, there is the examination of a model for the dynamic adjustment of randomness for the improvement of its balance across players and player evolution. Alongside this, there is an overview of the implementation within a testbed game and its technical particularities. Lastly, through telemetry within this implementation and a developed questionnaire about the PX, experimental validation is carried out. The results of which are presented and discussed.

1.5 Document outline

This work is split into five chapters, the introduction, related work, the case study, its evaluation, and the conclusion.

First, the introduction gives a brief exposition of this project's topic, hypotheses, and contributions. Following that, the related work examines existing research on the topics of randomness, game design, player experience, and its evaluation. The third chapter describes the case study. This chapter starts with an overview of the test-bed game, followed by this work's model and its implementation, along with the necessary tools. Chapter four presents the experiments of this thesis, from the preliminary trials to construct and validate the evaluation tools, to the main experiment to collect data on the implementation and ending with the data analysis. The final chapter discusses this work's conclusions and presents potential future work stemming from our findings.

At the end of the document, there are two bibliographies, the second of which includes all games referenced within this work with brief descriptions of their respective gameplays. Following these, there is an appendix comprised of the questionnaire developed for the experiment of this thesis.

2

Related Work

Contents

This chapter will go over concepts related to this work, as well as similar works previously published with pertinent conclusions. The concept of randomness will be explored, as well as how humans relate to it and how it is used within the context of game design.

2.1 The human perception of randomness

2.1.1 Randomness

This is a complex concept that often eludes us, but it can be central to many aspects of our lives and is an important tool in crafting experiences in games.

First, randomness is often described as noise, a truly random sequence cannot be predicted. Hence why most of the randomness we utilize in Computer Science is pseudo-randomness, not the actual thing. True randomness is very difficult to achieve, so, we resort to techniques that produce results that are outstandingly hard to predict. Nevertheless, though it would be impossible for us to observe the patterns left behind by those mechanisms, powerful computers can. Interestingly, the reverse can also be true, as we are prone to classifying unpredictability as randomness. Whether it is a lack of knowledge, experience, or both, in a given situation, it is not uncommon for a person to believe an outcome random, when it could have been predicted if only they had all the necessary tools. Within this work, we will adhere to Burgun's [1] definition of randomness:

> "[The] information that enters the game state which is not supposed to ever be predictable".

With this, randomness differs fundamentally from unpredictability based on the intent of the designer. Unpredictable is that which the player does not have enough experience or information to predict, the more play-time they accrue, the better their understanding will be.

Furthermore, randomness is a tool that can be employed in different ways based on the desired impact on the experience. Burgun outlines the distinction between "output" and "input" randomness:

- Output Randomness - "noise injected between the player's decision and the outcome" (e.g.: rolling a die to determine if an attack hits the target);
- Input Randomness - "informs the player before he makes his decision" (e.g.: procedural level generation).

It is important to underline this distinction as the two forms have different use cases and impact on the PX. In this work, we will focus on output randomness.

It is worth mentioning 'controlled randomness' as well. Though the notion might seem contradictory at first, it is crucial when designing games, because of how the human perception of randomness is

[1] K. Burgun, Randomness and Game Design. https://www.gamasutra.com/blogs/\KeithBurgun/20141015/227740/. Last accessed 6 Nov 2020

skewed. Khandaker [2] argues that our cognitive biases ultimately make true randomness seem unfair to us. For that reason, it is not only important to pick the right source for randomness (e.g.: rolling a 6-sided die or picking a card from a deck of 6 numbered cards is both mechanically and mathematically different), but also to chose when to skew that randomness in the interest of fairness and a better experience. One such situation is Tetris' [14] generation of blocks. Most likely, the majority of people would say it is random when asked, it does fall into the definition we established. Nevertheless, it is an example of 'controlled randomness', as by design the player does not receive the same piece an inordinate amount of times in a row, nor do other such situations that might detract from the PX occur.

2.1.2 The Perception of Randomness

As human beings, we have millions of years of evolution that have conditioned us with the goal of ensuring our survival, and yet, there are situations in which this evolution can hinder us.

For one, we must understand how it is that people perceive randomness. Many studies have delved into this and have sought to gain an understanding through two main tools:

- Production tasks - in which a subject is asked, for example, to produce a random sequence of numbers or symbols from a given set;
- Judgement tasks - which center around discerning which of similar example outputs was randomly generated.

Through the analysis of many previous studies, Bar-Hillel et al [1], were able to consolidate conclusions and cement understanding on this topic.

From their research, it can be stated that humans overestimate alternation in randomness. In the case of binary sequences, as may be generated through coin tosses, any 'long' sequence of repeated results ("Gambler's Fallacy") or observed symmetry will be taken as evidence of non-random generation. Likewise, in production tasks, it was observed that people attempt to avoid such streaks or patterns, though they would be found with true random generation. When tested, researchers observed that humans are most likely to perceive as random and generate to be random, binary sequences that have an alternation rate of 0.6. Nevertheless, they point out that despite being a systematical difference, this incongruity between our inherent notion of randomness and the reality is not a radical difference (in a binary sequence with 0.6 alternation rate, the next entry can be accurately predicted 52% of the time, whereas one with a 0.5 rate would allow for a correct prediction 50% of the time).

Given the above result, the study concludes that how we perceive randomness is paradoxical with the concept. For us to view a sequence as random, it cannot afford to be truly random. This is compatible

[2] M. Khandaker, Games, Randomness and the problem with being Human. https://www.gamasutra.com/view/news/124357/. Last accessed 7 Nov 2020

with the adopted definition, though. What is important is not that the mathematical characteristics of randomness are observed but that the outcome is intentionally unpredictable. For this reason, player perception is more significant.

Despite this, it is of note that the idea that humans are not well suited to perceiving or producing randomness may not be as well-founded as some claim. Nickerson [2] in his paper points to several explanations as to the observed discrepancies between reality and human perception of the concept. The researcher points out that several factors, such as unclear explanations of both production and judgment tasks, lack of concrete understanding of the concept by the researchers themselves, or pre-conceived notions of the participants might have skewed results. This is worth mentioning as it means players may detect emergent patterns, those not intentionally introduced. Which, he proposes, can be especially true when the circumstances are compatible with those of scenarios we have evolved to handle.

2.1.3 Randomness and Game Design

A game is an experience built from the ground by a team of individuals, for this reason, it is different from everyday life in that the laws of its world have a discernible purpose.

As discussed before, the way randomness is introduced can greatly impact its reception. In his article about “The Illusion of Randomness in Games”, Compton [3] , gives a brief analysis of how different mechanisms of randomness in board games impact the players. Of particular interest, dice are often seen as less trustworthy, perhaps because of their association with gambling, but, beyond that, Compton gives the example of a failed Star Wars Trading Card Game (TCG). He posits that the reason for its failure is that players felt they lacked agency. Cards representing entities were placed on the board and could attack opposing cards, to do so, dice had to be rolled, namely, for damage calculation. And, as he explains it, the stronger the card the more dice had to be rolled, at times more than ten. Players said this made the game feel ‘swingy’, with so many dice, the result felt incredibly random. This perception is mistaken, because in reality, the more dice involved, the more the distribution of results approximated a normal distribution. This can easily be reasoned by considering specific examples, such as how many possibilities there are of obtaining a 7 compared to a 2 when rolling two dice.

On the other hand, cards are pointed to as a more appreciated source of randomness. In the article, the author argues that this might be associated with the increased feeling of control in some card games, which can be a result of the player building their deck, or of all players drawing their hand at the start and playing from there. This point arcs back to the input versus output randomness, player’s see input randomness as preferable, perhaps fairer. They all start from uncertainty, and make the best of the cards

[3] C. Compton, Take a Chance: The Illusion of Randomness in Games. https://remptongames.com/2017/09/23/take-a-chance-the-illusion-of-randomness-in-games/. Last accessed 21 Nov 2020

they were dealt, instead of seeing how the best decision they could make still leads to failure because of a roll that did not go well. Of course, when drawing cards, a player may also be unlucky; however, games that involve deck building appear to mitigate feelings of frustration resulting from bad draws.

Moreover, in Schell's [3] book about game design, chance is one of the six main game mechanics explored. As he frames it, chance is the broadest of mechanics as it intertwines with all others in some way and is an essential part of the fun for the uncertainty it creates. He goes so far as to say that surprises are the secret ingredient of fun. And so, he too defends that the designer must have a good mathematical understanding of randomness along with a grasp of the player's perception of it. This is because, in facing chance, the player will have to determine the 'expected value' of the actions available. The value of an action is commensurate to the probability that it will have the intended result. In turn, the player's perception is relevant to this because it can skew their calculations of the expected value(s) and influence their actions.

With this in mind, we must also mention how skill factors into the interaction. Schell [3] points out that deducing/estimating the chance of a given outcome is a skill that players must nurture. One must also keep in mind that actions such as parrying a blow in a Role-Playing Game (RPG) have an inherent chance of success, informed by the player's ability to execute that action and also the adversary's ability to thwart the attempt. The estimation of the latter also a relevant skill. Hence, why in some tabletop games, there are dice rolls to determine whether an attack hits, or how much damage it does. This randomness is a proxy to the expansive and harder to quantify factors that would dictate the outcome in such adversarial interactions. (e.g.: Strength and direction of the wind, when shooting an arrow.)

On the matter of skill, Cook [4] focuses on how randomness is noise and through practice, users develop the tools to navigate it. From his perspective, if it is noise, then it has an underlying model, first unknown to the player. This model has a learning curve and, likely, variables that affect it, which if recognized can help understand it. And so, like Schell, he believes that there are skills that players can develop to handle these scenarios. Cook makes mention of some of the same mathematical notions as Schell and seems to agree on how important it is for both designer and player to understand them. Beyond those, realizing the value of sampling and being able to discern the cost/benefit relation in each attempt are also noted. As humans, when faced with new situations and deciding how to act, we leverage whatever previous experience we have to optimize our behavior, but often we will not have adequate existing heuristics.

In summation, randomness is pivotal to game design and needs to be carefully inserted in the crafted experience, as it has broad impact. As Schell and Cook laid out, the player will cultivate the skills to handle it through playing. And it is the designers' job to decide how it affects the game, where skill is meant to be more impactful or where it will have little bearing. From situations where mathematics and

[4] D. Cook, Understanding Randomness in terms of Mastery. https://lostgarden.home.blog/2012/12/31/understanding-randomness-in-terms-of-mastery/. Last accessed 23 Nov 2020

reasoning lead to the best answer, provided the player has modeled the scenario, to others in which they will have to decide on taking a risk. From unpredictability to randomness.

2.2 Balancing and Randomness in Games

Randomness is a versatile and powerful tool used to shape many different facets of games, impacting many genres, both positively and negatively.

2.2.1 Balancing for different players

Digital games cater to many different people, this is of course narrowed down when discussing a specific genre or, even more so, a particular game, but amongst those, there is still variation. These differences between players must be considered and are the reason for difficulty and accessibility settings. However, the most common solutions do not effectively tackle this balancing problem as it is increasingly difficult to adapt a game, during development, to a wider, more disparate, player base.

Presently, the work of balancing a game is a highly iterative process, which involves not just the game developers, but also play-testers. As is common in many fields, Quality Assurance Testing (QA) provides important data for fine-tuning the PX. For this reason, the group of people participating in these tests needs to be diverse, in the hopes of achieving a widely satisfactory experience. And, though at first glance video games might seem very different in goals and design, from other software applications, the importance of User-Centered Design (UCD) cannot be overstated [4].

To include players in the development process is proven to have a positive impact on the game's design and as seen in other fields, validating designers' intuitions through testing with users is just as helpful. However, it is increasingly clear that these methodologies cannot fully address these issues and so other solutions are being studied. Dynamic Difficulty Adjustment (DDA), is one such case. The idea is that beyond the balancing done through testing amidst development, developers would integrate systems into their games, capable of dynamically adjusting aspects of it to the current player. H. Fernandez et al. [5] developed and tested a system that used player performance and Electroencephalography (EEG) to adapt the game's procedural generation to change its difficulty to match the users perceived skill. Their results indicated the system was successful in adapting the PX and they proposed that this could be applied to genres other than platformers or adapted for purposes other than procedural generation. They noted, however; that given the use of EEG technology, their approach would only be feasible in a playtesting context to further help to balance.

Continuing on the topic of DDA, solutions solely based on player performance and, thus, feasible for application in released games, have also been explored. The adaptability is the most important factor, even simple alterations can affect the experience positively. A. Denisova et al. [6] studied the changes perceived by users when dynamically adjusting the challenge they played through based on their performance. The researchers set out to explore if such a change would rob participants of the sense of challenge, instead, they found that it made for a more immersive experience. In the experiment they devised, a shooter type of game was played and the goal was to score 300 points or more within 90 seconds. Players were divided into two groups: a control group playing the non-adaptive experience and another where the timer would speed up or slow down based on whether the player was doing better than average(established from the pilot study) or worse, respectively. All participants were unaware of the mechanism. Beyond the already mentioned increase of immersion, it was also observed that there was an improvement of performance in 'good' players from the experimental group in comparison to those in the control group. This observation is speculated to be due to the increased pressure of the shorter time limit, despite the fact participants were not aware of it.

An adequate adaptive balancing solution is, evidently, a promising avenue in preparing a game for a more diverse audience. Artificial Inteligence (AI) is a powerful tool in similar scenarios, but as seen above, even simple adjustments can have a meaningful impact on the PX. Players are often blind to the changes made and retain their sense of agency [7].

2.2.2 Input randomness in games

When randomness shapes, not the unfolding of interactions between player and game, but rather the context of those interactions, it is labeled input randomness.

One frequent use of this type of randomness is procedural generation, a tool developers have relied upon to create changing and re-playable experiences. Simply, it is the creation of procedures to be carried out by the game engine. A controlled form of randomness used in generating environments, and many other scenarios, using prepared assets.

Most famously, Minecraft [15] uses one such mechanism to create voxel worlds for players to explore and transform. When playing in its original mode (Survival Mode), the player starts near the origin point of the world's generation and can find themselves in a myriad of different biomes. This world grows as the player explores it, it is formed according to the procedural generation which not only decides if next, they will find an ocean or desert but also what is found within that area. Figure 2.1 shows the generated area of a Minecraft world, each color representing a different biome and the red arrow, the player.

In a more simplistic, and yet effective, approach, the indie game The Binding of Isaac [16] uses procedural generation to make its levels. A universal characteristic of rogue-likes, each level is built from pre-made assets, from mere objects to entire rooms, that can be arranged in many layouts (see Fig. 2.2). This incentivizes exploration and plays a big part in the re-playability of this genre. The controlled randomness allows for an ever-changing experience that is, nevertheless, consistent.

Though its applications in digital games are impressive, input randomness has been present in

Figure 2.1: A Minecraft world's map of the generated area.

Figure 2.2: The generated layout of a level in The Binding of Isaac.

games for centuries. Looking at card games, we can readily draw parallels to the applications mentioned above. Shuffling a deck of cards, we introduce the input randomness into whichever game is to be played, much like Isaac has its rooms, a French-suited deck has its four suites, each with its 13 cards. Impressively, it is often overlooked just how many different combinations those 52 cards can make, a number so large it is highly likely that each time a deck is shuffled, that order has never before occurred.

2.2.3 Output randomness in games

Output randomness, in contrast, informs how interactions unfold between the player and the game.

Perhaps one of the first iterations of the concept, randomness has long been used in Role-Playing Games (RPGs) to determine the results of actions. In table-top games, there is the widely known example of Dungeons and Dragons (DnD) [17], a game the relies heavily on dice to simulate the variability present in a player's actions. This mechanism can determine everything from how many projectiles an attack includes, which of those hit and which miss, and then, even how much damage each that hits its target does. But RPGs are games that most often reproduce many aspects of our lives, seeking to transport us to the worlds they create, this means that combat is far from the only interaction to simulate. Here too, DnD has proved their methodology successful as players can train even their charm or eloquence, and the same mechanisms that shape combat encounters, shape conversations.

More recently, Disco Elysium [18] used this successful formula and applied it in what has been considered a great success. Many other games have taken inspiration from the foundation DnD set, indeed, the RPG genre is full of such examples. And yet, it can be said that with the evolution of digital games these have become incredibly complex, with many overlaying systems that may cloud the impact of those roots. Disco Elysium is a novel reminder of how this randomness can be effectively utilized to create excitement, build anticipation, and shape narrative. All the while, this game does not take away from the player's sense of control or impact in the game's environment. With little in the way of combat, the game allows players to (re)build the personality of an amnesiac main-character and the traits they invest in, and the thoughts they pursue. It is the points invested and gained for these traits that determine the likelihood of success in many interactions. In a conversation, a very persuasive individual is more likely to succeed in, for example, convincing another to share a secret, but no matter their cunning, external circumstance can lead them to failure (see Fig. 2.3). It is in applying randomness to these interactions, grounded in such rationale, that both of the above examples manage to satisfy players.

Figure 2.3: A failed skill check in Disco Elysium.

A similar application can be found in Strategy games. In the XCOM [19] franchise, player's are tasked with coordinating the actions of several characters each turn, to advance into enemy territory and either wipe-out the adversaries or some other similar goal. Here too, many actions have an outcome determined by chance (see Fig. 2.4). Players are informed of the probability of success, which is determined by factors such as visibility to the target when aiming a shot or the suitability of their weaponry in face of the target's defenses. Lacking the certainty of the moves in a game of chance, players can become frustrated as outcomes do not go their way, even though probabilities might have been in their favor. If with one character, one the player identifies with, they can justify the uncertainties faced just like the ones we face daily. When orchestrating the behavior of several, a few bad outcomes can quickly create a feeling of impotence.

Another use of output randomness lies in rewards. Just as procedural generation creates re-playability by varying the content played or at least its order, varying the reward of an action or activity gives the player motivation to engage in it more often.

Games in many different genres rely on controlled randomness to reward players, RPGs are no exception. However, one of, if not the most, prominent case is that of Looter Shooters. Games like the Borderlands [20] entice players with the possibility of obtaining more powerful weapons or armor, which allows them to face stronger more challenging foes or tasks, that in turn reward them with resources to further improve themselves. Simply put, these games create a compulsion loop [5] that grabs players.

[5] Compulsion Loop (or Core Loop) - a habitual chain of activities which by being repeated leads to the perpetuation of the loop.

Figure 2.4: XCOM action success and critical hit probabilities.

The introduction of variability to this type of content means that users will often not get the reward they sought and must replay it. Interestingly, Massive Multiplayer Online (MMO) RPGs can add a further layer of randomness to the process of collecting loot, this time, in the interest of fairness. In these games, such as Final Fantasy XIV (FFXIV) [21], the fact that most endeavors are carried out by a group of players, raises the question of how to fairly split it. Beyond just the question of whether the specific item a player wants appearing as a reward or not, under the multiplayer scenario, there is the question of whether that player will acquire it. In FFXIV, this is resolved through randomness, there is a system in place that lets players choose if they need it (only if that item is usable by their character/class), if they want it (Greed) or do not want it (Pass). The first two options will then lead to the random generation of a number between 1 and 100, with Need tending towards higher results (see Fig. 2.5). Whoever lands the highest value gets the item, the intention being that players feel it is fair, as everyone is bound to the same rules, instead of relying on trust and attempting to divide the spoils of whatever challenge was faced with people you do not know.

Returning to rogue-likes, this genre also utilizes randomness to control what items or consumables players obtain. But, the inherent nature of this genre, in which a run [6] is played and at its end, the player state is reset, the motivation is different. Originally, after dying or completing a run, within this genre, a player could then begin another with the previous having no influence on future ones. Nowadays, many games have implemented mechanics that introduce an overarching progression. Nevertheless, in most cases, within a given run, the rewards a player obtains are randomized and this is because they are intended to inform that instance of the experience. The PX is largely informed by the environment the player finds themselves in and how they can impact it, the former, in rogue-likes as covered, varies

[6] Run - a term used to refer, in games of the rogue-like/rogue-light genre, to an attempt at clearing the game's re-playable journey.

Figure 2.5: FFXIV reward window. Items with 0 the player chose to 'Pass' on. The others, either 'Need' or 'Greed' was selected.

through the use of procedural generation and the latter through output randomness. Just as they have many rooms to populate their levels, rogue-likes have many weapons/items/powers a player can obtain to navigate through the challenge the game presents, each contributing to a different experience than last time.

The last instance we will analyze of output randomness, still regarding rewards, is the use of Loot Boxes. Although they are not significantly different in the context of randomness from other reward systems, just like the two previous examples, they differ in intention and, thus, we would be remiss in omitting them. Loot Boxes are a mechanism for monetization that offers players, through purchase, a chance at rewards that may only be cosmetic or that can impact gameplay too. This system takes advantage of the same mentality associated with slot machines and in some sense as introduced in Looter Shooters. The important distinction being that in that other genre, the reward variability is meant to encourage continued play and here, much like in casinos, it seeks to increase revenue by having players make repeated attempts at obtaining the result they desire.

2.3 Evaluating the game experience

In this project, we set out to study the balancing of randomness within games, and so, any experiment's goal must be the evaluation of the player experience as altered by the proposed solution. Previous cited works [4–6] have shown that balancing a game is essential as it shapes the player's experience and, most importantly, their enjoyment.

For several decades, the theory of Flow [8] has seen near-ubiquitous adherence in the study of enjoyment. M. Csikszentmihalyi, with his proposal of this theory, enumerated several fundamental aspects of enjoyment. Flow has since been utilized in many areas of study and game research is no exception. And yet, for our purposes, it is too broad of a framework. So, its value lies instead in highlighting the most important points of analysis in our context. Thus, amongst the aforementioned components of enjoyment, two stand out: facing "tasks with a reasonable chance of completion" and "sense of control over our actions".

The factors above are themselves objects of frequent study, in essence, they constitute player agency.

2.3.1 Player Agency

As games are interactive experiences, there is an expectation that players can impact the experience and are in control of the impact they have.

As presented in C. Klimmt et al.'s [9] work, effectance is an important factor when choosing to play a video game. To feel capable of enacting change reinforces our desire to engage in that activity. Some experiences are enjoyable for us, because of our ability to control how they unfold, to change the environment. Taking away the player's perception that their will matters and impacts the experience, reduces their enjoyment and diminishes motivation to participate.

In the same vein, if the results of player's actions are too unpredictable, it will not be possible for them to create a reliable mental model of how they can interact with the game's world. Humans construct representations of all activities they engage in to improve their performance with time and experience. If a reliable model cannot be reached, then nor can competence, this frustrates the individual. C. Klimmt et al's [9] work also supports this, it showed players picked games that they believed themselves able to reasonably master. The feeling of improvement, of developing competence to achieve self-efficacy [7] , is also a significant motivator when picking a game.

The combination of these factors has been described as the feeling of agency, hence its importance for this study. These are aspects intrinsically linked to how randomness impacts the PX, as many previous examples have indicated. Particularly, as C. Compton's article (Section 2.1.3) already pointed out, even just the perception of randomness can make the player feel a lack of control. In his example of the failed TCG, Compton underlines how the use of dice impacted the game's reception. It is then clear that if mere perception can be this significant, the design and implementation is all the more so.

Given all this, agency is, perhaps, the most important aspect of the PX when evaluating the effects of randomness.

[7] Self-Efficacy - where competence or efficacy is one's ability to effect change, self-efficacy is the belief one has of being (un)able to effect change.

2.3.2 Measuring sense of agency

Having established the relevance of agency as an evaluation metric for this project, what tools can be used for this evaluation must also be either researched or constructed.

The concept of Locus of Control (LoC) was first introduced by J. Rotter [10] in 1954, succinctly, it measures a person's perception of control over their lives. On one end, some individuals believe themselves to be in complete control of their lives, what transpires in their lives is above all a consequence of their actions/omissions. Opposite are those who value most the influence of external factors, all the aspects they could not shape and, ultimately, conditioned their decisions. Respectively, these are people with a strong internal LoC or external LoC.

By itself, LoC does not measure agency, it focuses not on a particular instance in a person's life but rather on their general outlook. Nevertheless, Rotter's LoC Scale, specifically, its questionnaire provides a robust foundation for a potential evaluation tool. The existing scale provides an evaluation of an individual's general belief of control, so, an adaption could be made to focus on a specific event. To collect data on the participant's perception of control within the context of an experiment, the generalized language employed in the original could be specified.

As seen, the experience of agency originates in the control one has in a particular situation, within the context of this work: in a game. Through its 29 item questionnaire, Rotter's Scale presents participants with generic scenarios and, for each, 2 possible answers, designed to reveal their perception of LoC. Therefore, by compiling relevant items, suited for alteration and contextualizing them within a game experience, it would be possible to evaluate the player's feelings of agency. With this shift, respondents with answers aligned with those of individuals with strong external LoC would have not experienced significant agency. On the other hand, if a player's responses were coherent with those of a strong internal LoC, it could be said the experience provided them with a sense of control.

2.4 Discussion

Randomness is a versatile and useful tool for game design, but the concept's complexity means it is not trivial to implement.

Through my analysis of existing work on the perception of randomness, it is clear that human perception is skewed. To start, our evolution as a species led to the development of mechanisms meant to aid in our survival. However, applying these innate capabilities to problems they are not suited for can instead be to our detriment. To survive, we became better at discerning patterns and creating connections, for that reason, randomness confounds us as it defies those expectations. In looking for patterns where by definition there are none, or, in the case of pseudo-randomness, they are too complex to discern, humans can make bad decisions.

Considering the above, one might think that randomness would be something to eliminate from experiences such as games, but it has long proven its value. Precisely because it defies our expectations, the introduction of randomness in games, from the uncertain outcomes of our actions to the construction of similar but unpredictable circumstances, creates novelty and excitement. Despite our instincts that fight for survival, we crave experiences with risk, as knowing every outcome leads to monotony. A game like chess does not introduce randomness through any mechanics, and yet, it can be very exciting as we can only try to predict what the opponent will do next.

It is trivial to say that excitement and novelty are common motivations for engaging in an activity. Nevertheless, from the work referenced, it has been shown that feeling in control, being capable of enacting change, more precisely, the change we intend, upon an environment, are motivators as well. Choosing to play a given game, much like a sport or other hobby, can be tied to our feelings of competence in that task and our perception of improvement. When failing repeatedly at something we become frustrated, all the more so if, with each attempt, there is no sense of progression. For this reason, randomness is tough to balance, as a poor implementation can mean the player feels their actions are inconsequential.

As discussed, it is a tool that can create novel experiences from the same building blocks (procedural generation), and perhaps, from a player's perspective, this is the least frustrating approach. Though their starting circumstances are different each time, the outcome is tangibly a result of their actions. Conversely, using it to make the outcomes of player actions uncertain is much more nuanced. A player can accept that at times they may be unlucky, but not feeling they lack agency. That could be why, much like we cannot predict with certainty the actions of an opponent in chess, the use of randomness to shape the behavior of adversaries not only is exciting but also is deemed fair in the player's eyes.

To evaluate a game's use of randomness, it is the player experience that must be analyzed. After all, games, digital or otherwise, are experiences crafted for the player's enjoyment. Agency is perhaps the most relevant component of that experience when evaluating the implementation of randomness. In many mediums, the consumer is a passive agent, their only role to observe, but games are interactive experiences, and though in many cases the overall narrative is unchanged by player action, there is much enjoyment in furthering it. People enjoy books and movies, regardless of the lack of interactivity as the LoC was never in question. When playing a game, the LoC should be internal, lest players be victims rather than participants.

In conclusion, randomness is an important game mechanic when used effectively. Controlled randomness, for one, can increase replayability, but the implementation of output randomness requires complex balancing. As has been done with difficulty, one answer to this problem might lie in dynamically adjusting it. Not all players are alike, and so, testing and balancing while in development can only accomplish so much. An implementation that alters how random the experience is, based on player performance, may improve the PX.

3

Case Study

Contents

In this chapter, an overview of the testbed game is presented and the model for the adaptive balancing of randomness will be detailed.

In this work, the exploration of this thesis will be focused on the impact of our approach within a confrontation between player and an adversarial Non-Player Character (NPC), particularly a Boss NPC. As such, the application of the model will center on the random factors within the NPCs decision-making and all other related systems which are similarly random.

Moreover, the testbed game for this thesis will be Enter the Gungeon and it is briefly described in this chapter's first section. The game was made in Unity 3D and the necessary modifications for the implemented solution (see Section 3.4) were developed using open-source tools and the analysis of its decompiled codebase.

As for the model, the aim is to provide players with a better experience by evaluating how they are performing and adjusting randomness-affected mechanics. In particular, in the event a player is losing, it will reduce randomness so as to limit the chance of occurrences that may feel as unfair or out of the players control, at time in which they may already feel some degree of powerlessness. Conversely, if they are performing well, it will add to the unpredictability, in an effort to promote new and unexpected scenarios, increasing excitement.

3.1 Testbed Game

For this work, we will be modifying the game Enter the Gungeon (EtG) [22]. This 'bullet hell' rogue-like has been highly rated by players, having an "Overwhelmingly Positive" rating on the PC platform Steam. Its gameplay experience is well established and does not require further validation. Though not all described mechanics or systems are relevant to this work, this section aims to provide a complete view to give the reader a full understanding of the changes made.

Figure 3.1: Enter the Gungeon - Title Screen

The game, available on multiple platforms, was made in the Unity 3D engine, which uses C# as a scripting language. The developers took advantage of the inherent 3D capabilities of the engine to achieve a 2.5D look and feel with sprites rendered in a pixel-art style. The core gameplay loop has the player choose a character, each of which starts with different equipment. They then navigate several floors of the titular 'Gungeon' to reach its innermost level. Each floor has a procedurally generated layout, composed of rooms of several kinds, mainly filled with enemies that the player must fight using whatever weapons they have acquired. The floors culminate in a large room where a Boss awaits, it must be defeated to progress. Setting special and hidden floors aside, the game features five floors at the end of which, if they have survived, the player will have completed a 'run'.

The player character is defined by a variety of attributes, what weapons they have collected, and what items, both passive and active, that they have acquired. In terms of attributes, the most relevant is health, which has a set maximum that can be increased in several ways. There is also 'armor' which in essence counts as one hit-point that cannot be replenished and does not count towards the maximum health. Then, there are the weapons, most of which guns, the outstanding majority ranged, which the player can collect while progressing through the run, they can hold as many as they want at any given time but only wield one at a time, generally. Most weapons also have a finite amount of ammunition, this can be replenished by finding 'Ammo Boxes' during play. Finally, there are two kinds of items, passive, which give the player on-going bonuses, and active, which must be activated by the press of a button and usually have a cool-down associated.

Throughout the game, the player will face several obstacles, traps, or holes in the floor, for example, but also adversarial entities, enemies, and bosses. The enemies encountered vary greatly in abilities and behaviors, though they all share the common theme of the game, guns, and are in a lot of cases fashioned to look like bullets. Likewise, the attacks of most are ranged, these go from simply shooting a single bullet to complex patterns. Because of this, and as a result of the number of enemies that the player finds in a given room, a lot of projectiles are usually traveling around them, this categorizes it as a 'bullet hell'. Much in the same way, all bosses utilize projectiles both directly and as part of creative mechanics to fight the player, for instance creating a barrage of bullets that must be avoided by timing dodges precisely.

Lastly, the player has another two important mechanics available to them, dodging and 'blanking'. All of the player characters can execute a dodge-roll during which they are invulnerable to damage. This is a tool used to both avoid damage and move around as it is necessary to, for instance, to cross over gaps on the floor. Similarly, blanks (following the bullet theme) are consumables the player can use which destroy all enemy bullets within a radius around the player when used and, so, are an alternative form of damage avoidance/mitigation.

With all this in mind, though the developers have not themselves provided the tools for it, the game's

Figure 3.2: Gatling Gull Boss Encounter (Experiment Version) - Enter the Gungeon

player community has and continues to make solutions for the modification and addition of content. Software currently available allows for user-created Dynamic-Link Libraries (DLLs) to be run with the game. These additions are coded in C#, much as if creating scripts for the game while developing it in the engine (Unity 3D), which are then made into the aforementioned DLLs.

3.1.1 A Boss Encounter

Given bosses are our case study, a base level understanding of the Boss encounters in EtG is pivotal. Thus, in this subsection, we will go over the game's original implementation.

From a design perspective, encounters of this kind serve as a hurdle the player must overcome to advance in the game, a test. In EtG, they gate the player's path to the next level by presenting a mechanical challenge. As such, defeating them is first and foremost motivated by the will to press on. And, overcoming them requires familiarity with the game's main mechanics, namely: dodging, shooting, and the enemy's abilities.

In this thesis, however; their suitability as a case study stems from their decision-making. Commonly, in games, simple enemies do not execute complex algorithms to pick their next action, relying instead on more basic solutions such as Random Number Generation (RNG) [1] . Here EtG is no exception and many NPC adversaries utilise this type of solution. In particular, the behavior of bosses is composed of three main different types:

- Targeting - Which dictate how the enemy can perceive and manage its targets (i.e.: the player(s));
- Movement - Which control the Boss' movement, from wandering in the room to pursuing a player;
- Attacking - Which are responsible for deciding what action or attack to execute next.

[1] Random Number Generation (RNG) - a process by which a random number is generated, often used in games as means to randomize a parameter of a mechanic.

Each of these types is polled, in order of priority (top to bottom), and if a behavior of that type is underway, its step is executed. Then, lower priority behaviors are executed in tandem, provided the higher priority ones are non-blocking.

On the matter of the attacking behavior, the Boss chosen for this study implements a RNG based solution, with each available action having an assigned probability and criteria. When deciding on a next action, the process is to:

1. Determine available actions (i.e.: not on cooldown and execution criteria are fulfilled);

2. Use RNG to select one of the available actions, considering the probability of each.

3.2 Modding

Modding the testbed game is how this work will construct the environment to test its hypothesis, thus to understand its implementation, this section will give an outline and high-level explanation of this process.

To contextualize, EtG was developed using the Unity 3D game engine, a fact favorable to modding efforts. This engine works off of the C# programming language and does little in the way of optimization when building games. Moreover, Unity does not change or hide variable names, making decompiled code easy to understand, provided the developers have not taken precautions. As such, it can be stated that games like EtG are much more easily modded.

Knowing this, a segment of the game's community strived to create tools that would allow for a wider, less knowledgeable, set of individuals to make their own modifications. From these efforts originated Mod the Gungeon (MtG), a program that runs on the game's executable, altering it to add multiple functionalities, namely, the injection of user alterations. Alongside it, lower-level tools have also been developed to add rooms, weapons, items, or modify enemy behavior, player stats, and more. In this work, given its objectives, the tools used were:

- MtG[2] - which enables the running of mods for the EtG;

- EnemyAPI[3] - provides a simple interface for a mod to make alterations of enemy behavior;

- CustomRooms[4] - to inject new rooms into the game's level generation system.

Of these, the second was integrated within the mod developed in this thesis, whilst the third is a standalone mod and was included in the modification package required to run this work's experiment (see Fig. 3.3).

[2]Mod the Gungeon - Tool Link
[3]Enemy API - API Link
[4]Custom Rooms - Mod Link

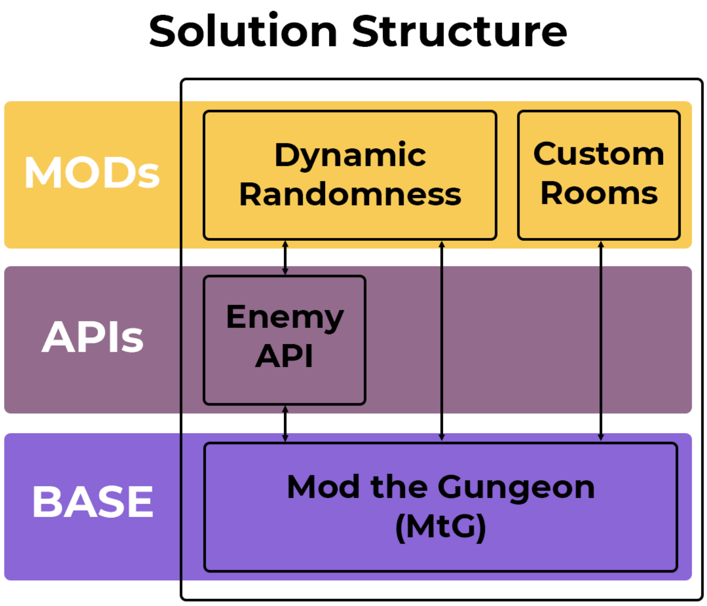

Figure 3.3: Solution structure

On a technical level, the modifications made to the game were introduced through three main methodologies. Having been developed with C#, most of these resulted from core aspects of object-oriented programming, particularly extension of classes and overriding of methods. The most used method, given its versatility, was class extension. As MtG enables the introduction of new classes, and even extends classes from the game's assembly, modifying behavior is versatile.

However, the alterations required were often of a much smaller scale, so simply creating an override for a method would suffice. Thus, the second most frequent alteration tool was "hooks", methods introduced through the mod, which MtG could tether to existing methods of classes from the game's assembly. The most frequent use case was the changing of arguments of the overridden method or even its inner workings. Beyond this, by carefully selecting which method to "hook", encountering a specific instance of a class was simple. This was useful when different alterations were needed for varying instances or at multiple points throughout the experiment.

Lastly, as mods for MtG are structured, a "Module" class exists that serves to set up the mod, and a controller-class can be developed and instanced for needs that might span a playthrough. This controller was relevant most of all in the data logging for the experiment, both in regards to the collection of the data throughout and saving it at its conclusion.

3.3 Boss Behavior Model

The Boss NPC studied will, until the encounter concludes, cycle through the following model:

1. Evaluate its present resources;
2. Evaluate the players resources;
3. Compare the two evaluations to decide on one of three scenarios:
    (a) It is loosing to the player;
    (b) It and the player are on equal-footing;
    (c) It is winning.
4. Based on the determined scenario, opt, respectively, for one of the following behaviors:
    (a) Chaotic - the next action is chosen arbitrarily, with all choices available equally likely to be picked (only the action just executed is excluded);
    (b) Balanced - the next action is chosen randomly, with each having an assigned likelihood, making some more frequent than others;
    (c) Sequential - a predetermined sequence of actions exists which the Boss will follow, thus, in this mode, the behavior is entirely deterministic.
5. Based on the selected behavior, decide on an action;
6. Once the action is concluded, repeat from 1. until the fight concludes.

3.4 Implementation

In this section, the implementation of the solution will be explained by breaking it down into three core facets. This examination will start with the player character's alterations, then progress to the environment and flow necessary for the experiment, and, finally, the modifications the the Boss behavior, the cornerstone of this work. For each of these, the current state of the game will be described, followed by the implemented changes to meet both the requirements of this work and create the conditions to test them.

3.4.1 Player Character

Players interact with the game through a character with resources and abilities which shape those interactions. In this section, it is explained what was available in EtG and how the implemented solution sought to constrain participants, given the wide choice available, in order to construct a consistent experiment scenario across players.

3.4.1.A Game State

As a rogue-like, a core tenet of the experience is the randomization in each playthrough of everything from level-layout to loot. And yet, it is now common that games such as EtG implement systems which give the player some level of control over their initial circumstances. On such example is the choice of different starting equipment, or indeed, characters.

When starting a new playthrough through EtG's hub, players can choose from a selection of different characters, each starting with a different gun (or guns), equipment and statistics. These different characters provide variety, but also allow for players with different play styles to find what improves their experience. However, across all characters some fundamental aspects are consistent:

- the ability to dodge;
- a starting weapon with infinite ammunition;
- one passive item.

From this base, all characters introduce some variety by changing one of these points. Moreover, they are all possessed of some amount of health points, with some presenting twists on that point as well. In this work, the character's weapon and health point are the focal points of its loadout. These resources (i.e: health and ammunition) can be found randomly through normal gameplay to replenish what the player as lost or spent.

3.4.1.B Implemented Solution

With too much inherent variability in the original design, the decision regarding the player character was that it ought to be standardized such that its resources were adequate for participants to engage with the experiment.

On this matter, allowing participants to pick different characters is incompatible with this experiment's design. As they are only meant to experience on playthrough, consistent across participants, such variety is undesirable. And, though it could be argued that different play styles would benefit from other characters, the experience must only be made to vary through the independent variable. To this end, for the "Quick Start" route, the player character was hard-coded for participants, and in the hub all other characters were disabled.

In terms of the resources, the process was iterative and much consideration went into balancing the experience. Given the aforementioned randomness of resource acquisition during normal play, mimicking reward attribution in the test level would not be viable and the alternative settled on was to use preliminary tests with users to balance the scenario.

Ultimately, participants played as the “Pilot”, one of Gungeon’s starting player characters. This choice stemmed from its simplicity and the fact it is the game’s starting character. As for its equipment, neither of its starting items would impact the experiment. Both had interactions with mechanics that did not play a part in the experiment scenario. Lastly, as referred, its health was edited from 3 starting “hearts”, or a total of 6 health points, to 10 health points. Its starting gun had the magazine size increased to reduce the frequency of reloads.

In terms of data collection in relation to the player character, the only point logged through the experiment was the amount of health points with which a battle was concluded. Given that the player’s overall aptitude was not under scrutiny, measures such as ammunition spent or number of dodges was not considered.

3.4.2 Experiment Flow and Environment

This section lays out the changes made to the normal flow of the game in order to accommodate this work’s experiment. The two main objectives of these alterations were to, first, guarantee players of all experience levels would be given the necessary knowledge to participate. And, second, to create an environment in which noise would not be introduced into the collected data through systems or mechanics extraneous to the present work.

3.4.2.A Tutorial Game State

When playing for the first time, a player of EtG would be met with its tutorial, though only if no pre-existing save file existed. Which is to say, any player that had previously played, as long ago as it might have been, or as little as they might have played, would have the option but not in any way be encouraged to play through it. Further, if multiple people participated using the same computer, at best, only the first would experience the tutorial. To guarantee a base level of knowledge for everyone playing, this had to be changed.

On the other hand, the tutorial itself, as originally implemented, served our purposes without fault. It explains the basic commands to anyone new and introduces them to the game’s most fundamental mechanics, like dodging or “blanks”. The singular potential caveat was its introduction of aspects of the game not relevant to this experiment, which padded out its length.

3.4.2.B Tutorial Implemented Solution

To present the tutorial to every participant, two points needed addressing - entering and leaving - each of which had multiple scenarios to account for.

For entering the tutorial, we had to assure that any path that would direct the player to a normal playthrough is redirected. That being the case, there exist two routes to account for:

- the "Quick Start", which is what the experiment will ask participants to follow;
- and the game's hub, which needs to be considered in case there is a failure in following instructions.

In either case, placing players at the tutorial's entrance was the main goal. For the second, avoiding the game's hub was also important to remove the possibility of tampering with the experiment. The game uses this space to house several NPCs the player can interact with and alter their next playthrough. For instance, its in this hub that the player would find and be able to select one of the alternative player characters.

Likewise, when exiting the tutorial, a player would normally be loaded into the game's hub either:

- by returning through its entrance, the set of stairs at its beginning;
- or the elevator at its conclusion.

Once more, the main objective of these alterations is avoiding the hub, but it is also of note that they simplify the flow of the experiment.

For the purposes of determining the player's level of familiarity with the game, along with the self-reporting measures, logging was added to the tutorial to flag whether a participant had completed it or elected to skip it.

3.4.2.C Level Game State

In turn, EtG's levels are procedurally generated to create a web of connected rooms, from a starting room to a "Boss" room, with rooms in-between mostly containing enemies or equipment for the player. As discussed, all this variation is not conducive to this work, thus the level generation too had to be altered in order to accommodate our design. The alterations would have to allow for the construction of the layout previously discussed and the replaying of encounters by participants.

Given the above, the game, in its original form, required that the access of players to its hub and respective NPCs be significantly limited, players must be placed in the tutorial, with the option to skip it, and the level generation needed to be removed, with a custom test level replacing it.

3.4.2.D Implemented Test Level Solution

The levels in EtG are procedurally generated, as is a staple of the rogue-like genre. But for the test level, a standard disposition was needed.

In the process of creating our test level, the first step was to create the necessary rooms to be used in making said level. This was accomplished through a Room Editor community-created tool, and, at first, only a room for the Boss encounters was created. In the interest of limiting deviation from the original game design, the designs explored were replications of the rooms already implemented for this fight. The choice of which to use hinged on our priorities, particularly that the player be given ample opportunity to observe the entity's behavior. By the end, rather than a custom solution for the Boss room, we decided on one of the rooms designed for this particular Boss included in the game.

Each room used in level generation can be of several different types, such as Treasure, Trap, Boss rooms, and the like. The editor including functionality to select these proved important as they are often tied to relevant functionalities. For our use case, a Boss room was inadequate because it is programmed to provide players with rewards at the end which would be randomized and differentiate participant's experiences from one another. However, a room being marked as such is what signaled for the display of the Boss' health bar. Here, the solution was to mark rooms as Mini-Boss rooms, which despite being meant for different entities than the one we would use, included the health bar and did not carry the same issues.

The second step then was creating a level generation procedure that would make a test level to our specifications with those rooms. At the start, a corridor layout was tested, participants would face each variation of the encounter in sequence. In order to remove potential bias resulting from the chosen order, we used Latin Square to select three possible patterns (Table 3.1), of which one would be randomly assigned to the player at the start of the experiment. During development a flaw arose, however; the fights, apart from the first variant presented, are not easily replayable in this format. As a result, the second iteration shifted to a branching layout (Fig. 3.4) with a central room in which the player starts and leads to each of the variations of the encounter. In the final experiment, participants were instructed to face them in alphabetical order as displayed in Figure 3.4.

Sequence ID	A	B	C
0	Sequential	Chaotic	Dynamic
1	Chaotic	Sequential	Dynamic
2	Dynamic	Chaotic	Sequential

Table 3.1: Latin Square based fight sequences

Though participants would be asked in the questionnaire which Sequence ID they were attributed, for the sake of redundancy this information was also logged. No other data points were collected regarding the Test Level.

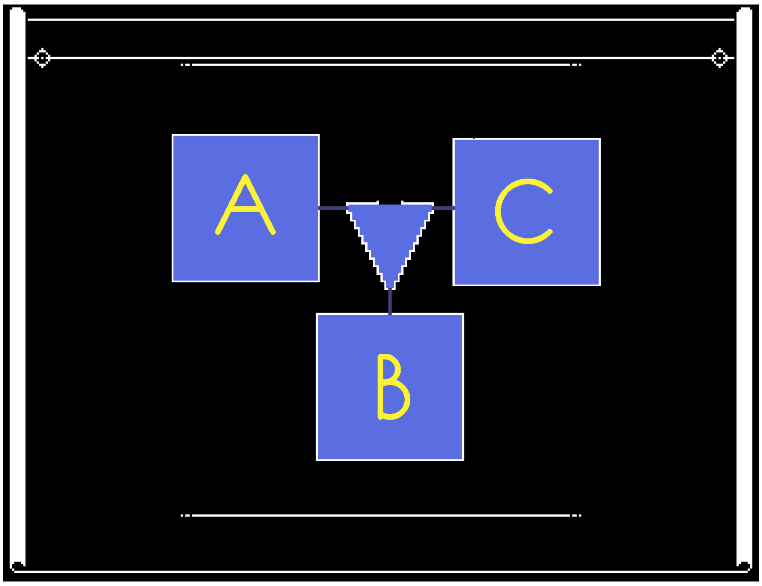

Figure 3.4: Test Floor Layout

Finally, replayability was a big consideration, participants are meant to be able to replay each encounter as they please. The game already implements a restart option which begins a playthrough from the start, resetting all progress and generating new levels. But, as we will examine ahead, the mechanism through which each of the copies of the Boss is made to have a different behavior is dependant on the order in which they are instantiated. This means that in order to use the restart option to allow players to replay a fight, on restart the Boss variants could not shift between the three rooms. Implementation wise, the solution to this was to force the generation seed to always be the same, this guarantees two things:

- the rooms will be created in the same order every time, resulting in the variants maintaining their distribution;
- and, the only other randomness-based mechanics impacted, such as chest rewards, do not affect the test level or the experiment.

Because the game utilizes separate RNG solutions for the NPC behavior, and some other systems, this locking of the generation seed does not impact the randomness within Boss encounters.

3.4.3 Boss Encounters

The following section reviews the Boss behavior in EtG going into more detail regarding action selection than previously explained (Sec. 3.1). Adding to this, the three variants designed and introduced are described.

3.4.3.A Original In-Game Behavior

Boss encounters in EtG are varied, but most follow the same basic decision-making patterns. For this work, the existing pattern will serve as a mid-point within the model.

In terms of behavior, most of the analyzed Gungeon Bosses can be described as having multiple isolated abilities or attacks that are executed, one after the other, throughout an encounter. How the next ability is selected varies slightly, but, by and large, it is random. Each ability is assigned a likelihood and may have conditions for being executable and, once the previous finishes executing, a list is made of the actions whose conditions have been met from which one is selected.

Looking specifically to the Boss chosen for this study, the "Gatling Gull", it is a first-floor Boss with a total of six abilities implemented. As described, this Boss would then use random number generation to obtain a float, which in conjunction with the values of each ability, would result in the selection of an action from those available. It is also of note that, for this Boss, multiple attacks used RNG within their execution.

Action Name	Description	Weight	Condition(s)
Walk and Shoot	The Boss follows the player while shooting a random spray of projectiles.	3	None
Fan Spray	The Boss, while stationary, shoots a wide spray of rapid bullets towards the player.	2	None
Big Shot	The Boss shoots a large slow bullet towards the player, upon impact it explodes into multiple projectiles which go in all directions.	1.5	None
Waves	The Boss, while stationary, shoots two waves of bullets towards the player with a small interval between them.	1.5	None
Leap	The Boss jumps to one of several pre-assigned spots in the room.	1.5	Available leap-spots within the room.
Rockets	The Boss executes Leap and then shoots rockets towards the ceiling which come down in spots marked by red cross-airs on the ground and explode.	3	Can only execute after Leap.

Table 3.2: Base Gatling Gull actions and respective description

Note: the Boss has one more action available, a melee attack, but this one cannot be selected in the above process and is instead executed reflexively when a player comes into range.

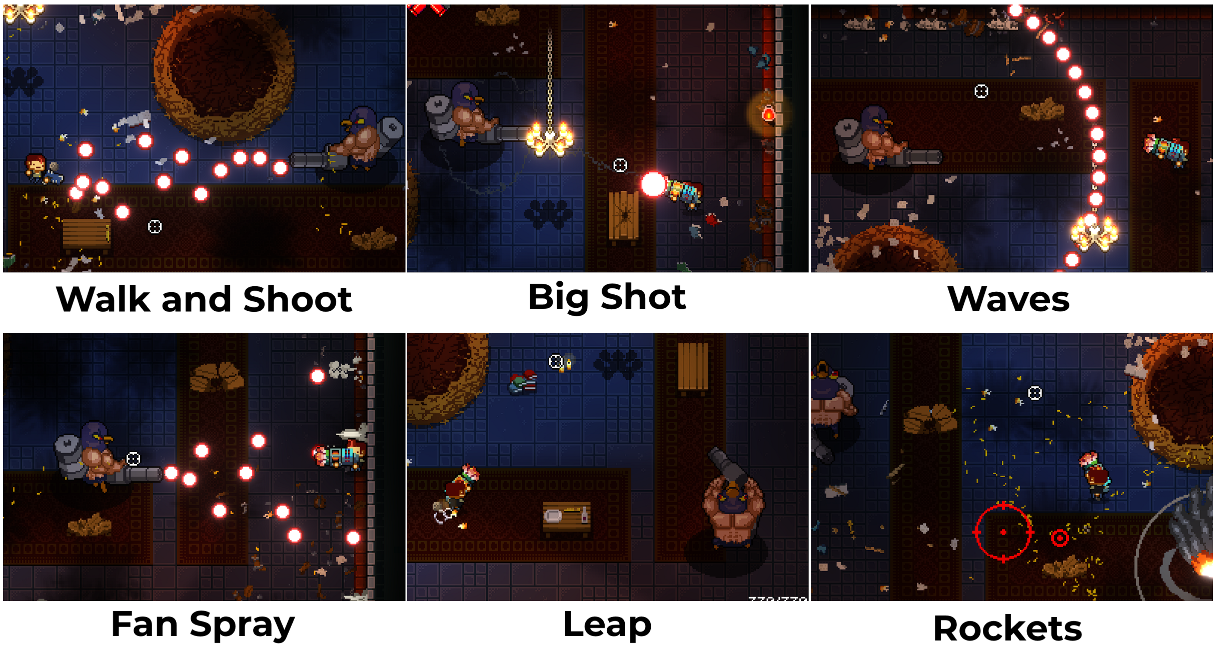

Figure 3.5: Base Gatling Gull action pictures

3.4.3.B Implemented Solution

Considering the existing implementation of Boss behaviors, a new solution was necessary that allowed for a Boss to pivot based on the criteria outlined in the model (see Sec. 3.3). Nevertheless, the existing functionality was useful in creating the chaotic pattern and within the dynamic pattern for both the chaotic and middle profiles.

A – Chaotic For all three of the Boss instances, both the overriding and architecture were very similar. But above all else, for the chaotic variant, the alterations were very limited. Knowing the selection process currently employed, as this variant needed only to give all options an equal opportunity, at first the only change was to make all ability weights equal to 1. As the development continued, the only other modification was to impede the execution of the same attack back to back, as a result of preliminary testing.

B – Sequential In contrast, the sequential variant required a completely custom solution for its decision-making. The game contained what seemed to be an already implemented ‘sequential’ pattern, so this was the first explored solution. And yet, it quickly became apparent that this was only a partial implementation, likely scrapped. This being the case, an override to the existing functionality that iterates through a predetermined sequence was developed. A list of actions is iterated upon, and after an action is complete, the next is already predetermined. The exception to this is the requirements for an action being unmet, resulting in skipping it. The action sequence present in the final solution [5] was designed based on the original weights for this Boss' actions, previously presented (see Table 3.2).

C – Dynamic At last, this works cornerstone, the dynamic variant, combined the previously outlined solutions, both original and of this work.

As laid out in the model (see Sec. 3.3), this behavior pattern assesses the player's and Boss' resources to determine who is winning and choose a behavior profile. When designing this evaluation, multiple approaches were pondered, with, ultimately, the choice of a simple and perhaps naive approach: focusing only on health percentages. If the two's health percentages differed more than 7.5% then the Boss would operate in one of the two 'extremes'. In sequence, if its health was highest, or chaotically if it was lowest. That 15% window would have the Boss behaving with a 'moderate' level of randomness (see Table 3.3).

Action Name	Walk and Shoot	Fan Spray	Big Shot	Waves	Leap	Rockets
New Weight	3	0	1.5	1.5	0	3

Table 3.3: Action weights under the Moderate profile

On the one hand, for both the chaotic and sequential profiles, the Boss behavior followed the outlined for the respective standalone patterns. On the other hand, the 'moderate' profile uses the original implementation as a middle ground. Given that it utilizes random selection with assigned weights for each action, it achieves a semblance of predictability without being so.

D – Additional Alterations Having set out how the behaviors operate, it is necessary to tackle the actions they employ.

For all variants, preliminary testing showed that behavior variability could be detrimental given the limited time of the experiment. For this reason, from the original set of actions, two were excluded. When selecting which to cut, 'Walk and Shoot' was never considered as it is integral to the Boss movement. Likewise, the 'Rockets' attack stayed as it provides a window of opportunity for the player, so removing it could cause accessibility issues for less experienced players. Of the other four, the two excluded were 'Fan Spray' and 'Leap', the former for its similarity to 'Walk and Shoot'. And, the latter for its low impact in the fight, rarely being used outside of the 'Rockets' action. And, when used without Rockets as a follow-up, it also confused inexperienced participants as it often placed the Boss out of view with no visual queue after.

With the action set finalized, development focus turned to the balancing of the RNG within actions themselves. It followed, given this work's motivation, that the model ought to be applied throughout. In

[5] Sequential behavior followed the sequence: 0, 2, 3, 0, 5. (Indexes of actions in the order listed on Table 3.2)

other words, the adaptative randomness should be present not only in the decision-making process but in all other relevant aspects of the Boss's behavior, such as actions. To that end, two actions needed modifications given their reliance on RNG:

- Walk and Shoot - an angle range dictates the spread of fired projectiles, with RNG selecting the individual angle of each shot;
- Rockets - a random vector is applied to the player's position at the time of a shot, the resulting coordinates are where the next rocket will fall.

For both of these attacks, the calculations were altered so that the level of randomness could adapt. For the sequential pattern, the angle range and vector magnitude were decreased. While, for the chaotic pattern, both increased. These changes were, likewise, activated for the respective profiles within the dynamic pattern. With this, not only would decision-making predictability increase or decrease according to the behavior, but also action results themselves.

3.5 Conclusion

In summary, the modded version of EtG described in this chapter provides a foundation for this work's experiment. Locking participants to the specified character, the 'Pilot,' and presenting them to the tutorial, their resources are set up and the relevant know-how provided. Within the test level, the three Boss behavior variants are laid out for the player to fight. There are three variations on the sequence of the fights to avoid biasing the results. The chaotic and sequential variants provided a point of comparison for this work's model, both being static encounters in regards to the decision-making process. Each constitutes a different end of the randomness spectrum. Finally, the dynamic variant, which implements our model, adjusts between levels of randomness, from its action selection to action results themselves, based on player and Boss resources.

4

Evaluation

Contents

In this chapter, the experiments conducted for this thesis are examined. To begin, we will go over the preliminary experiment, from objectives, to procedure and ending on the alterations it prompted for the main experiment. Then we will put forth the goals and procedure for the main experiment, the results of which will then be analyzed in the following section.

4.1 Preliminary Experiment

4.1.1 Objectives

Before setting out to test this thesis' hypothesis, there were several aspects which required verification and validation. There are two main motivators for these checks: first, no questionnaire was found that went over all points of analysis deemed relevant, which lead to the creation of an original; second, the present circumstances due to COVID restrictions prompted a remote design that made researcher supervision and intervention unnecessary.

4.1.1.A Questionnaire Construction

Given this thesis research question, it follows that player experience would be a central aspect of the evaluation. Further, as examined in the related work (see Chapter 2) another fundamental component is agency, a player's sense of control and impact. Moreover, this evaluation must consider the time and comprehension required from players, given the objectives for the experiment design.

First, outlining the questionnaire's structure, the decision was beginning with demographic data. These dimensions included the usual questions of gender and age, which could prove to influence results, given previously observed tendencies. Then, extending to the frequency of play and familiarity with the testbed or comparable games. Finally, reaching a LoC assessment, included as a result of the research on the importance of agency for this work's hypothesis. However, seeing as the length was a constraint, alternatives to Rotter's Scale [10] were investigated, leading to the use of the IE-4 [11]. This tool is a short-scale for the assessment of LoC consisting of 4 Likert-scale items with seven levels.

Onto the player experience, evaluating the contribution of this work's model required a tool that covered its main aspects, something other research has explored. Throughout the development, multiple questionnaires were considered, such as the Game Engagement Questionnaire (GEQ) [12] and the Intrinsic Motivation Inventory (IMI) [13]. Not only have these been validated through existing works and employed thereafter in other research, but also they focus on pertinent points for this analysis, such as enjoyment. Looking at the requirements of this thesis, IMI was deemed most suitable. This decision resulted from its extensive modularity, allowing for the selection of the relevant dimensions and respective items. More specifically, of its several instruments, the first, which evaluates several different dimensions (e.g.: enjoyment, effort, and others), was helpful towards addressing concerns for the overall length of the experiment per participant. In its final iteration, the questionnaire includes one item per included dimension (see Table 4.1). Though, in earlier stages, up to three items per dimension were explored, especially to include a positive and negative statement for each.

Further, pertaining to the evaluation of agency, the inclusion of the above dimensions from IMI already provided an analysis on competence. The factor, per prior research (see Section 2.3.1), already speaks to Effectance, the effective interaction with one's environment. And yet, there was another component to agency, Self-Efficacy. Which, to reiterate, referred to one's perception of their capacity to effect change. For this dimension, two items, one negative and one positive were adapted from existing LoC items. Lastly, a further two were added related to agency, both speaking more simply to control, in order to add redundancy and for the sake of clarity. For all of these dimensions, adapted from IMI or Rotter's Scale, items were evaluated on 5-point Likert-scales.

To end, in regards to self-reported measures, participants had to select which two encounters they found most similar. Using two alternative behavior patterns, in essence, included in the model, the goal was to ascertain how the players perceived them in direct comparison. For this reason too, as a final question, they were asked what made the third encounter stand out from the remaining two.

Dimension	Item	Origin
Enjoyment	The Boss fight was enjoyable.	IMI
Competence	I felt competent during this Boss fight.	IMI
Effort	The Boss fight mattered to me and I put a lot of effort into it.	IMI
Tension	I felt very tense during the Boss fight.	IMI
Self-Efficacy (+)	My performance could improve with experience.	Rotter's Scale
Self-Efficacy (-)	I cannot develop the skills to win the Boss fight.	Rotter's Scale
Control (+)	During the fight I felt in control.	Rotter's Scale
Control(-)	The Boss' behaviour did not allow me to plan ahead.	Rotter's Scale

Table 4.1: Boss encounter evaluation items.

4.1.1.B Independent Participation

In many experiments, participants are observed throughout to be given help, if needed, and for data collection. For this work, it became apparent that such a system would not function.

The reasons for the above were two-fold. Number one, an in-person experiment, while allowing for easy assistance of participants, would face many challenges given COVID restrictions. Number two, as the expected length of the trials, when accounting for the tutorial, fighting each variant, and answering the questionnaire, sitting with each player would undoubtedly be inefficient and slow down data collection.

Considering the above, the implementation and questionnaire were both developed with easy and precise language in mind. Before the questions regarding the encounters, included in the questionnaire, were several sections explaining installation. After, instructions were written to walk participants through the procedure, leaving the least ambiguity possible.

4.1.2 Procedure

Participating in these preliminary tests was done by invitation. For each participant, participants were scheduled a time for their test. Contrary to the main experiment, these were monitored to collect observations and feedback. All occurred remotely.

To initiate the experiment, volunteers were provided with a link to an online form. As stated, within this form, the necessary information for running the experiment without intervention was presented. Then, participants were instructed to follow the indications therein and ask for clarification whenever they required it.

On a conceptual level[1], the form follows the below structure:

1. Disclaimers;
2. Demographic Data collection;
3. Instalation instructions;
4. Gameplay Test procedure;
5. Playthrough data collection.

After the tests were concluded, each participant was asked for feedback on the overall experience. In particular, if there had been any difficulties of note or anything that caused frustration.

4.1.3 Resulting Alterations

From the observations, modifications were made to the form and to the mod itself. For the former, these centered on language, while the latter would require further balancing.

As a result of the remote and unsupervised nature of the experiment design and accessibility concerns, especially as English may not be every participant's primary language, the wording had to be revised. Likewise, images were added to help clarify, for example, the corresponding room of each questionnaire section in the Boss evaluation. This point was pivotal as data for the self-reported measures must be correctly paired with the variant experienced.

[1] For the complete form, as presented to participants, see Appendix A.

In regards to the modification made to the game, some balancing issues also came to light. As referenced in the implementation section (see Section 3.4), the main issues to resolve were:

- Player starting resources, as they would not acquire items or consumables before facing the Boss variants;
- Boss room layout, specifically as it pertained to finding cover without outright separating player and Boss;
- Ease of replayability, to allow for repeated experience with a variant, to form more cohesive opinions;
- Decreasing variability of Boss actions to simplify behavior under the limited timeframe of the tests.

These, along with several minor implementation bugs, were addressed and a second version was created.

Because of the corrections and changes made, the second batch of preliminary tests was carried out. For these, my supervision had two principal goals: confirming all fixes had been successful and balancing had improved. With overall positive feedback and no further changes made from this version, the data collected was included with data from this work's main experiment.

4.2 Main Experiment

4.2.1 Objectives

The goal of the main experiment of this work was to find evidence that the devised model for the balancing of randomness, within the selected case study, positively impacted the PX. As such, this experiment was tailored towards finding evidence of increased enjoyment, perception of agency, or other such benefits to the player.

4.2.2 Procedure

For its procedure, as described before, players followed the experiment guide constructed with the assistance of preliminary testing (see A).

Having achieved an experiment design suitable for the execution of tests without supervision, participants were primarily sourced online. Accompanied by the link to the experiment's form, explanatory posts were made on multiple online platforms with connection to the testbed game. In these, prospective volunteers had outlined to them the nature of the research, without indication of its hypothesis or research question, and were asked to follow the form's instructions in order to participate.

Further more, just as participants for the final preliminary tests had been, more players were recruited on an individual basis. Nevertheless, with care to not disclose information that might bias perspectives and with safety in mind, as these tests too were not conducted with participant and researcher in the same physical space.

Between the two recruitment methods above, the objective was to gather a diverse data set, with volunteers of different backgrounds and experience. In particular, when recruiting in platforms related to the discussion of the testbed game and mods thereof, there was an expectation that volunteers would have a higher level of ability when playing. Considering this, recruiting through other avenues was necessary to diversify the samples.

For all participants, regardless of recruitment method, the entirety of the procedure remained the same, all the more for the lack of researcher involvement. And, so, all participants would follow the steps as described in the previous section.

4.3 Data Analysis

4.3.1 Demographic Results

In total, between the last round or preliminary tests and the main experiment, data was collected on the playthroughs of 20 volunteers. For each of these participants, their experiment version was randomly assigned upon starting. The data collected for the analysis was entirely self-reported.

Below, a breakdown of players by gender and age can be seen (see Fig. 4.1-4.2). As it shows, the population can be said to be primarily males in the 16-24 age range. This matches expectations given the testbed game with which this work was conducted.

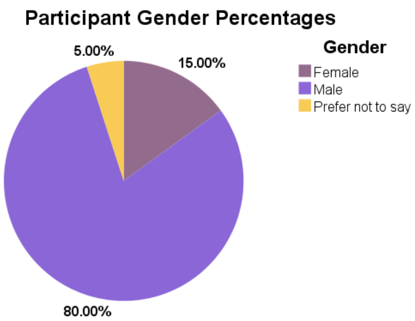

Figure 4.1: Participant Gender Percentages

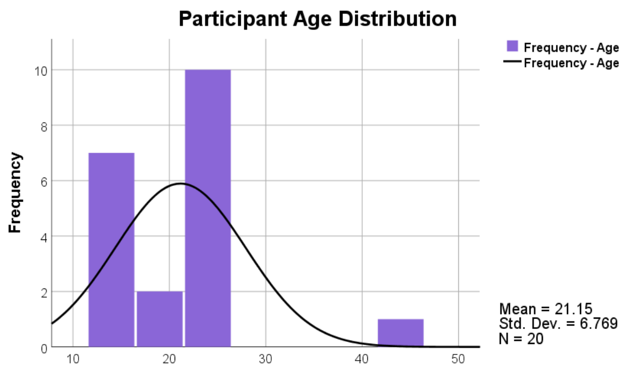

Figure 4.2: Participant Age distribution

Likewise, most volunteers reported that they made time in their schedules to play video games and were familiar with the testbed game, having played it multiple times (see Fig. 4.3-4.4). These questions were included as, though this work did not set forth to look into difficulty or how it impacts the PX, from the research, it became apparent that the impact of this works model would have repercussions involving it.

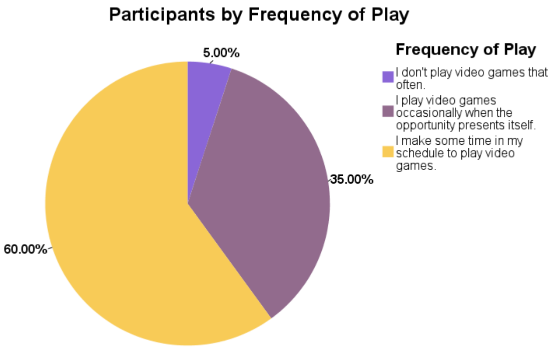

Figure 4.3: Participant frequency of play

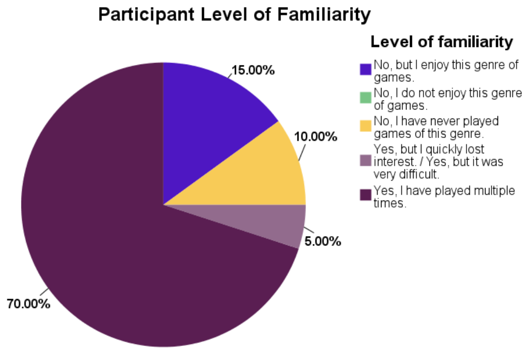

Figure 4.4: Participant familiarity with EtG

To conclude, the assessment of participant's LoC reflected a good diversity of participants. From the 4 5-level Likert scale items, the scores are summed in relation to the External-Internal locus scale (i.e.: a negatively worded item related to an internal LoC, will affect the score towards external the more strongly the participant agreed with it), resulting in a LoC score. For these scores, it can be confidently said it is normally distributed (Kolmogorov-Smirnov Sig.=0.200 and Shapiro-Wilk Sig.=0.384).

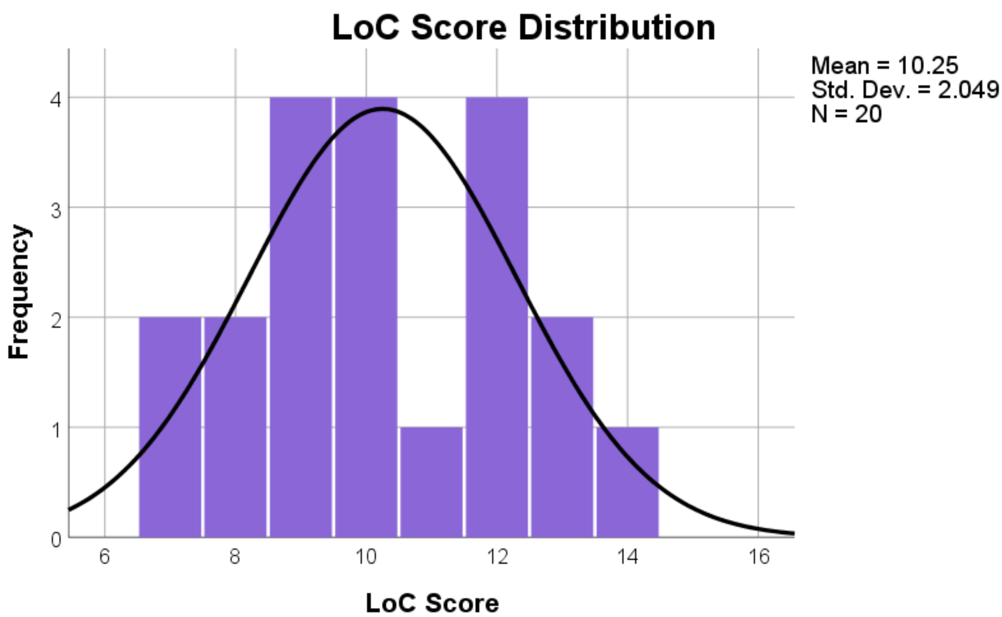

Figure 4.5: LoC Score Distribution

4.3.2 Playthrough Data Analysis

In this subsection, the variables related to each participant's playthrough will be analyzed. Amongst these, there are self-reported measures and data logged during play. First, descriptive statistics will be included to contextualize information ahead. Then, an analysis will be made on data validity. Finally, analysis that resulted in statistically relevant observations is presented.

Starting with the contextualization data, it is necessary to look at the division of participants between the experiment variants outlined previously. Each player, upon turning on the testbed game, was randomly assigned a variant of the tests. From the 20 participants gathered, 5 played variant 0, 10 variant 1, and 5 variant 2. It must be noted that 3 participants did not correctly share the data logged by the game, which means analyses on those variables include data from 17 participants only. For example, regarding the tutorial, the data shows 8 individuals played it, 9 skipped it, and for the remaining 3, the data is missing.

As shown, the distribution of players between the three sequences is far from equal. However, for all the results within this subsection, the possibility of the sequence played biasing results was considered. For none of them was evidence found that this was the case. Further, this analysis of the encounters in the order they were faced resulted only in one significant result, and another close to significance. For Boss health, a statistically significant difference was found (Friedman $p = 0.028$). Pairing encounters for analysis revealed that for the first encounter when the Boss prevailed, it did so with significantly higher health. Respectively, the average was 116, 57, and 45 health points from first to last (Wilcoxon $p = 0.028$ and $p = 0.046$). This result points to a possible learning effect that would support the inclusion of the three different sequences, given there was a worse performance for the first encounter. Moreover, self-reported measures showed that, despite this, participant perception was not significantly impacted. Which in turn, could point to a player's expectation of improvement, explaining a less harsh view of poorer performance at the start. For the PX, only one dimension had a close-to-significant difference. Enjoyment averaged scores of 4.8, 5.35, and 5.30 for the sequence of fights. These results correspond to a Friedman $p = 0.065$, and Wilcoxon $p = 0.056$ for the first and second encounters.

4.3.2.A Model Analysis Results

From analyzing the collected data on the Boss encounters, self-reported and logged, several interesting results surfaced.

First, on the impact of balancing randomness for the PX, results revealed there was a statistically relevant difference in perceived competence based on the Boss behavior variant, $X^2(2) = 7.042$, $p = 0.030$. Comparing the variants two by two, the disparity exists between the chaotic and sequential behavior patterns, $Z = -1.956$, $p = 0.05$. This seems to support that the balancing of randomness affects player perception of competence. And, this matches the expectations that the chaotic version is the hardest, while the sequential is the easiest.

	Mean Rank
[S] I felt competent during this Boss fight.	2.33
[D] I felt competent during this Boss fight.	2.00
[C] I felt competent during this Boss fight.	1.68

Table 4.2: Perceived Competence - Friedman Ranks

N	20
Chi-Square	7.042
df	2
Asymp. Sig.	$.030$

Table 4.3: Perceived Competence - Test Statistics

	[S] I felt competent during this Boss fight. [D] I felt competent during this Boss fight.	[D] I felt competent during this Boss fight. [C] I felt competent during this Boss fight.	[C] I felt competent during this Boss fight. [S] I felt competent during this Boss fight.
Z	-1.231	-1.462	-1.956
Asymp. Sig. (2-tailed)	.218	.144	.050

Table 4.4: Perceived Competence - Wilcoxon Signed Rank Test Statistics

Amongst the Boss evaluations items, no other presented a statistically significant difference. And yet, one came close to it, as Negative Self-Efficacy, on a Friedman test, resulted in an $X^2(2) = 5.059$, and $p = 0.080$. Investigating further through a Wilcoxon Signed-Rank test between the pairs of variants revealed a statistically significant difference between the chaotic and dynamic behaviors ($Z = -2.070$, $p = 0.038$). Once more, this result is in line with expectations regarding the chaotic version's difficulty and player perception.

	Mean Rank
[S] I cannot develop the skills to win the Boss fight.	2.03
[D] I cannot develop the skills to win the Boss fight.	1.83
[C] I cannot develop the skills to win the Boss fight.	2.15

Table 4.5: Negative Self-Efficacy - Friedman Ranks

N	20
Chi-Square	5.059
df	2
Asymp. Sig.	$.080$

Table 4.6: Negative Self-Efficacy - Test Statistics

Beyond this, tests conducted on participants' LoC scores showed multiple correlations with dimensions of the encounters' evaluations. First, LoC scores correlated strongly and positively with the Enjoyment of the sequential variant (Spearman $rho = .593$, $p = .006$). Suggesting that the more internal a player's LoC is, the more enjoyment they derived from experiences with little to no randomness. From the log data, three correlations were found between LoC and dimensions of the playthrough: a positive correlation with player death count (Spearman $rho = .558$, $p = .020$); and, for the chaotic variant, a negative correlation with player health (Spearman $rho = -.513$, $p = .035$) and a positive one with Boss

	[S] I cannot develop the skills to win the Boss fight. [D] I cannot develop the skills to win the Boss fight.	[D] I cannot develop the skills to win the Boss fight. [C] I cannot develop the skills to win the Boss fight.	[C] I cannot develop the skills to win the Boss fight. [S] I cannot develop the skills to win the Boss fight.
Z	-1.134	-2.070	-1.069
Asymp. Sig. (2-tailed)	.257	.038	.285

Table 4.7: Negative Self-Efficacy - Wilcoxon Signed Rank Test Statistics

health (Spearman $rho = .505$, $p = .039$). The former two seem to indicate a worse performance on the more randomized fight for participants who believe themselves more in control.

	[S] The Boss fight was enjoyable.	Death Count	[C] Player Health	[C] Boss Health
Pearson Correlation	.550	.457	-.513	.443
Sig. (2-tailed)	.012	.065	.035	.075
N	20	17	17	17

Table 4.8: LoC Correlations

4.3.3 Discussion

As we have seen, this experiment garnered the participation of 20 individuals. Each participant faced each of the three variants and reported on their experience.

Regarding observations on the Boss behavior variants, it was demonstrated that the chaotic variant was generally perceived as harder, matching prior expectations. Not only did we observe that Perceived Competence was scored significantly lower by participants, but also, players scored the Negative Self-Efficacy item higher for the chaotic variant. And though this latter observation did not show statistical significance outright, it was found that there was a significant disparity between chaotic and dynamic variants on this item. This is especially important as it does not just suggest that the chaotic variant is the one that least elicited a sense of self-efficacy in the player. It also points to the dynamic variant as the one most successful on that front. Which is to say, that variant is the one in which players most believed themselves able to improve and achieve positive results.

As for correlations with participants' LoC scores, the first of the results stood out. Given the structure of LoC calculations, the closer a score was to 0, the more external the player's LoC is, while the closer to 20, the more internal it would be. Finding a strong and positive correlation between this score and the scoring of Enjoyment for the sequential variant was unexpected. This variant has little to no RNG involvement, and thus, this result seems to indicate that volunteers with a firmer sense of control over their lives preferred it. In other words, those who feel they guide their own lives appear to prefer an experience that reinforces that perception.

Furthermore, there were negative and positive correlations with the player health and Boss health for the chaotic variants, respectively. Both of these results could indicate that player's who believe themselves in control, perform worse in circumstances in which those notions are challenged. Without, however, reporting less enjoyment of them. This point would need to be the focus of future research.

Lastly, there was a positive correlation with Death Count. Here, it could be speculated that participants with an internal LoC simply have worse performance. Instead, it is possible that fueled by their perception of control and responsibility, these individuals fought the variants more times. Their motivation is a desire to improve, viewing each result as a product of their performance, regardless of or despite external circumstances.

To conclude, his work did not find conclusive evidence supporting our hypothesis, but none refuting it either. We also found evidence of players with disparate Locus of Control scores approaching and experiencing the game differently, which should be taken into account in future research. As such, it would be particularly important to re-examine the possibility with both a larger sample size and, perhaps, an approach using starker behavior variants.

5

Conclusion

Contents

5.1 Conclusions

This work presents a possible approach to the dynamic adjustment of randomness within a game. Different players have varied play styles, and after extensive hours dedicated to one game, a player can shift play styles, making balancing difficult. In an attempt to manipulate the magnitude of randomness the player faces, finding an adjusting measure was necessary. To that end, we looked at performance and resolved to estimate it through player and adversary resources (i.e.: health points).

Before arriving at this process, we reviewed existing research on randomness and its place in games. Through it, its influence on the player experience was established and how it could be measured. Based on this knowledge, a questionnaire was developed, from existing tools, to evaluate and validate my hypothesis. The questions' main aim was to profile participants based on Locus of Control and assess their perception of the different boss encounters to evaluate the dynamic model.

Regarding implementation, a mod was created for Enter the Gungeon. It guarantees all participants have the necessary knowledge (e.g.: controls and basic mechanics), constructs the experiment's environment (i.e.: test level), alters boss behavior, and collects data on the playthroughs. These altered boss behaviors include dynamically adjusted behavior, an entirely unpredictable variant, and a deterministic alternative. The latter two were included based on the experiment design, aiming to provide contrast for participants through two static approaches.

In this work's experiment, participants faced the three boss variants in one of three different sequences each. At the start of a playthrough, one order was randomly assigned to avoid introducing biases in the data. And, after fighting each boss as many times as they chose, players reported on their experience through items adapted from IMI and Rotter's Scale.

From the gathered data, two types of results were observed. First, there were connections between boss variant and player reported experience. Second, correlations appeared between player Locus of Control and their performance playing the experiment.

Regarding the player experience, players felt significantly more competent with the predictable variant than the unpredictable one. This matches prior expectations regarding the relationship between predictability and difficulty, with the chaotic version being the most difficult. Moreover, they reported feeling more capable of improving with the dynamic variant when compared with the unpredictable behavior. Of all results, this one is the most promising, seemingly giving some support to the model.

As for the Locus of Control, those with an internal mindset reported enjoying the predictable version more and performed worse on the chaotic variant. The unpredictable variant was expected to be more difficult, particularly in comparison to the predictable alternative. This difference in performance was also coherent with the perceived competence result aforementioned. Finally, these players also had a higher number of deaths, though this point seems connected with a willingness to replay each encounter before reporting on it.

Overall, the results obtained showed promise in the approach outlined. With that, it is important to consider that as there were a total of 20 participants, in regards to LoC-related observations, further investigation with more participants is warranted. And, on the design of the dynamic variant, feedback from participants in comparing the variants indicates that it may change too subtly, especially between its ‘moderate’ and chaotic profiles.

5.2 Future Work

As previously mentioned, in the execution of this research, multiple threads arose which elicit further investigation. These include:

- broader testing of potential links between player Locus of Control and player performance in experiences of varying design, specifically concerning player control (i.e.: level of randomness, freedom of choice);
- investigation of the impact of a starker or subtler profile change, in regards to random number generation magnitude, could have on the reported experience;
- exploration of the extension of the model to multiplayer scenarios;
- testing of the application of this or analog models within other genres of game;
- research the impact of the conceptual model within other mechanics, such as reward attribution;

Bibliography

[1] M. Bar-Hillel and W. A. Wagenaar, "The perception of randomness," *Adv. Appl. Math.*, vol. 12, no. 4, p. 428–454, Dec. 1991. [Online]. Available: https://doi.org/10.1016/0196-8858(91)90029-I

[2] R. Nickerson, "The production and perception of randomness," *Psychological Review*, vol. 109, no. 2, p. 330–357, May 2002. [Online]. Available: https://doi.org/10.1037/0033-295X.109.2.330

[3] J. Schell, *The Art of Game Design: A Book of Lenses*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2008.

[4] J. Sykes and M. Federoff, "Player-centred game design," in *CHI '06 Extended Abstracts on Human Factors in Computing Systems*, ser. CHI EA '06. New York, NY, USA: Association for Computing Machinery, 2006, p. 1731–1734. [Online]. Available: https://doi.org/10.1145/1125451.1125774

[5] H. D. Fernandez B., K. Mikami, and K. Kondo, "Adaptable game experience based on player's performance and eeg," in *2017 Nicograph International (NicoInt)*, June 2017, pp. 1–8.

[6] A. Denisova and P. Cairns, "Adaptation in digital games: The effect of challenge adjustment on player performance and experience," in *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play*, ser. CHI PLAY '15. New York, NY, USA: Association for Computing Machinery, 2015, p. 97–101. [Online]. Available: https://doi.org/10.1145/2793107.2793141

[7] R. Hunicke, "The case for dynamic difficulty adjustment in games," in *Proceedings of the 2005 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology*, ser. ACE '05. New York, NY, USA: Association for Computing Machinery, 2005, p. 429–433. [Online]. Available: https://doi.org/10.1145/1178477.1178573

[8] M. Csikszentmihalyi, *Flow: The Psychology of Optimal Experience*. Harper Perennial, 01 1990.

[9] C. Klimmt and T. Hartmann, "Effectance, self-efficacy, and the motivation to play video games," *Playing Video Games: Motives, Responses, and Consequences*, pp. 133–145, 01 2006.

[10] J. B. Rotter, "Generalized expectancies for internal versus external control of reinforcement," *Psychological Monographs: General and Applied*, vol. 80, no. 1, pp. 1–28, 1966. [Online]. Available: https://doi.org/10.1037/h0092976

[11] A. Kovaleva, *The IE-4: Construction and Validation of a Short Scale for the Assessment of Locus of Control*, ser. GESIS-Schriftenreihe. Köln: GESIS - Leibniz-Institut für Sozialwissenschaften, 2012, vol. 9.

[12] K. Poels, Y. de Kort, and W. IJsselsteijn, *D3.3 : Game Experience Questionnaire: development of a self-report measure to assess the psychological impact of digital games*. Technische Universiteit Eindhoven, 2007.

[13] E. McAuley, T. Duncan, and V. V. Tammen, "Psychometric properties of the intrinsic motivation inventory in a competitive sport setting: A confirmatory factor analysis," *Research Quarterly for Exercise and Sport*, vol. 60, no. 1, pp. 48–58, 1989, pMID: 2489825. [Online]. Available: https://doi.org/10.1080/02701367.1989.10607413

Game References

[14] A. SA, “Tetris,” Electronika 60, 1984, the player must position differently shaped blocks as they fall in order to complete lines. When a line is completed, it disappears and the player earns points. The game can end either when blocks are stacked as high as the boards upper limit, leading to a loss, or after time elapses and they win. [Online]. Available: https://tetris.com/

[15] Mojang, “Minecraft (pc),” Self-Published, 2011, 3D Sandbox game which gives players the freedom to choose how to play. From exploration to construction, the procedurally generated voxel world provides the player with an environment designed to give many different individuals a satisfactory experience or the space to craft their own. [Online]. Available: https://www.minecraft.net

[16] F. H. E. McMillen and D. Baranowsky, “The binding of isaac,” Self-Published, 2011, 2D Rogue-like top-down dungeon-crawler. The player enbodies Isaac, or one of several other unlockable characters with different attributes, and navigates the titular character’s basement and further. [Online]. Available: https://bindingofisaac.com/

[17] W. of the Coast, “Dungeons and dragons,” Self-Published, 1997, fantasy table-top RPG. Players create their characters based on a set of guidelines, including customization of starting traits. Players are taken through a journey orchestrated by another individual, the game master, whose responsibilities go from giving guidance to players and controlling all aspects of the game world, from npcs to enemies. Dice are a main component as they are rolled to, for instance, determine action success. [Online]. Available: https://dnd.wizards.com/

[18] ZA/UM, “Disco elysium,” Self-Published, 2019, rPG. In this game, players control an amnesiac character from a top-down perspective in a 3D environment. Mechanically, the game revolves around dialogue and character improvement follows the metaphor of rebuilding the character’s personality. With many similarities to DnD, this game performs ’roll checks’ to determine if actions are successful, which can be affected by player traits, collected information (from dialogue exploration) or equipment. The player must resolve a murder mystery. [Online]. Available: https://discoelysium.com/

[19] F. Games, "Xcom: Enemy unknown, xcom 2 and xcom: Chimera squad," 2K Games, 2012, 2016 and 2020, respectively, tactical RPG with turn-based tactics. Players control the actions of a squad of characters in turn-based combat and must strategize to overtake the enemy in each mission. In-between missions, the player can improve a hub which in turn allows them to progress through the game by creating better equipment, training characters, among other features, introducing an element of resource management. [Online]. Available: https://xcom.com/

[20] G. Software, "Borderlands, borderlands 2, borderlands: The pre-sequel, borderlands 3," 2K Games, 2009, 2012, 2014 and 2019 respectively, first-person RPG looter shooter. In each entry to the franchise, the player can choose from several characters with different abilities, with similarities to archetypes of RPGs, use projectile weapons from handguns to canons and grenades to defeat enemies. As the player advances through the game they find more powerful and fancier gear to defeat increasingly tough and over-the-top adversaries. [Online]. Available: https://borderlands.com

[21] S. Enix, "Final fantasy xiv," Self-Published, 2013, mMORPG. Players create a character choosing a class and will then level it up, gaining new skills, to use in combat, as is common in many MMORPGS. Interestingly, in this game a single character can learn every class, though not accumulate their respective skills at once. Like in most MMOs, players can strive towards many goals, from besting challenging combat scenarios to building a home. [Online]. Available: https://www.finalfantasyxiv.com/

[22] D. Roll, "Enter the gungeon," Devolver Digital, 2016, a bullet hell rogue-like in which you pick from a several characters and attempt to reach the lowest depths of the titular Gungeon. Gameplay centers around the use of many different guns to destroy enemies and evade their attacks with dodge rolls and other tools. [Online]. Available: https://dodgeroll.com/gungeon/

A

Experiment Form

Changing Enter the Gungeon Boss Behaviour

This questionnaire is part of a Master's Thesis of a student of Instituto Superior Técnico, University of Lisbon.

All the data collected through this questionnaire will only be used within the context of the thesis of which it is a part of. If you have any questions or concerns, please contact me through: manuelmpcorreia@tecnico.ulisboa.pt

Before agreeing to take part in this study, please read the following statements attentively:

1. I have read and understood all information provided on this questionnaire. I had opportunity to ask questions, if necessary, and have them answered.

2. I understand that participation in this study is voluntary and I may opt-out at any point, without needing to justify my decision. In which case, I will not be penalized and any data collected on me will be erased.

3. All data will be stored privately, and will be destroyed after five years in accordance with GDPR.

4. I authorize the processing of the data collected within the context of this project for analysis, investigation and dissemination of results in scientific publications or conferences related to this project, by the author.
5. I understand the data collected in this study will be used as explained above.
6. In accordance with the details presented, I authorize my participation in this study and accept its conditions.

I agree:
☐ Confirm

A.1 Demographic Data

This section of the questionnaire aims to collect data about you as a person and as a video game player to look into possible connections between who you are and how you felt about the experience.

How old are you?

____

What is your gender?

- ☐ Female
- ☐ Trangender Female
- ☐ Male
- ☐ Transgender Male
- ☐ Non-Conforming
- ☐ Prefer not to say
- ☐______________________________

A.1.1 Locus of Control

For each of the following statements, pick the option you agree with the most.

Rate each item:

Item	Doesn't apply at all	Applies a bit	Applies somewhat	Applies mostly	Applies Completely
If I work hard, I will succeed.	☐	☐	☐	☐	☐
I'm my own boss.	☐	☐	☐	☐	☐
Whether at work or in my private life: What I do is mainly determined by others.	☐	☐	☐	☐	☐
Fate often gets in the way of my plans.	☐	☐	☐	☐	☐

A.1.2 Player Profile

In this sub-section the questions are intended to evaluate your familiarity with video games.

Do you play video games?

☐ Yes

☐ No (Continue to the Mod Installation Section)

How often do you play video games?

- ☐ I make some time in my schedule to play video games.
- ☐ I make some time in my schedule to play video games.
- ☐ I play video games occasionally when the opportunity presents itself.
- ☐ I don't play video games that often.

Have you played the game 'Enter the Gungeon'?

- ☐ Yes, I have played multiple times.
- ☐ Yes, but it was very difficult.
- ☐ Yes, but I quickly lost interest.
- ☐ No, but I enjoy this genre of games.
- ☐ No, I do not enjoy this genre of games.
- ☐ No, I have never played games of this genre.

A.2 Mod Installation

Please download the Mod installation package below:

Download Link

Within the ZIP you will find all the necessary files, along with a README with instructions for the installation. To start you will need to have had already downloaded and installed the game on your computer.

After completing the installation, advance to the next section.

If you have any issues, you can contact me:

- on Discord, Ghoster#3426;
- through email, manuelmpcorreia@tecnico.ulisboa.pt;

A.3 Starting the Experiment

Please read this section before starting the game.

When opening Enter the Gungeon, after some loading, the game will give you the option to "Quick-start" as can be seen in the picture below:

Figure A.1: Quick start screen

Depending on how you are playing you can activate quickstart through different keys/buttons:

- On a controller, press the "Y" button, or equivalent;
- On a keyboard, press "Q".

Please proceed to the next section of the Questionnaire.

If you don't press the key in time and get to the game's lobby, The Breach, please go to the door directly above you from where you start.

You may now start the game and, after the quickstart, continue to the next section.

A.4 The tutorial

If everything is functioning correctly, you will be here:

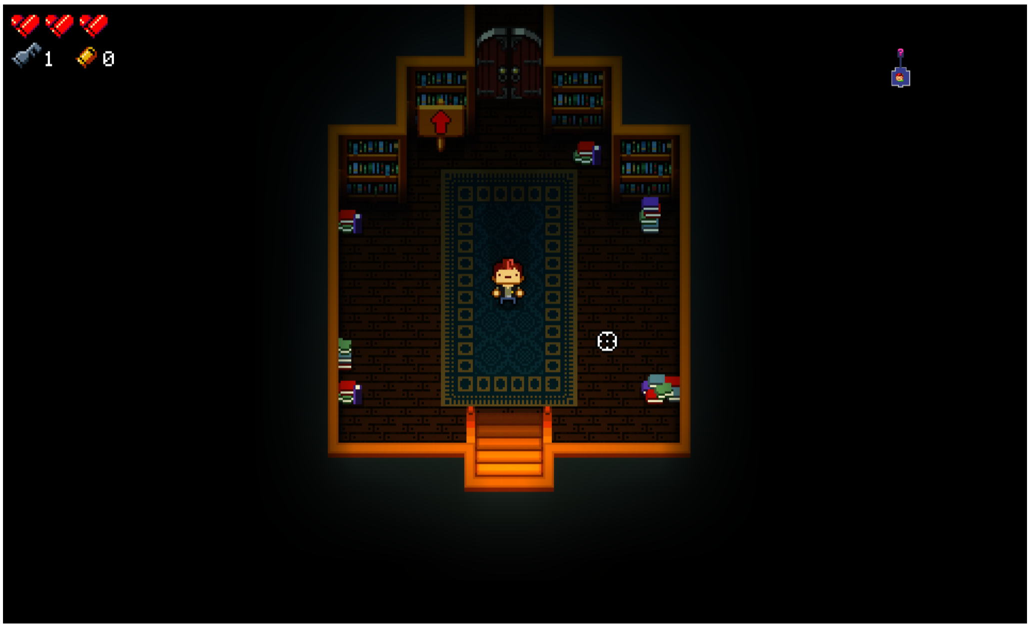

Figure A.2: Tutorial entrance

This is the entrance to the tutorial and you should continue through the door above you to start it. The experiment itself will only begin after the tutorial has been completed.

If you have played Enter the Gungeon before AND you do not feel you need to play through the tutorial, you can use the stairs below you to proceed to the experiment.

(When leaving the tutorial through the stairs, on occasion, you will be placed back in the tutorial. Go to the stairs again and the experiment will progress.)

After completing / skipping the tutorial, proceed to the next section of the Questionnaire.

A.5 The Experiment

After the tutorial you will be taken to this experiment's test level, the layout of which can be seen below:

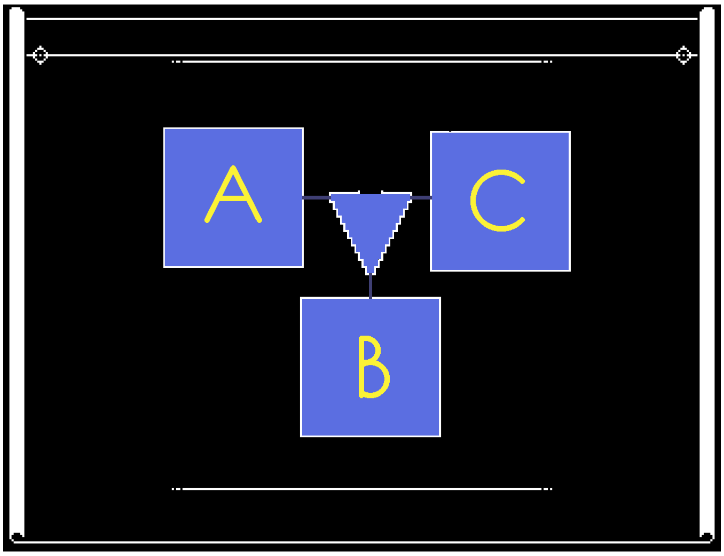

Figure A.3: Test floor map

You will be at the center room at the start of the experiment.

Each room has the same Boss fight, with some changes. Please fight them in order A - B - C.

After each fight, before you start the next one, please return to the questionnaire and complete the questions about that fight.

IF YOU LOSE, the level will reload and you can continue the experiment. You do not need to win against a given variant before filling out its section. Just make sure you have played it enough to form an opinion.

IF YOU WANT TO REPLAY, please press Escape (ESC) to open the 'Pause Menu' and click the option "Quick Restart". The level will reload and you can fight any/all variants again.

A.5.1 Experiment Variant

Each participant will have a slightly different experiment scenario. Please press "F2" in-game to open the console.

You will see the message: "Experiment Variant ->#"

Where '#' is a number, please insert the number displayed below.

Experiment Variant:

- ☐ 0
- ☐ 1
- ☐ 2

You may now continue to the next section.

A.5.2 Boss A Evaluation

At any point, you can go to previous sections and update your answers.

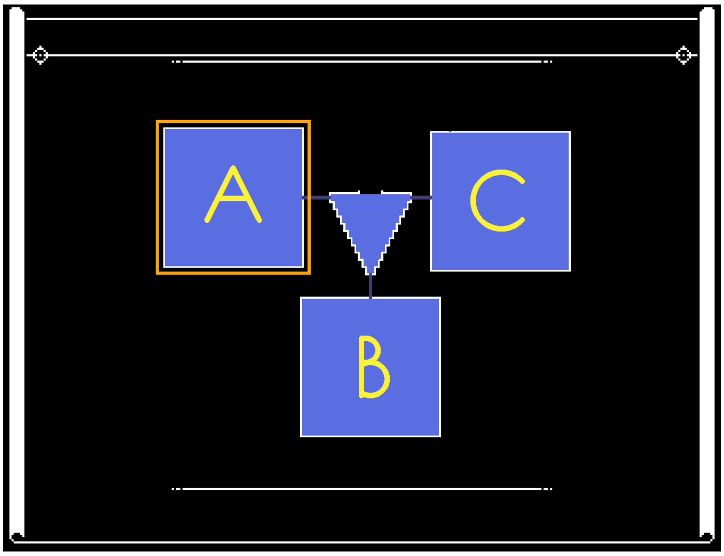

Figure A.4: Boss Room A

Rate each item:

Item	Not true at all	2	3	Somewhat true	5	6	Very True
The boss fight was enjoyable.	☐	☐	☐	☐	☐	☐	☐
I felt competent during this boss fight.	☐	☐	☐	☐	☐	☐	☐
The boss fight mattered to me and I put a lot of effort into it.	☐	☐	☐	☐	☐	☐	☐
I felt very tense during the boss fight.	☐	☐	☐	☐	☐	☐	☐
My performance could improve with experience.	☐	☐	☐	☐	☐	☐	☐
I cannot develop the skills to win the boss fight.	☐	☐	☐	☐	☐	☐	☐
During the fight I felt in control.	☐	☐	☐	☐	☐	☐	☐
The boss' behaviour did not allow me to plan ahead.	☐	☐	☐	☐	☐	☐	☐

A.5.3 Boss B Evaluation

At any point, you can go to previous sections and update your answers.

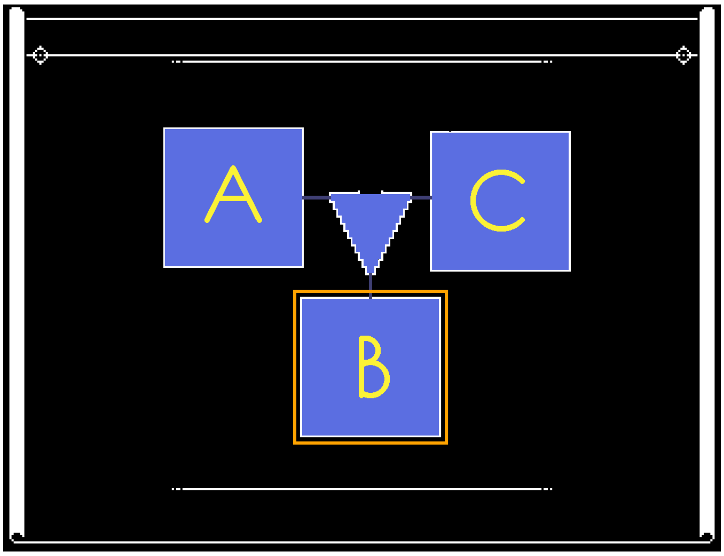

Figure A.5: Boss Room A

Rate each item:

Item	Not true at all	2	3	Somewhat true	5	6	Very True
The boss fight was enjoyable.	☐	☐	☐	☐	☐	☐	☐
I felt competent during this boss fight.	☐	☐	☐	☐	☐	☐	☐
The boss fight mattered to me and I put a lot of effort into it.	☐	☐	☐	☐	☐	☐	☐
I felt very tense during the boss fight.	☐	☐	☐	☐	☐	☐	☐
My performance could improve with experience.	☐	☐	☐	☐	☐	☐	☐
I cannot develop the skills to win the boss fight.	☐	☐	☐	☐	☐	☐	☐
During the fight I felt in control.	☐	☐	☐	☐	☐	☐	☐
The boss' behaviour did not allow me to plan ahead.	☐	☐	☐	☐	☐	☐	☐

A.5.4 Boss A Evaluation

At any point, you can go to previous sections and update your answers.

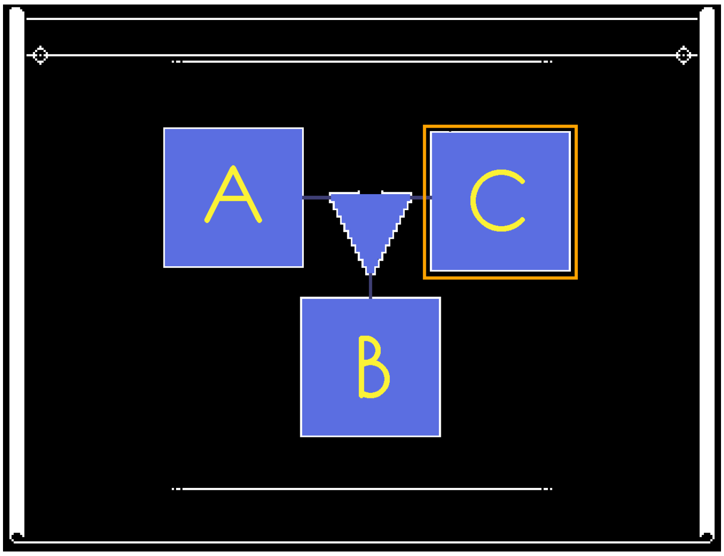

Figure A.6: Boss Room C

Rate each item:

Item	Not true at all	2	3	Somewhat true	5	6	Very True
The boss fight was enjoyable.	☐	☐	☐	☐	☐	☐	☐
I felt competent during this boss fight.	☐	☐	☐	☐	☐	☐	☐
The boss fight mattered to me and I put a lot of effort into it.	☐	☐	☐	☐	☐	☐	☐
I felt very tense during the boss fight.	☐	☐	☐	☐	☐	☐	☐
My performance could improve with experience.	☐	☐	☐	☐	☐	☐	☐
I cannot develop the skills to win the boss fight.	☐	☐	☐	☐	☐	☐	☐
During the fight I felt in control.	☐	☐	☐	☐	☐	☐	☐
The boss' behaviour did not allow me to plan ahead.	☐	☐	☐	☐	☐	☐	☐

A.5.5 Boss Comparison

Which two boss fights were most alike?

- ☐ A and B
- ☐ A and C
- ☐ B and C

What differentiated the third fight from the other two, in your opinion?

____

A.5.6 Playthrough Log

Throughout the experiment, the Mod saved some information of your playthrough. It will be saved to a file when you close the game.

Close the game and go to your Enter the Gungeon installation folder. In this folder there will be a file named: "DRMod_log.txt"

Google does not allow file upload for anonymous questionnaires. For this reason, please open this file and copy its contents to the field below. It will be a continuous line of text.

Log File Content:

A.6 Thank you!

Your participation in this study is much appreciated, if you have any questions or would like to be notified about results from this study, please contact me through:

manuelmpcorreia@tecnico.ulisboa.pt

To uninstall the modifications made for this experiment, delete the folders:

- Mods
- CustomRoomData

within your Enter the Gungeon installation folder. To completely unmodify the game, I would recommend a fresh installation.

Thank you,

Manuel Correia