

How is Player Enjoyment and Challenge Affected When Implementing Dynamic Difficulty into a Top-Down Shooter?

Adrian Carter - 1703170

Abstract—With the rapid growth of the video games industry and its popularity, players are more varied and diverse than ever before. With this diversity comes a plethora of problems for game developers. One such problem is the effectiveness of static difficulty settings. Static difficulty settings have been standard in most mainstream commercial games. However more and more research is being conducted with dynamic difficulty adjustment systems that can adjust difficulty settings during runtime. This proposal aims to test the effect on perceived player experience of a dynamic difficulty system when compared to a static difficulty system. Specifically within the context of a singleplayer, top down shooter level with an average play time of under 3 minutes. Effectiveness will be measured based on player experience feedback using questionnaires and telemetry data during gameplay.

I. INTRODUCTION

THIS research project intends to investigate the effects Dynamic Difficulty Adjustment (DDA) can have on Player Experience. With a more diverse variety of people playing video games than ever before [1] more care has to be taken when designing difficulty in video games. However, with the advancements of computational systems, it is becoming easier for developers to manage and adjust a wide variety of parameters in video games, even while the game is running. This has led to the expansion of DDA in recent years, enabling the designer to have more control over different difficulty factors. To adjust the difficulty, many heuristics can be considered. Heuristics ranging from affective feedback [2] to player performance [3]. These heuristics will be discussed in this paper in order to select appropriate methods that might be utilized in the creation of a bespoke dynamic difficulty system. This paper hypothesises that if one were to create two difficulty settings for a 2D top down shooter, one static and the other dynamic, the dynamic setting will be preferred by players. The paper also hypothesizes that the dynamic setting will have less variation in perceived challenge, since the dynamic setting should adapt better to players of varying skill levels. Section II of the paper will look at some of the theories and concepts that are relevant to the research question. The most prominent aspects that need to be considered when designing a dynamic difficulty system include player experience, flow theory, and player modelling. Section III discusses some of the most notable implementations of dynamic difficulty and why some of those techniques would be useful in designing a dynamic difficulty system for a 2D top down shooter with short levels. Section IV will discuss the methodology by which this

paper intends to answer the proposed research question and test the hypotheses.

II. RELATED WORK

A. Player Experience

Player experience in games is often described using terms such as flow, immersion, engagement, fun, and presence [4] [5] [6]. Measuring this is no easy task due to the subjective nature of player experience. Although tools can be used to objectively measure some quantitative aspects of the player experience such as biosensors to measure heart rates or electrodermal activity (measure of sweat perspiration)[7] [8] [2], they still require evaluation based on the subjective experiences of the individual. Questionnaires can be particularly useful in measuring the subjective experiences of players in a quantifiable manner, due to the fact they include a number of standardised tests that can be aligned with what the researcher intends to measure about said experience [9].

To ensure that player experience is enjoyable, the idea of Flow must be considered. Csikszentmihalyi [5] [10] uses the term “Optimal Experience” to describe the moments in which a person feels a deep sense of enjoyment or exhilaration and states that these moments are, more often than not, due to events in which that person’s body or mind is stretched to its limits in an effort to achieve something difficult or challenging. The experience of Flow, or “Optimal Experience”, in games is the experience gamers describe when immersed in video games, becoming entranced and totally focussed on their actions and losing track of time and other worries. Whether a player enjoys a video game is largely dependent on whether that game provides a Flow experience. In Flow theory, the idea of the Flow Zone (see figure 1) is the notion that in order to maintain a player’s Flow experience, the game must ideally balance challenge and the ability of the player to overcome that challenge. If the game becomes too challenging, the player becomes disheartened and anxious [11]. On the other hand, if the game becomes too easy, the player grows bored and disinterested [12]. The art of good game design, is therefore in the ability of the developer to balance challenge and player ability or skill for the duration of a game experience [13]. This becomes more difficult the larger the game grows and also the larger the potential audience becomes. With a larger potential audience, the fact that each person experiences activities in different ways becomes problematic for game designers. Every player is different, with varying skill levels and expects a

different level of challenge in order to experience Flow [5]. The most widely used solution for this problem in video games has been the implementation of various difficulty modes (Easy, Medium, Hard etc) [14]. One of the problems with this is the lack of detail the player is given when choosing a difficulty level. How easy is Easy? How hard is Hard? Without playing through the game, the player won't really be able to know which mode is best suited for their skills, since difficulty ratings are generally subjective [15]. This is where the notion of DDA becomes very useful to developers.

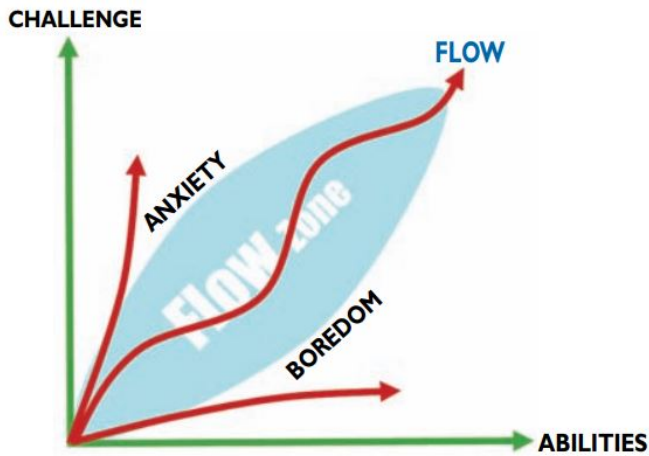


Fig. 1. Chen, Flow in Games [16].

B. Player Experience Modelling

While there has been much research on what constitutes fun in game design and how player engagement lends itself to other aspects of player experience [17] [13], much of this research is qualitative in nature and serves mainly to aid in game design conceptually. Less research exists in the field regarding quantitative measurement of player experience and player modelling. The primary objective of player experience and player modelling research, is to describe how players interact with and experience a game.

Player Experience Modelling, in the context of this paper, refers to the modelling of complex and dynamic aspects of a player interacting with the world during gameplay. Specifically the study and use of AI techniques to create computational models of player experience. This is not to be confused with player profiling, which deals with the types of players you are designing the game for [18].

Yannakakis et al [19] argue that one of the main aims of studying players in games is to gain an understanding of players' cognitive, affective and behavioural patterns. This is strongly related to the field of Human Computer Interaction [20] [21] and usability.

The simplest way of generating a model of player experience is to ask the players themselves and build a model based on the collected data. Yannakakis et al [22] describe this as Subjective Player Experience Modelling. Using short questionnaires like this seems like an effective method to

model player experience during the experiment. This study [23] used gameplay interactions to construct a computational model of player experience using statistical features such as how often and when the player jumped, ran, died etc. Whilst it is out of the scope of this paper to collect data such as this during development in order to adapt the dynamic algorithm design, this data collection method would be useful in measuring the effectiveness of said design once complete.

The creators of Polymorph [24] used methods from machine learning and procedural content generation to create a 2D platformer that generates upcoming parts of the level based on the current skill of the player. In order to create the system, a model of difficulty was required. Instead of modelling difficulty based on subjective game design ideas, the creators designed a mass data collection tool to gather information about level segments directly from players. Each player would play a short segment of the level and then give feedback by rating the perceived difficulty on a scale from 1 to 6. The tool would also collect telemetry data on the players behaviour while playing the level segment which could be applied to a machine learning algorithm in order to model player skill. Again, there are clear differences in this study to the one in this paper, however, the collection of player behavioural data is something worth considering when testing the effectiveness of the dynamic and static difficulty settings.

III. DYNAMIC DIFFICULTY ADJUSTMENTS

The challenge of designing a difficulty curve for players of varying skill level is a big problem even experienced designers with strong domain knowledge have difficulty solving. In contrast to static difficulty, dynamic difficulty adjustment (DDA) attempts to address this issue.

Currently, research in the field of DDA is in a state that does not allow for particularly complex systems, and as such, cannot be used as a framework for DDA systems in other games. Although some research is being done to establish DDA frameworks [25], they generally cover only certain basic elements of gameplay, usually within 2D games [26] and more complex variants need to be implemented in bespoke ways, depending on the genre and game elements that need to be adjusted.

The ways in which these methods adjust difficulty may differ, but they all exhibit a common theme in prediction and intervention.

A. Dynamic Scripting

Dynamic scripting (DS) is an online learning approach that maximises an AI players chance of winning a game by determining what tactics will lead to the solution, when presented with a certain problem and domain [27]. Using dynamic scripting, the AI player adapts its tactics based on the opponents' tactics using domain-specific knowledge and several rulebases that categorize various opponent types. Each script rule is assigned a weighting value that is adjusted based on the success or failure rates of said rule.

The AI learns and adapts progressively using this technique. The AI will update the weight value of the current rules

being used during a confrontation with the player, based on the outcome of said confrontation. If the resulting outcome is beneficial to the AI, the rules that lead to that outcome will have their weight values increased, whereas the rules that were counter to that outcome will have their weight values decreased. If the outcome is detrimental to the AI then the inverse occurs. Remaining rules adjust accordingly to keep the sum of all weighted values in the rulebase constant. Eventually, the AI would become too difficult for human players to compete with and so would need to be balanced properly [28]. In order to balance the opponents ability with the players, there are three common ways to modify dynamic scripting:

- 1) High-fitness penalizing: Increase or decrease weight values proportional to a fitness value. Reward suboptimal behaviour and penalize superior behaviour to essentially make the AI less efficient.
- 2) Weight clipping: Uses a maximum weight value to limit the rules that are chosen from the rulebase. The higher the maximum value the greater the disparity between the most optimal and least optimal rules. Lowering the maximum weight value during gameplay would make less optimal rules more likely to be chosen, thereby decreasing the effective difficulty of the opponent, and vice versa.
- 3) Weight culling: As with weight clipping, uses a maximum weight value limiter. However, the weight value of rules are allowed to extend above the maximum value. When a rule has a weight above maximum, the AI is unable to select said rule, thus automatically balancing the AI strength in proportion to the player.

The commercial viability of this Adaptive AI has been dismissed by some developers in the past [29] [30]. The reason for this is the inherent difficulty in creating game AI with online learning capabilities without the fear that the AI will learn inferior behaviour. This is a large concern, particularly to publishers, in the games industry. Simple games with less computational risk have been seen to work [31] and research has been done to effectively implement Adaptive AI using DS into existing commercial games [32].

B. Resident Evil 4

Resident Evil 4 [33] was created with a hidden grading system that judged the players performance in the background unbeknownst to the player. The system grades the player from 1 to 10 with higher grades leading to a higher challenge. If the player plays well by scoring headshots often and surviving for long periods without dying very often, the player is given a high grading which means more enemy spawns and more powerful enemies with higher health and better accuracy etc. If the player dies a lot or is inaccurate and generally not playing very well, the system lowers the players grade, and with that the enemies become fewer and less dangerous. The constant adjustment of difficulty without the player even knowing this system was in place, lead to a great experience for the players. Knowing that a game has dynamic difficulty adjustments in place can sometimes affect the experience for the players. In this way, a DDA system should adapt to the player in an

inconspicuous manner, so the player is not aware of the outer intervention. While the techniques in this system are bespoke to the game, the general idea of a player rating system to model ability is one that is prevalent in the field of DDA. The heuristics used would also be useful in the creation of a dynamic difficulty system.

C. Homeworld

One example of potential abuse of a difficulty system is seen in *Homeworld* [34]. The game is a space strategy simulator that involves the player collecting resources and building units to fight an AI team of similar units. At the end of a mission, the system would rate the power of your current fleet size and adjust the number of units that would be spawned when you started the next mission, effectively raising or lowering the difficulty. Unfortunately, once players realised this, they looked for ways to abuse the system by scrapping their entire fleet before ending the mission. This would make the next mission much easier since there would be fewer enemies spawned but by scrapping their fleet, they still retained the resources needed to quickly build another large and powerful fleet and “steamroll” the enemy AI fleet.

D. Left 4 Dead AI Director

The *Left 4 Dead* games created by Valve are 4 player cooperative first-person shooter games [35] [36]. The Artificial Intelligence (AI) director found within both *Left 4 Dead* games is focussed on maintaining pace: having an intensity that peaks and troughs throughout play [37]. This intensity fluctuates in real time and alters the frequency and strength of enemy waves during a given game. This keeps players on edge throughout the game and ensures the players make progress towards an end goal. If the players stay in any one area for too long, the director sends more enemies their way to encourage them to keep moving. This method makes the game far more replayable than games with more traditional, static spawning systems due to the uniquely unpredictable enemy spawns.

In order to determine each players “skill” the AI director closely monitors the performance of each player. Determining a player’s skill level quantitatively becomes problematic since we tend to judge skill based on qualitative measures. However, there are many attributes that could be employed to determine a players skill value in an fps: their overall health, damage ratio (received vs delivered), enemy kills, health packs used, shot accuracy, ammunition consumption rate and many more. These attributes can be used to help evaluate player skill and would be suitable heuristics to use in developing a dynamic difficulty system.

The AI director also attempts to model player emotional intensity, or “stress” level, throughout each game. This helps determine how intense each player is finding the game at that time. This is quantitatively measured using factors such as: enemy proximity to player, number of enemies in level, number of enemies actively attacking player, enemy proximity to player on enemy death etc. In the game, special enemies have the ability to incapacitate the player, if this happens, the emotional intensity of the player reaches it’s maximum value.

As well as normal enemy spawns, the director also has the ability to spawn a “mob” of 10 - 30 normal enemies behind or to the sides of the player route in order to surprise the player. There are limits to the number of enemies that can in the level at any one time and cooldowns to these “mob” spawn events so as not to completely overwhelm the players. The idea of using an emotional intensity value is very intriguing. It would be very interesting to see if this could be scaled down and adapted to suit a game with shorter levels.

Both player “skill” and “stress” values, along with other pieces of information such as how long the player has spent in the current area, are used by the AI director to manage the pace of the game. The director will influence the pacing by controlling where enemies are spawned, how many enemies are spawned, what enemies are spawned, how enemies will act, ammo and health spawns among many other factors. For example, a typical level will start with a small number of normal enemies, increase the spawn rates of enemies gradually and introduce special enemies until the players stress levels reach a maximum value. Then, at the peak of game intensity, the AI director will cease spawning enemies and give the players a brief respite of around 30 seconds to organize themselves before moving forward. Overall, I believe this system to create pacing, peaks and troughs of intense gameplay would be an interesting route to go down to see if it could be adapted to suit shorter play times.

E. Hamlet

Hamlet [38] is a DDA tool built by *Hunicke and Chapman* that uses methods taken from *Operations Research and Inventory Management*. The tool would adjust the difficulty of the first person shooter *Half-Life* [39] as the player progressed through the level. Algorithms were created to measure the difficulty based on the speed at which the player progressed and what resources the player currently had. It essentially manipulates the game difficulty by studying and adjusting the supply and demand of the players inventory.

IV. METHODOLOGY

The philosophy underpinning this paper is largely based on Empiricism. The view that information is primarily gained through sensory experience. Hence my methodology will mainly talk about how to quantify subjective data using likert scales and using telemetry data in order to empirically evaluate the effectiveness of my computing artefact.

A. Hypotheses

When designing a dynamic difficulty system there are several objectives one can hope to achieve. One objective is to create a system that leads to increased replayability of a game. One might also wish to implement DDA to make a game more accessible to players of varying skill levels. Another aim could be to create a system that enhances overall player enjoyment. For the scope of this paper, the focus is on creating a dynamic system that is effective in a game with shorter levels. Effective can mean a variety of things, in this case an effective dynamic

difficulty system would be one that is more enjoyable and engaging than that of a standard static difficulty system. Xue et al [40] found that when DDA techniques were applied to multiple mobile games, player engagement increased by up to 9 percent. These findings suggest that dynamic difficulty could increase player engagement in other games. While engagement encompasses multiple factors, This paper attempts to look at enjoyment specifically. Therefore the first hypothesis is that players will find playing the dynamic difficulty version of the game more fun than when playing the static difficulty.

The second hypothesis considers the fact that in the literature, dynamic difficulty adjustments are said to help solve the problem of scaling difficulty for different players with different abilities [41] [38]. In that regard, a well designed dynamic difficulty setting that takes individual player skill as a variable should, in theory, be perceived as equally challenging to all players, whereas a static difficulty system should be more challenging to less skilled players and less challenging to more skilled players. Therefore the second hypothesis of this paper is that there will be less variance in the challenge that different players perceive when using the dynamic difficulty setting compared with the static difficulty setting.

The third Hypothesis is based on the belief that players will not be able to identify which level used a dynamic difficulty system. Research suggests that a good DDA system should be hidden from the player and that most players should not realise that they are playing a game which uses dynamic difficulty. [CITE REF HERE] Therefore, if the DDA Algorithm is performing adequately, it would be interesting to see if this is the case here.

The study

B. Sample Group

In order to acquire the required sample size within a short timeframe, The Games Academy studio at Falmouth University was the most convenient locale to gather players for research. Students are randomly selected, ensuring a sample with a variety of gaming abilities, although it must be noted that the proportion of experienced gamers is likely to be high since the Games Academy consists entirely of game development students. In order to mitigate this, the participants are questioned on their prior gaming experience. Future studies would benefit from a more diverse selection of participants but due to time and funding limitations it was deemed as a necessary limitation. Three *a priori* power analyses were conducted for this experiment. The first analysis consisted of a one tail T-test with an expected effect size of $d = 0.5$, a Bonferonni-corrected alpha error probability of $\alpha = 0.017$ and a power of $1 - \beta = 0.95$. The resulting recommended sample size was $N = 59$. The second power analysis consisted of a one tail F-test of variance with an expected ratio variance of 3, a Bonferonni-corrected alpha error probability of $\alpha = 0.017$ and a power of $1 - \beta = 0.95$. The resulting recommended sample size was $N = 50$. The third power analysis consisted of a Chi-Square test with an expected effect size of $d = 0.6$, a Bonferonni-corrected alpha error probability of $\alpha = 0.017$ and a power of $1 - \beta = 0.95$. The resulting recommended

TABLE I
HYPOTHESES TABLE

	Hypothesis	Null Hypothesis	Data Source
1	Players will enjoy playing the version with dynamic difficulty more than the version with static difficulty	Players will not enjoy playing the version with dynamic difficulty more than the version with static difficulty	Questionnaire
2	There will be more variation in how challenged players were playing with static difficulty than when playing with dynamic difficulty	There will not be more variation in how challenged players were playing with static difficulty than when playing with dynamic difficulty	Questionnaire Gameplay Statistics
3	Players will be unable to reliably judge which version used dynamic difficulty	Players will be able to reliably judge which version used dynamic difficulty	Questionnaire

sample size was $N = 46$. Therefore the minimum number of samples required is $N = 59$. The sample actually acquired during this study was $N = 61$.

The sample consists of one group of players, each player playtested both difficulty settings during the experiment. Each player will be given a consent form as well as an information sheet with only vital information such as control layouts, so as to limit bias. Each playtest will last no longer than 3 minutes per level. This means that each experiment should last between 15 to 20 minutes depending on the skill of the player.

C. Playtest Session

The playtesting of the game requires each player to play through a short tutorial to begin with. This tutorial will involve completing several tasks. The first task will be traverse through a set of short waypoints. The next task will be to fire on and defeat several stationary enemy targets. A task will also be given to use any special abilities the player may have. This brief tutorial should help the player to become acclimated to the basic controls of the game as well as the general game systems. This will help to limit the practice effect the player might experience later on during testing of the two systems.

Once the tutorial is finished, the game will randomly load one of the two level versions. One level version will include the static difficulty setting, the other will include the dynamic difficulty setting. The difficulty setting that the player is testing will be unknown to both the researcher and the player. The double-blind nature of the research will help to reduce observer bias as shown in Hrbjartsson *et al* [42].

The player will have 2 minutes 30 seconds to complete the level. This will be unknown to the player since this may affect gameplay decisions by introducing unwanted pressure bias. This completion time will be measured by the game. If the player completes the level within the time limit, the completion time, number of kills and number of deaths will be collected automatically and an in-game questionnaire will appear asking the player questions about their experience playing the level. If the player does not complete the level within the allotted 10 minutes then the level will end automatically as if the player had completed the level, but the game will indicate in the statistics gathered that the player did not finish the level. Once the player has finished the first version, they will then play the alternate version, following the same process as the first. The anonymous results of both playtests will be automatically sent to a secure university network drive.

Questionnaires are the most widely used research method, consisting of a well defined set of questions to which an

individual is asked to respond [43]. The quality of the questions in a questionnaire correlates directly to the usefulness of a questionnaire, as ill-defined or vague questions can result in bad data [44]. Questionnaires can also be subject to confirmation bias; if a question written by the researcher about an experience seems loaded rather than being as neutral as possible then the player response won't be as useful, skewing the results in favour of the researchers preconceptions about the experience [45]. This is a common pitfall with interview and focus group questions as well. A web-based questionnaire has the ability to disseminate the test to large swathes of a population regardless of geographical location and gather user responses relatively quickly and are noted to be preferred over paper questionnaires [46].

In the questionnaire players are asked whether or not they agree with each statement ranging from 0 (not at all) to 100 (extremely). The questions were originally set to a 5 point likert-scale as is used in the Core module questions in the Game Experience Questionnaire (GEQ) [9]. However, in order to gather the most detailed information, I will be gathering ratio data from 0 to 100 as ratio scaling provides the most detailed information quantitatively. The GEQ includes seven components:

- 1) *Imaginative and sensory immersion*
- 2) *Competence*
- 3) *Flow*
- 4) *Tension*
- 5) *Positive affect - fun and enjoyment*
- 6) *Negative affect*
- 7) *Challenge/Suspense*

Since immersion is mainly considering the games story and aesthetics I have elected to not include questions about this component. Each statement in the GEQ matches one of these Player Experience Components. There are several statements relating to each component and the order in which the statements/questions are placed in the questionnaire is mixed. The mean score of each components results are then used to inform the results on that particular part of the player experience. This is to limit acquiescence response bias and to get more realistic and correct data [47]. The majority of my testers will be students, taking time out of their day to playtest my game and answer the questionnaire I create. Ideally the questionnaire will be much shorter than 27 questions, especially if I ask them to play both versions of the game and fill out a questionnaire for each. Therefore I have narrowed down these questions. Clearly, under other circumstances, more questions would

provide more reliable results, but it has been shown that if questionnaires are too long they can cause the user to stop answering thoughtfully which can skew results and introduce bad data. I believe under the conditions I will be playtesting and with whom will be participating in the playtests, it is best to be prudent and not draw out the tests to interminable lengths. The questions can be seen in Table II.

D. Analysis Plan

The results of the experiment should give an indication as to how player experience was affected by the two difficulty systems. Results will be gathered by way of questionnaires built into the artefact and telemetry data such as player deaths/kills. This data will be stored in a CSV file and analysed using RStudio.

I will conduct a paired samples t-test to compare the means between questionnaire results of the two difficulty systems.

```
> View(testdat)
> t.test(testdat$Fun2, testdat$Fun1, paired=TRUE, alternative="greater")

Paired t-test

data:  testdat$Fun2 and testdat$Fun1
t = 5.3785, df = 19, p-value = 1.719e-05
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 0.8820609      Inf
sample estimates:
mean of the differences
      1.3
```

Fig. 2. An example t-test in RStudio using mock data

The example above is using mock data that I created in order to determine what tests to carry out using R. Where “Fun2” refers to the “*It felt fun*” (Table II) question asked after playing the dynamic difficulty setting and “Fun1” refers to the same question asked after playing the static difficulty setting. If my hypothesis is correct, this should show differences in player enjoyment, specifically a higher mean value for answers regarding “Fun2”. I will repeat these tests for the different questions in my questionnaires with regards to their various player experience factors. I will also be able to compare the differences in number of player deaths between the two difficulty systems. This should also help evaluate my second hypothesis, that the challenge players claim to experience when playing the static difficulty system will be much more varied than the dynamic system, due to the nature of the dynamic system adapting to the players skill level. This hypothesis suggests that the standard deviation amongst the challenge category questions should be higher in the static version than in the dynamic version.

E. Dynamic Difficulty Algorithm

In order to create my Dynamic Difficulty Algorithm, I will need to iteratively design and playtest throughout development to validate and verify my Artefact [48]. I will therefore need to ensure that anyone I have playtest the game, is not in the final experiment, since this may bias my results. I will be able to playtest the difficulty settings myself for most of the development, but nearing the end I will do a small pilot

study of a few people just to ensure the two systems are not completely unbalanced and work as I intended. Once

The plan is to use an AI Director and player grading system similar to those used in *Left 4 Dead* [35] and *Resident Evil 4* [33]. This would involve monitoring heuristics and altering heuristics as the game is being played. Currently the Artefact uses a modified version of the group team game as a base. The Artefact is being built using Unity 2018.4.2f1. This includes a twin stick player character, enemy spawners, and enemies (figure 3). I have created a script called AIDirector.cs that alters spawn rates and enemy statistics such as speed based on an emotional intensity value that ranges from 0 to 100. This value increases or decreases based on a number of factors, including but not limited to player skill, player health, number of enemies and proximity of enemies. The aim is to eventually increase the complexity of this AI Director and the challenge will be in designing the algorithm with level flow and challenge in mind. If my design is successful then there is an opportunity to use the knowledge gained in this project to design a simple dynamic difficulty system that could be utilized in my group project but this is beyond the current scope of this paper.



Fig. 3. Screenshot of the initial prototype Artefact

F. Ethical Considerations

There should be little to no ethical risks involved in this study. I am continually referencing the British Computer Society Code of Conduct [49] and its guidance on duty and integrity when conducting my studies and considering any ethical implications. Participants will be sampled from members of the Games Academy and will be subject to the Code of Research Ethics at Falmouth University. Each participant will be given a consent form as well as an information sheet discussing the reasons for the study. It will be stated that all participants have the right to opt-out of the study any any time without consequences and will have all the rights afforded to them as stated in the General Data Protection Regulation [50] if applicable, however, personal information such as name, date of birth, gender will not be needed in this study. All data gathered will be anonymous and confidential, with any data

being stored on a secure university network drive. Participants will be asked to pause the study and take a break in the unlikely event that the experiment lasts more than 30 minutes. This is to prevent any possible bouts of nausea, eye strain, or general discomfort from looking at computer screens for too long. The software will be designed to account for this possibility.

V. ARTEFACT

The computing artefact is a top-down 2D shooting game wherein the player defends themselves from an unending wave of zombie enemies. The player has several abilities or mechanics in-built. The player has eight-directional movement using the WASD keys and can use the mouse to aim the cursor and thus aim the gun. The player has a cone of vision as shown in *fig 4* that aims in the direction between the player character and cursor. This aiming cone shrinks and expands based on movement. The cone expands to a maximum of 25 degrees when the player moves to simulate less accuracy as is used in most first person shooters. The cone shrinks to a minimum of 5 degrees when the player is not moving or when the player 'aims down the sights' (ADS) using the right click mouse button or Left Shift, however the player movement speed is altered by a factor of 0.7, thus slowing the player and making them more vulnerable. The player can shoot the gun using the left click mouse button. On click, a random vector within the aiming cone field of view is chosen and a ray is cast in that direction to simulate the bullet.



Fig. 4. Screenshot of the Level Layout

VI. ARTEFACT DEVELOPMENT

The Artefact was built using Unity 2018.4.2f1¹. The DDA Algorithm was written in C# using Visual Studio 2019² as was the Top Down Shooting Game that the Algorithm would be used in.

The enemies move using the free system "A* Pathfinding Project"³ found on the Unity Asset Store.

¹<https://unity.com/>

²<https://visualstudio.microsoft.com/vs/>

³<https://arongranberg.com/astar/>

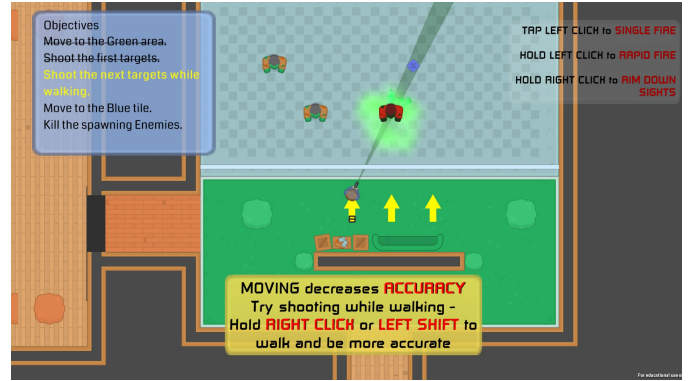


Fig. 5. Screenshot of the Tutorial

The game art was procured for free using kenney assets⁴. The music played during the game is called fiberitron and is a royalty free track found online⁵. The sound effects used are all free sounds found online⁶.

A. System Development Life Cycle

The software used in this experiment was developed in its entirety by one person. The initial prototype was created in conjunction with a video game that was in development by a team of developers including the Researcher, however, it was decided early on to discontinue this and to develop an entirely bespoke game for the project. This is mainly due to the potential risks involved in having to depend on the group based project reaching certain development milestones and not blocking the Researcher's continued development of the Artefact. Also, designing a bespoke DDA system was deemed to be much easier if the Researcher had full control of the design decisions that went into designing the game that it was to be used with.

When developing the Artefact, the initial life cycle method proposed was the Agile Method, since team collaboration and communication would have been essential. After the decision to create a bespoke game was made, however, the development life cycle that was most suited to the project became The Prototyping Model [reference book - Information System Development Life Cycle Models]. An experimental approach to The Prototyping Model was most suited to the project since, like Agile, it is an iterative process that is more responsive to user needs. The user needs, such as the flow of the game and the effectiveness of the DDA Algorithm could be proposed and evaluated through pilot testing and AI simulations. Then once feedback is attained, the system could be refined to attempt to improve and revise the Prototype to then be tested again until the Prototype satisfies all user requirements.

B. Software Testing

The DDA Algorithm functionality was tested using a combination of pilot testing and a Player Artificial Intelligence

⁴<https://kenney.nl/assets/topdown-shooter>

⁵<https://www.dl-sounds.com/royalty-free/fiberitron-loop/>

⁶<https://www.zapsplat.com/sound-effect-categories/>

Agent that simulates the player. Initially, each iteration of the Artefact was to be tested using human players. However, this seemed to be an impractical method of testing the DDA Algorithm. The development of a working dynamic difficulty system requires many refinements and iterations to be tested, which would mean a lot of participants would be required to do pilot tests throughout development. There are ways of limiting the impracticality of this, such as with posting the Artefact online to get feedback from players, on the other hand, these methods also come with drawbacks and would likely slow development considerably while waiting for feedback. Thus, a Player AI was developed to simulate the player so that testing of the DDA Algorithm could be much faster.

The Player AI has fairly basic pathfinding abilities and as such cannot be said to be as human-like as can be but serves the purposes of testing the difficulty system. The Player AI moves between the four corners of the map using the same A* Pathfinding that the enemies in the game use. The Player AI will interrupt it's pathfinding and move away from enemies that get too close however, somewhat simulating a human player. The Player AI has most of the same systems as the Player Prefab in-game. To adjust the Skill Level of the Player AI, the shots per second and accuracy values can be adjusted. As the player has an accuracy cone system built into the design, the Player AI uses this when aiming at the nearest enemy with a clear line of sight and therefore the Skill Level of the AI can be adjusted by adjusting the spread of this cone, thus adjusting the shot accuracy of the AI. Pilot tests were carried out to assess the Player AI against human players and adjust the shots per second and accuracy values in order to get as accurate a representation of human gameplay as possible. A static difficulty curve with an unchanging spawn rate was used in the pilot test. The average Kill/Death Ratios (KDR) of the five pilot testers was used as a base to aim towards when adjusting the shot accuracy and shooting rate of the Player AI. When the Player AI had achieved a similar enough KDR in simulations with the same static difficulty, it was deemed as being satisfactorily functional for the purposes of testing the DDA Algorithm.

VII. RESULTS AND ANALYSIS

The data gathered from the experiment were output into a csv file on the university network drive. RStudio 1.2.5019⁷ was used to collate and process the results. The R package ggplot2 was used to generate the graphs.

A. Hypothesis 1 - Players will enjoy playing the version with dynamic difficulty more than the version with static difficulty

This hypothesis was intended to judge whether or not players enjoyed playing the dynamic version of the game over the static version. Figure 6 shows us that there is a slight preference for the dynamic version. A t-test was carried out analysing the differences in how the players scored the two difficulty versions on the basis of player enjoyment. This test produced a p-value of 0.001638 and a t-value of 3.0634. The

p-value is very low and therefore the null hypothesis can be rejected. The mean value of the differences was 6.08 showing that on average players enjoyed the dynamic version 6 percent more than the static version.

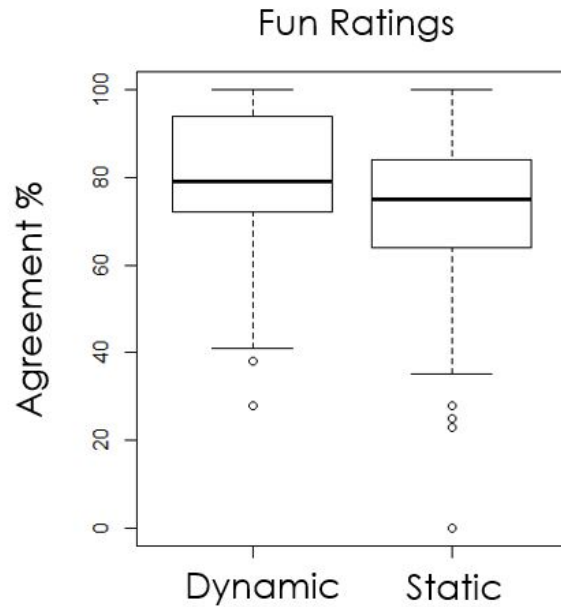


Fig. 6. Boxplot of Enjoyment Perceptions

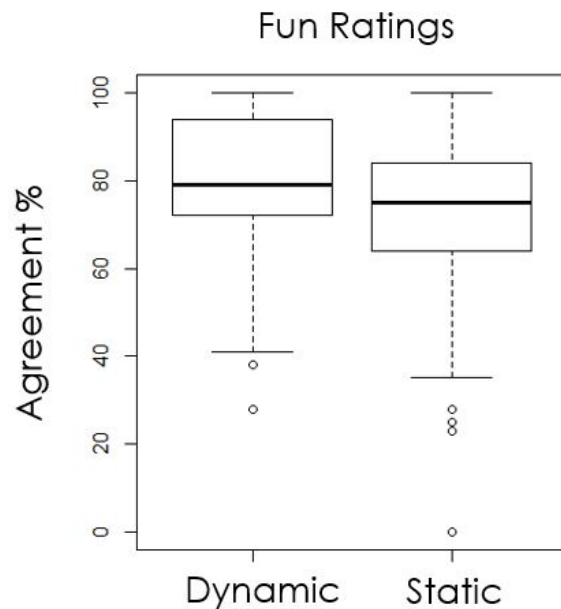


Fig. 7. Boxplot of Enjoyment Perceptions

B. Hypothesis 2 - There will be more variation in how challenged players were playing with static difficulty than when playing with dynamic difficulty

An F-Test was carried out to determine the variance of player deaths between the two difficulty versions. This is

⁷<https://rstudio.com/products/rstudio/download/>



Fig. 8. Scatter plot of Player Enjoyment vs Player Kill/Death Ratio

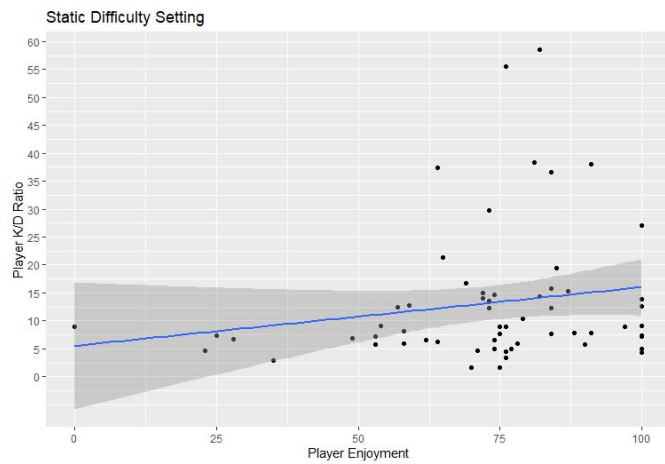


Fig. 9. Static Scatter plot of Player Enjoyment vs Player Kill/Death Ratio

deemed by the researcher to be a valid assessment of difficulty. The test produced a p-value of $2.277e-09$. This p-value is extremely low and therefore the null hypothesis can be rejected. The ratio of variances was 4.9226.

Interestingly, player perception of difficulty may not be as varied as can be seen in Figure 11. Therefore a second F-Test was carried out in order to determine the variance in challenge scores the two difficulty versions received by each player. This test analyses player perception of difficulty and therefore is tangentially related to the hypothesis in question. This second test produced a p-value of 0.4927. This p-value is too high and therefore the null hypothesis would not be rejected. The ratio of variances was 1.0047.

This seems to indicate that although the difficulty appears to be more varied based on player deaths with the static difficulty, the perception of difficulty among players is less varied than one would expect.

C. Hypothesis 3 - Players will be unable to reliably judge which version used dynamic difficulty

The data showed that 36 out of the 61 participants in the study correctly guessed which level used the dynamic difficulty system. A Chi-squared test was performed to determine

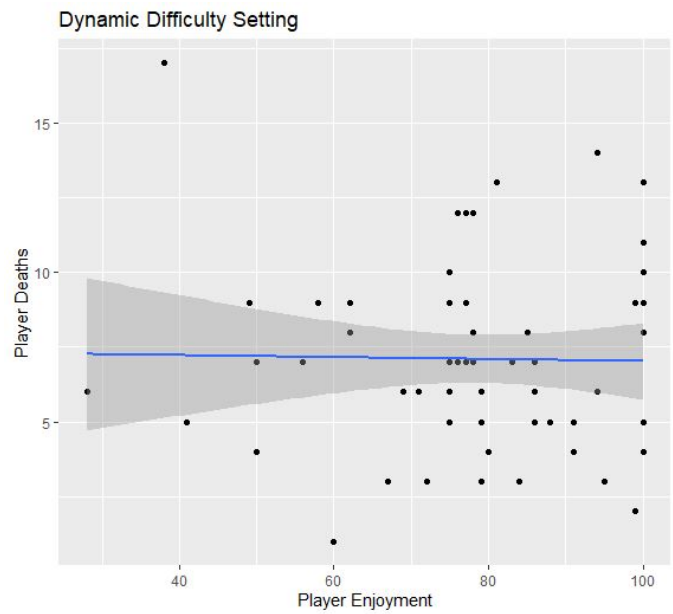


Fig. 10. Dynamic Scatter plot of Player Enjoyment vs Player Deaths

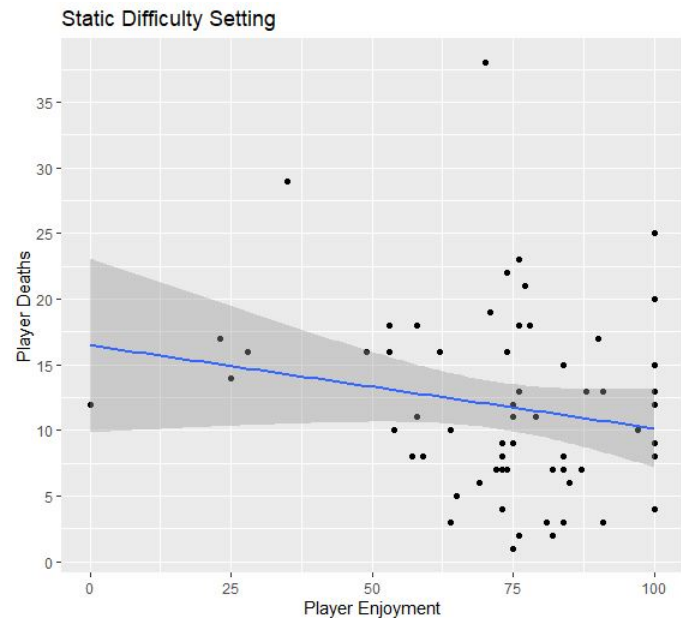


Fig. 11. Static Scatter plot of Player Enjoyment vs Player Deaths

the probability that this was higher than random chance. The test produced a p-value of 0.159. This p-value is low enough to discount the null hypothesis. This suggests that players were not able to reliably identify which version used dynamic difficulty.

VIII. DISCUSSION

IX. LIMITATIONS

X. FURTHER RESEARCH

REFERENCES

- [1] J. Juul, *A casual revolution: Reinventing video games and their players*. MIT press, 2010.

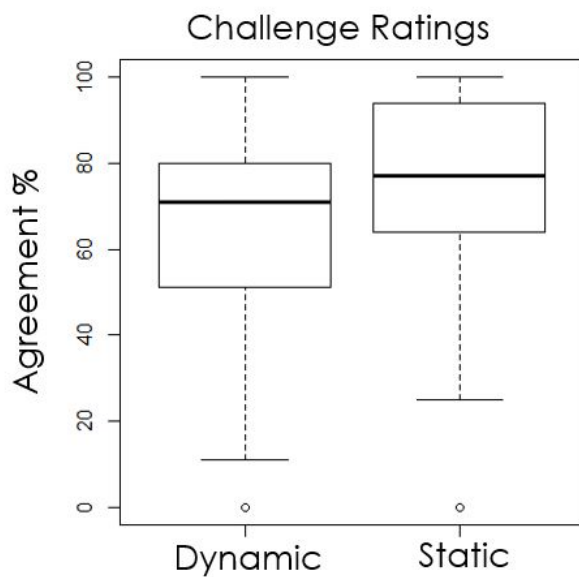


Fig. 12. Boxplot of Challenge Perceptions

- [2] C. Liu, P. Agrawal, N. Sarkar, and S. Chen, "Dynamic difficulty adjustment in computer games through real-time anxiety-based affective feedback," *International Journal of Human-Computer Interaction*, vol. 25, no. 6, pp. 506–529, 2009.
- [3] 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*. ACM, 2015, pp. 97–101.
- [4] K. Isbister and N. Schaffer, *Game usability: Advancing the player experience*. CRC press, 2008.
- [5] M. Csikszentmihalyi, *Flow: The psychology of optimal experience*. New York: Harper & Row, 1990.
- [6] A. McMahan, "Immersion, engagement, and presence: A method for analyzing 3-d video games," in *The video game theory reader*. Routledge, 2013, pp. 89–108.
- [7] K. Mikami, K. Kondo *et al.*, "Adaptable game experience based on player's performance and eeg," in *2017 Nicograph International (NicoInt)*. IEEE, 2017, pp. 1–8.
- [8] M. Ambinder, "Biofeedback in gameplay: How valve measures physiology to enhance gaming experience," in *game developers conference*, vol. 2011, 2011.
- [9] W. IJsselstein, Y. De Kort, and K. Poels, "The game experience questionnaire," *Eindhoven: Technische Universiteit Eindhoven*, 2013.
- [10] M. Csikszentmihalyi, "Toward a psychology of optimal experience," in *Flow and the foundations of positive psychology*. Springer, 2014, pp. 209–226.
- [11] K. M. Gilleade and A. Dix, "Using frustration in the design of adaptive videogames," in *Proceedings of the 2004 ACM SIGCHI International Conference on Advances in computer entertainment technology*. ACM, 2004, pp. 228–232.
- [12] G. Chanel, C. Rebetez, M. Bétrancourt, and T. Pun, "Boredom, engagement and anxiety as indicators for adaptation to difficulty in games," in *Proceedings of the 12th international conference on Entertainment and media in the ubiquitous era*. ACM, 2008, pp. 13–17.
- [13] R. Koster, *Theory of fun for game design*. "O'Reilly Media, Inc.", 2013.
- [14] J. T. Alexander, J. Sear, and A. Oikonomou, "An investigation of the effects of game difficulty on player enjoyment," *Entertainment computing*, vol. 4, no. 1, pp. 53–62, 2013.
- [15] H. Qin, P.-L. P. Rau, and G. Salvendy, "Effects of different scenarios of game difficulty on player immersion," *Interacting with Computers*, vol. 22, no. 3, pp. 230–239, 2009.
- [16] J. Chen, "Flow in games (and everything else)," *Commun. ACM*, vol. 50, no. 4, pp. 31–34, Apr. 2007. [Online]. Available: <http://doi.acm.org/10.1145/1232743.1232769>
- [17] K. Salen, K. S. Tekinbaş, and E. Zimmerman, *Rules of play: Game design fundamentals*. MIT press, 2004.
- [18] R. Bartle, "Hearts, clubs, diamonds, spades: Players who suit muds," *Journal of MUD research*, vol. 1, no. 1, p. 19, 1996.
- [19] G. N. Yannakakis, P. Spronck, D. Loiacono, and E. André, "Player modeling," Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2013.
- [20] J. Preece, Y. Rogers, H. Sharp, D. Benyon, S. Holland, and T. Carey, *Human-computer interaction*. Addison-Wesley Longman Ltd., 1994.
- [21] G. Fischer, "User modeling in human-computer interaction," *User modeling and user-adapted interaction*, vol. 11, no. 1-2, pp. 65–86, 2001.
- [22] G. N. Yannakakis, "Game ai revisited," in *Proceedings of the 9th conference on Computing Frontiers*. ACM, 2012, pp. 285–292.
- [23] C. Pedersen, J. Togelius, and G. N. Yannakakis, "Modeling player experience in super mario bros," in *2009 IEEE Symposium on Computational Intelligence and Games*. IEEE, 2009, pp. 132–139.
- [24] M. Jennings-Teats, G. Smith, and N. Wardrip-Fruin, "Polymorph: dynamic difficulty adjustment through level generation," in *Proceedings of the 2010 Workshop on Procedural Content Generation in Games*. ACM, 2010, p. 11.
- [25] M. Zohaib, "Dynamic difficulty adjustment (dda) in computer games: A review," *Advances in Human-Computer Interaction*, vol. 2018, 2018.
- [26] D. Wheat, "Dynamically adjusting game-play in 2d platformers using procedural level generation," 2013.
- [27] P. Spronck, I. Sprinkhuizen-Kuyper, and E. Postma, "Online adaptation of game opponent ai with dynamic scripting," *International Journal of Intelligent Games and Simulation*, vol. 3, no. 1, pp. 45–53, 2004.
- [28] L. Lidén, "Artificial stupidity: The art of intentional mistakes," *AI game programming wisdom*, vol. 2, pp. 41–48, 2003.
- [29] S. Woodcock, "The future of game ai: A personal view," *Game Developer Magazine*, vol. 7, no. 8, 2000.
- [30] S. Rabin, "Promising game ai techniques," *AI Game Programming Wisdom*, vol. 2, pp. 15–27, 2004.
- [31] P. Demasi and J. d. O. Adriano, "On-line coevolution for action games," *International Journal of Intelligent Games & Simulation*, vol. 2, no. 2, 2003.
- [32] P. Spronck, M. Ponsen, I. Sprinkhuizen-Kuyper, and E. Postma, "Adaptive game ai with dynamic scripting," *Machine Learning*, vol. 63, no. 3, pp. 217–248, 2006.
- [33] "Resident Evil 4," Capcom, 2005.
- [34] "Homeworld," Sierra Studios, 1999.
- [35] "Left 4 Dead," Valve, 2008.
- [36] "Left 4 Dead 2," Valve, 2009.
- [37] M. Booth, "The ai systems of left 4 dead," in *Artificial Intelligence and Interactive Digital Entertainment Conference at Stanford*, 2009, 2009.
- [38] 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*. ACM, 2005, pp. 429–433.
- [39] "Half-Life," Sierra Studios, 1998.
- [40] S. Xue, M. Wu, J. Kolen, N. Aghdaie, and K. A. Zaman, "Dynamic difficulty adjustment for maximized engagement in digital games," in *Proceedings of the 26th International Conference on World Wide Web Companion*. International World Wide Web Conferences Steering Committee, 2017, pp. 465–471.
- [41] O. Missura and T. Gärtner, "Player modeling for intelligent difficulty adjustment," in *International Conference on Discovery Science*. Springer, 2009, pp. 197–211.
- [42] A. Hróbjartsson, A. S. S. Thomsen, F. Emanuelsson, B. Tendal, J. Hilden, I. Boutron, P. Ravard, and S. Brorson, "Observer bias in randomized clinical trials with measurement scale outcomes: a systematic review of trials with both blinded and nonblinded assessors," *Cmaj*, vol. 185, no. 4, pp. E201–E211, 2013.
- [43] R. Goddard and P. Villanova, "Designing surveys and questionnaires for research," *The psychology research handbook: A guide for graduate students and research assistants*, pp. 114–125, 2006.
- [44] M. S. Litwin, *How to measure survey reliability and validity*. Sage, 1995, vol. 7.
- [45] F. J. Fowler Jr and F. J. Fowler, *Improving survey questions: Design and evaluation*. Sage, 1995.
- [46] J. R. Evans and A. Mathur, "The value of online surveys," *Internet research*, vol. 15, no. 2, pp. 195–219, 2005.
- [47] J. D. Winkler, D. E. Kanouse, and J. E. Ware, "Controlling for acquiescence response set in scale development," *Journal of Applied Psychology*, vol. 67, no. 5, p. 555, 1982.
- [48] T. Fullerton, C. Swain, and S. Hoffman, *Game design workshop: Designing, prototyping, & playtesting games*. CRC Press, 2004.

- [49] "Bcs code of conduct," 2019. [Online]. Available: <https://www.bcs.org/membership/become-a-member/bcs-code-of-conduct/>
- [50] "Guide to the general data protection regulation," 2018. [Online]. Available: <https://www.gov.uk/government/publications/guide-to-the-general-data-protection-regulation>

APPENDIX A

ADDITIONAL MATERIALS

Link to Repository:

<https://AC200905@gamesgit.falmouth.ac.uk/scm/s192013teamicop/punkey.git>

Currently the computing artefact has been created using the Team project as a base. I have been working on my dynamic difficulty algorithm in the dynamic difficulty branch. The specific script I have made for this artefact is AIDirector.cs.

Figure 4 shows the current iteration of the consent form. While currently it is designed to be a paper form, it may be advisable that I implement it into the beginning of the artefact itself for simplicity's sake.

Comparing Difficulty Systems in a Top-Down Shooter

User Number: 0

Adrian Carter, undergraduate on the BSc Computing for games Course is investigating the effects of different difficulty settings on player experience and would like to invite you to take part. Below is a description of the study and what taking part would involve. All participants have been randomly chosen from within the Games Academy.

For this study, you will be asked to play two short levels of a 2D top-down shooter game. Each level may be using a different difficulty setting I have created, otherwise the levels will be precisely similar. Each level should last less than 3 minutes. After each level, you will be asked to complete a short multiple choice questionnaire regarding your experience of the level. During gameplay, statistics such as 'player deaths' and 'enemies killed' will be gathered. From this data, I will evaluate the effectiveness of each difficulty setting. The purpose of this study is to gather insight into the design of different difficulty systems and how they affect player experience.

All data gathered is anonymous and as such cannot be used to identify you. Your participation in this study is subject to the British Computing Society Code of Conduct and the Code of Research Ethics of the higher education institution in which you are enrolled: Falmouth University. Data will be archived for the maximum period of time permitted under these regulations: 2 years.

Please note, it is not compulsory to take part in this study, and you can withdraw at any time without consequences.

If you have any queries or concerns at any time, feel free to contact ac200905@falmouth.ac.uk. If you have any complaints regarding the ethical aspects of this project, please contact the Falmouth University Research & Development Office at research@falmouth.ac.uk.

I have read the information provided above and I understand that the study is being conducted for the sole purpose of academic research. I acknowledge that participation is voluntary and any data I provide will be treated anonymously, confidentially and will be subject to rigorous data protection. I am also happy for my contributions to appear in a scholarly publication, on the condition that all such contributions are anonymous. Therefore, I consent to my involvement:

☐ I Consent

Fig. 13. Consent Form

Concerning the level you just played. Please mark how strongly you would agree with each statement.

<i>I felt skilled</i> Not at all Slightly Moderately Fairly Extremely 	<i>I felt bored</i> Not at all Slightly Moderately Fairly Extremely 	The blue slider icon marks your previously stated feelings on each statment	
<i>I thought it was too hard</i> Not at all Slightly Moderately Fairly Extremely 	<i>I felt frustrated</i> Not at all Slightly Moderately Fairly Extremely 		
<i>It became too difficult to enjoy</i> Not at all Slightly Moderately Fairly Extremely 	<i>I felt pressured</i> Not at all Slightly Moderately Fairly Extremely 		<i>I enjoyed it</i> Not at all Slightly Moderately Fairly Extremely
<i>I was able to overcome being swarmed</i> Not at all Slightly Moderately Fairly Extremely 	<i>I had to put a lot of effort into it</i> Not at all Slightly Moderately Fairly Extremely 		<i>I felt challenged</i> Not at all Slightly Moderately Fairly Extremely

Press ENTER to Submit Data and Play Next Level

One of the levels, neither of the levels or both of the levels you just played implemented dynamic difficulty adjustments to try and match the level of challenge to your skill level.

<i>Which do you think used dynamic difficulty?</i> First Level Second Level Neither Levels Both Levels 	<i>Which controller setup do you prefer to play video games with?</i> Keyboard + Mouse Gamepad No Preference
	<i>How many hours a week do you play video games? (Rounded to the nearest hour)</i> 0

Fig. 14. In-Game Questionnaire