MASTER THESIS

Adaptive Checkpoints for Flow State Optimization in Video Games

Author:

Zeb De Roover

Supervisor:

Thomas Buijtenweg MSc

Kasper Geeroms MA

This thesis is submitted in fulfilment of the requirements

for the degree of Master of Science in Game Technology

In the programme:

Professional Master Game Technology

Academy for AI, Games and Media

Breda University of Applied Sciences

27 May 2026

DECLARATION OF AUTHORSHIP

I, Zeb De Roover, declare that this thesis titled “Adaptive Checkpoints for Flow State Optimization in Video Games” and the work presented are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date: May 27, 2026

To me, game design and psychology go hand in hand and can meaningfully influence each other, which is something that has been proven again and again. I believe combining these two fields has the potential to create more unique and engaging player experiences, which feels especially important in an ever-expanding market, particularly with the rise of indie games.

This thesis is a stepping stone for me as a game designer and developer. I hope it contributes, even modestly, to the advancement of game design and player experience research.

Zeb De Roover

ABSTRACT

Academy for AI, Games and Media

Master of Game Technology

Adaptive Checkpoints for Flow State Optimization in Video Games

By Zeb De Roover

Dynamic Difficulty Adjustment (**DDA**) adapts game difficulty to maintain a balance between challenge and player ability. Flow Theory describes deep immersion and energized focus emerging when challenge matches skill. Both centre on this challenge-skill balance, and Flow Theory has informed some DDA systems accordingly. Checkpoints, a game mechanic determining where a player respawns upon failure, have not been examined as an adaptive element despite extensive DDA research. This study addresses he following question: to what extent do adaptive checkpoint systems affect the flow channel compared to fixed checkpoints across diverse player experience levels? A between-subjects design was used with 61 participants, a control group with fixed checkpoints, and an experiment group with adaptive checkpoints, in a modified version of the 2D platformer *VVVVVV*. Gameplay was tracked via telemetry and flow was measured using the Flow Short Scale, completed immediately after play. Independent samples t-tests compared groups overall and across three experience-based subgroups. No significant difference in overall flow intensity was found between groups. The experiment group accumulated significantly more deaths, spent more time, and activated fewer checkpoints, indicating a more demanding experience. Participants with general game experience but limited platformer experience in the experiment group reported significantly higher frustration ($p = .002$), while those with the least overall experience reported significantly higher enjoyment ($p = .006$), suggesting experience-dependent effects. The system demonstrated partial validity at experience extremes and warrants further investigation in games designed with adaptive checkpoints from the ground up.

ACKNOWLEDGEMENTS

I hereby declare that in writing this thesis, I made use of the AI-assisted tool Claude AI, as provided by BUas, for language and grammatical corrections, rephrasing, assisting in the programming of the game experiment, and the summarising of references. It was used as a tool for aid, and all ideas, content and conclusions are my own.

First and foremost, I would like to thank my supervisors for the feedback and help throughout the development of this thesis. I am grateful for the useful near-weekly meetings with Kasper and the interesting lectures and guiding from Thomas, as well as all other teachers and personnel at BUas who aided me.

I want to give a special thanks to everyone who took the time to participate in my experiment and spent the time to play through the game, even when the game proved to be challenging for them. These people helped make this research possible and allowed it to come to the conclusion that it did. I also want to thank Terry Cavanagh, the developer of *VVVVVV*, who gave me the permission to use his game as a basis for my experiment game.

I would like to give an extra special thanks to those in my personal life who continuously supported me throughout, as we navigated a difficult period together too. I want to thank my mother and my father, as well as my stepmother and stepfather. Lastly, I would like to thank my little sister.

CONTENTS

Declaration of Authorship ii
Abstract iv
Acknowledgements v
Contents vi
List of Figures viii
List of tables ix
List of Abbreviations x
1 Introduction 1
2 Literature Review: Flow, Failure, and Adaptability 3
2.1 Flow Theory in Digital Games 3
2.1.1 Theoretical Foundations 3
2.1.2 Applying Flow to Digital Games 6
2.1.3 Critiques of Flow and Design Implications 7
2.2 Dynamic Difficulty Adjustment 8
2.2.1 Understanding Dynamic Difficulty Adjustment 9
2.2.2 Approaching Adaptive Design 10
2.2.3 Critiques and Limitations 12
2.2.4 Notable DDA Applications 14
2.3 Checkpoint Systems 16
2.3.1 Defining Checkpoints 17
2.3.2 Checkpoints in Research 19
2.4 The Paradox of Failure 20
2.4.1 Why Players Need Failure 21
2.4.2 Failure as a Design Tool 22
2.5 Synthesis and Research Gap 23
2.6 Ethical Considerations 26
2.6.1 Ethical Implications of Flow-Based Game Design 26
2.6.2 Transparency and Agency for DDA 27
3 Methodology: Designing an Adaptive Checkpoint System 29
3.1 Approaching the Research Question 29
3.2 The Game 30
3.3 Designing the Systems 34
3.3.1 Influential Methodologies 34
3.3.2 The Adaptive Checkpoint System Design 36

3.3.3 Tracking Player Behaviour ... 42
3.4 The Post-Play Questionnaire ... 43
3.5 Ethical Considerations and Methodological Limitations ... 44
3.6 Experiment Procedure and Analysing Data ... 46
4 Experiment Data and Results ... 48
4.1 Data Collection ... 48
4.2 Analytical Approach ... 49
4.3 Descriptive Statistics ... 49
4.4 Hypothesis 1: Flow ... 51
4.5 Hypothesis 2: Completion ... 53
4.6 Hypothesis 3: Frustration ... 54
4.7 Exploratory Analysis ... 57
5 Discussion and Interpretation ... 59
5.1 Hypothesis 1: Flow ... 59
5.2 Hypothesis 2: Completion ... 60
5.3 Hypothesis 3: Frustration ... 61
5.4 Post-hoc Validation ... 63
5.5 Limitations ... 65
5.6 Implications ... 65
6 Conclusion and Implications ... 67
6.1 Findings and Conclusion ... 67
6.2 Contribution ... 68
6.3 Limitations ... 68
6.4 Future Directions ... 69
Appendix A ... 70
Extended Game Description ... 70
Adaptive Checkpoint Locations ... 71
Appendix B ... 72
Online Files ... 72
References ... 73

LIST OF FIGURES

Figure 2.1 Diagram Showing the Basics of Flow Theory 5
Figure 2.2 Flow State Model 5
Figure 2.3 Chen's Flow Diagram: Different Players Have Different Flow Zones 7
Figure 2.4 Chen's Graph Depicting Flow-Based Player Choices 12
Figure 2.5 Common Infected During the Build Up Phase in Left 4 Dead (2008) 14
Figure 2.6 Nanotech Crate From Ratchet and Clank (2002) 15
Figure 2.7 Aku Aku Mask from Crash Bandicoot (1996) 16
Figure 2.8 Checkpoint Crate from Crash Bandicoot (1996) 16
Figure 2.9 Lamp Post Checkpoint in Sonic the Hedgehog (1991) 18
Figure 2.10 Midway Checkpoint Flag from New Super Mario Bros. U Deluxe (2019) 18
Figure 2.11 Player Resting by a Bonfire in Dark Souls (2011) 19
Figure 2.12 Tally Screen in Shovel Knight (2014) Displaying Active, Broken and Inactive Checkpoints 19
Figure 2.13 Death Screen from Hades (2020) 23
Figure 3.1 VVVVVV game banner 31
Figure 3.2 The last room of the tutorial for the experimental version. After this the DDA system activates 32
Figure 3.3 The teleporter at the end of the first level, "Space Station". 33
Figure 3.4 A Gravity Wall placed above spikes in "The Laboratory". 33
Figure 3.5 A chart showing the different conditions for struggling. 37
Figure 3.6 A chart visualising the process of the DDA checkpoint system. 38
Figure 3.7 An example of the different amounts of checkpoints that can be present in a room based on current difficulty 40
Figure 3.8 The first room with DDA active. The player enters from above and heads right to Comms Relay, entering again via the bottom-right where DDA is activated. 41
Figure 3.9 Example data sent to the spreadsheet 43
Figure 4.1 Histograms of total deaths between the two groups 50
Figure 4.2 Histograms of total time between the two groups 51
Figure 4.3 FSS Total score distributions by group and experience level displayed as a boxplot. Boxes represent the interquartile range and horizontal lines indicate the median. 52
Figure 4.4 Checkpoints activated by group and experience level 54
Figure 4.5 Frustration scores by group and experience level 56
Figure 4.6 Enjoyment scores by group and experience level 56
Figure 4.7 Additional checkpoint desire comparison between groups and experience level 57
Figure 4.8 Difficulty increases by experience level in the experiment group 58
Figure 4.9 Difficulty decreases by experience level in the experiment group 58
Figure 5.1 Section of the post-hoc validation spreadsheet depicting the calibrated thresholds and the P70 values of both groups 63

LIST OF TABLES

Table 4.1 Descriptive game and questionnaire data comparison between control and experiment groups 49
Table 4.2 Independent samples t-test results for FSS Total by group and experience level 51
Table 4.3 Two-way ANOVA interaction effects (experience level × group) on FSS subscales 52
Table 4.4 Completion rates by group 53
Table 4.5 Behavioural measure comparisons between groups 53
Table 4.6 Frustration and enjoyment comparisons between groups overall and by experience level 54
Table 4.7 Additional checkpoint desire comparison between groups and experience level 55

LIST OF ABBREVIATIONS

DDA	Dynamic Difficulty Adjustment
FSS	Flow Short Scale
KPM	Key presses per minute
P70	70th Percentile

Dedicated to all who supported me throughout.

1 INTRODUCTION

Since the early stages of arcade games, video game difficulty has evolved from challenges designed to milk players for their quarters for maximising revenue, to sophisticated systems trying to provide catered experiences. As technology advanced, the ability to save progress revolutionised how players experienced challenge, with newer home consoles allowing more flexible approaches. Yet despite these advances, difficulty remains fundamental to gaming's appeal. The satisfaction of overcoming obstacles continues to drive player engagement across all genres, yet providing the right level of difficulty for every player remains one of game design's most persistent challenges.

Today's designers still face the fundamental question that existed decades ago: how do we create solid experiences that engage both casual gamers and experienced fanatics? With beginners wanting to just have a fun experience and hardcore gamers wanting to be challenged while displaying their skills, providing appropriate challenge for all player types is a persistent design problem. Difficulty misalignment with players, whether too easy or too challenging, remains a primary reason players abandon games. When players feel overwhelmed by challenge or bored by repetitive simplicity, they disengage. Dynamic Difficulty Adjustment (**DDA**) has existed since the arcade era, with games adjusting elements like enemy behaviour based on player performance. As the gaming industry expanded and technology advanced, DDA systems gained traction among developers seeking to accommodate increasingly diverse player bases and to address this issue.

DDA implementation varies widely across games and genres. Some systems rely on simple metrics like death counts, while others employ sophisticated analysis of player choices and movement patterns using artificial intelligence. However, many successful DDA implementations share a common theoretical foundation: Flow Theory. First introduced by Mihaly Csikszentmihalyi in 1975, flow describes a state of energised focus and deep enjoyment that emerges when challenge matches skill level. This optimal experience, often equated with fun in gaming contexts, has become central to psychologically-informed game design. Extensive research has documented and validated Flow Theory's application to games, establishing its value for creating engaging player experiences. This research is motivated by a personal interest in Flow Theory and its potential application to game design, particularly in how it might improve player experiences. This personal interest in Flow Theory, and how DDA might create conditions for sustained engagement, motivates the focus of this research on an overlooked element: checkpoints.

The game industry has implemented DDA across various genres. Notable examples include *Left 4 Dead (2008)*, *Ratchet & Clank (2002)* and *Crash Bandicoot (1996)*. Most DDA systems adjust parameters such as enemy health, damage output, or spawn rates, leaving checkpoints largely static by comparison. The exception is *Crash Bandicoot*, where players experiencing repeated failures receive help through the transformation of collectible crates into additional checkpoint crates. By controlling what sections a player needs to replay, checkpoints directly control the temporal cost of failure, which is the primary cause of player frustration.

Checkpoint design has not seen remarkable innovation throughout gaming history. Most implementations follow well-known patterns: placing checkpoints at level midpoints or right before or after difficult segments, such as boss encounters. Despite this body of research, existing research on DDA and flow in games has not examined how checkpoints could be integrated with Flow Theory to enhance this overlooked feature. The absence of research on adaptive checkpoint systems represents a significant gap, as checkpoints fundamentally shape how players experience failure and recovery, both of which are critical components of game difficulty and flow maintenance.

This thesis investigates the potential of adaptive checkpoint systems based on the principles of Flow Theory. The research addresses the following question: **To what extent do adaptive checkpoint systems affect the flow channel compared to fixed checkpoints across diverse player experience levels?**

The following chapter presents a literature review covering Flow Theory, DDA, checkpoint systems, and the role of failure in games, concluding with a synthesis that establishes the research gap and framework for this study.

2 LITERATURE REVIEW: FLOW, FAILURE, AND ADAPTABILITY

This chapter examines the theoretical foundations and existing research relevant to adaptive checkpoint systems. The review begins with Mihaly Csikszentmihalyi's Flow Theory as the psychological framework for understanding optimal player experience. Following this is Dynamic Difficulty Adjustment as the technical approach to maintaining appropriate challenge levels. Subsequent chapters explore checkpoint design, the role of failure in games, and how these elements can be synthesised into an adaptive system. The chapter concludes with ethical considerations that must guide responsible implementation.

2.1 FLOW THEORY IN DIGITAL GAMES

Flow is being completely involved in an activity for its own sake. The ego falls away. Time flies. Every action, movement, and thought follows inevitably from the previous one, like playing jazz.
— Mihaly Csikszentmihalyi, in an interview with *WIRED* (1996)

Flow Theory provides the psychological foundation for understanding how players maintain engagement during gameplay, making it central to designing effective adaptive systems.

2.1.1 Theoretical Foundations

Mihaly Csikszentmihalyi introduced Flow Theory first in 1975 with his book *Beyond Boredom and Anxiety*. Through extensive research involving artists, athletes, and professionals, he later identified Flow as a distinctive psychological state characterised by complete absorption and intrinsic motivation, or simply “the optimal experience”. However, flow represents more than mere enjoyment or satisfaction. Flow describes moments of deep, energised focus where consciousness becomes fully ordered and aligned with the activity at hand.

These flow experiences, Csikszentmihalyi argued, represent the most fulfilling moments in human life and serve as the foundation for growth and creativity. Crucially, flow requires voluntary participation. Without the intrinsic motivation, forced activities rarely produce this state. When information conflicts with our intentions or goals, we experience what Csikszentmihalyi termed “psychic entropy”. This describes a state of inner disorder, or mental chaos. Flow represents the opposite: a state where psychic energy flows freely without internal conflict, creating order in consciousness.

The phenomenology of flow encompasses eight distinct characteristics that collectively define this optimal experience:

- **A task with reasonable chance of completion**: The activity must be achievable within the individual's current or developing capabilities, providing a sense of possible success
- **Clear goals**: Well-defined objectives that provide direction and purpose, allowing individuals to understand exactly what needs to be accomplished and how
- **Immediate feedback**: Continuous information about progress and performance, enabling real-time adjustments and maintaining engagement
- **Deep but effortless involvement**: Complete immersion that removes awareness of everyday worries and frustrations, creating what Csikszentmihalyi called energised focus
- **Sense of control**: The paradoxical feeling of being in command of the situation while remaining fully responsive to its demands and challenges
- **Loss of self-consciousness**: The disappearance of self-awareness as the boundary between self and activity dissolves, losing sense of physiological needs such as hunger or fatigue
- **Transformation of time**: Altered temporal perception where hours pass like minutes or seconds expand into detailed awareness
- **Autotelic experience**: The activity becomes intrinsically rewarding, pursued for its own sake rather than external outcomes or rewards

Central to Flow Theory is the dynamic relationship between perceived challenges and personal skills. As illustrated in the graph in Figure 2.1, flow occurs within a specific channel where both challenge and skill operate above an individual's average levels. This threshold requirement distinguishes flow from simple task engagement. The flow channel represents the optimal corridor between anxiety (when challenges exceed skills) and boredom (when skills surpass challenges). The model evolved from Csikszentmihalyi's original three-state conception to include eight distinct experiential channels referred to as the Flow State Model (see Figure 2.2), representing the full spectrum of mental states from apathy to flow.

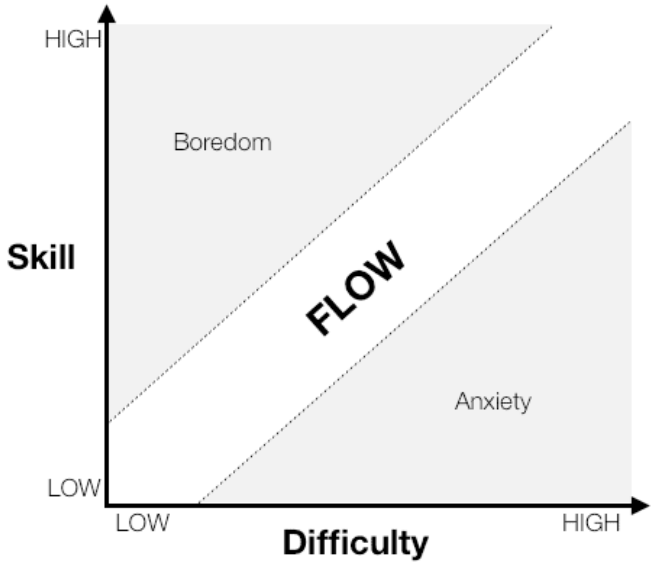

Figure 2.1 Diagram Showing the Basics of Flow Theory

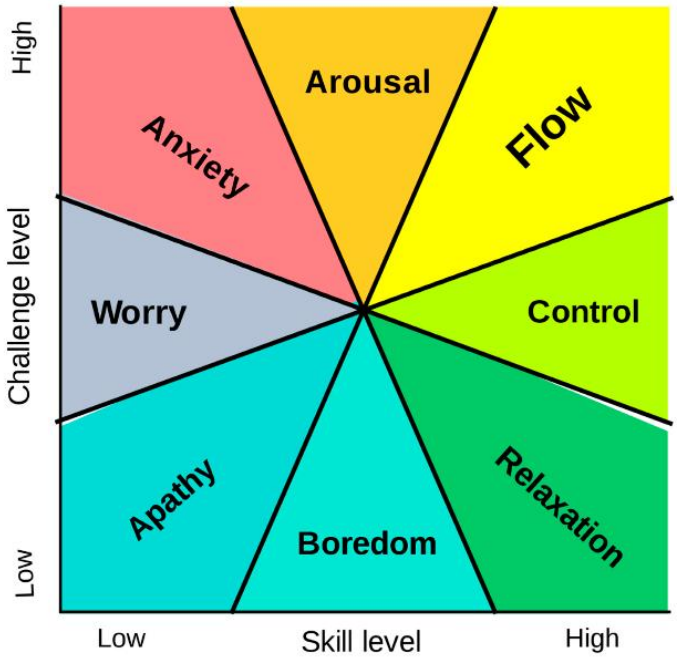

Figure 2.2 Flow State Model

This balance requires constant adjustment as individuals improve their skills and face new challenges. Flow Theory states that optimal experiences promote personal growth through two processes: differentiation (the development of increasingly complex skills through challenge) and integration (the unification of attention and action into a coherent, focused state). As skills increase, individuals seek greater challenges, creating a cycle of growth that makes flow both rewarding and developmentally beneficial. This dynamic is directly relevant to game design, where maintaining this cycle is the central challenge.

2.1.2 Applying Flow to Digital Games

The application of Flow Theory to video games gained significant academic attention with Sweetser and Wyeth's GameFlow model in 2005. They developed a structured framework that mapped Csikszentmihalyi's eight flow elements directly to concrete game design criteria:

- **Challenge**: Difficulty matching player skill levels
- **Clear Goals**: Missions and objectives guiding gameplay
- **Feedback**: Immediate visual and audio responses
- **Concentration**: Rich multi-sensory stimuli maintaining player attention
- **Control**: Player agency over actions and outcomes
- **Immersion**: Emotional involvement and loss of self-awareness
- **Player Skills**: Learning through play and skill development
- **Social Interaction**: Competition and cooperation (although not part of the original flow model, they argue this is crucial for the context of games)

This framework demonstrates the concrete applicability of Flow Theory to games, setting the stage for Chen's more design-focused interpretation. Sweetser and Wyeth's validation study comparing high and low-rated real-time strategy games demonstrated that successful games excel at creating these flow conditions, while poorly designed games fail to maintain the delicate balance required for optimal experience.

Flow equals fun in the context of video games, according to Chen's 2006 thesis *Flow for Games*, making maintenance of flow the primary design goal. He provided a practical implementation of Flow Theory in game design, building on the theoretical foundation by Sweetser and Wyeth. Chen identified a critical flaw in passive adjustment systems: they cannot accurately read player mental states through limited controller inputs. Instead, Chen advocated for player-oriented active systems where players navigate their own flow experience through embedded choices. This approach respects player agency while allowing natural self-adjustment when players feel bored or frustrated. Chen highlighted three fundamental conditions derived from Flow Theory for applying flow to games: intrinsic rewards, appropriate challenge-skill balance, and player sense of control.

Chen further reconceptualised flow as a dynamic, wave-like experience rather than a static state. He observed that players naturally fluctuate in and out of the flow channel during gameplay, as illustrated in Figure 2.3, with each player having a different Flow Zone comparable to a fingerprint, based on their individual skill. These exits occur because of curiosity for exploration, finishing a level, or seeking different experiences. These reactions represent natural engagement patterns, rather than design failures. Chen proposed expanding a game's "flow coverage" to accommodate diverse skill levels within the same experience. This

theoretical framework shaped Chen's approach to designing adaptive systems that serve multiple player types simultaneously without compromising individual flow experiences.

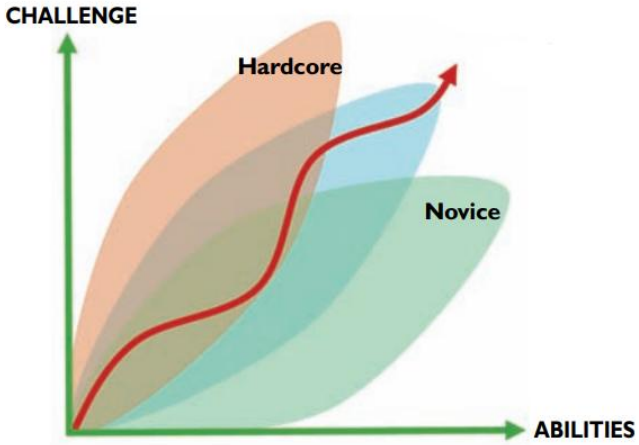

Figure 2.3 Chen's Flow Diagram: Different Players Have Different Flow Zones

Cowley et al. (2008) independently examined games as flow-inducing systems through their User-System-Experience (USE) model. They demonstrated that games function as information systems that create and resolve psychic entropy through structured challenges. Their analysis revealed that games facilitate flow more readily than traditional activities due to low investment thresholds, structured progression, and risk-free environments. For adaptive design, they proposed that understanding flow could inform real-time difficulty adjustments and dynamic content modification.

2.1.3 Critiques of Flow and Design Implications

Against Flow: video games and the flowing subject presents a provocative critique of Flow Theory's application to games. The central argument is that flow functions as an ideological framework that shapes player experiences in specific, limiting ways (Soderman, 2021). These arguments, however, should be understood as one perspective in an ongoing academic debate rather than as established fact, and do not undermine the practical value of flow as a design framework when applied with appropriate nuance.

Soderman argues that Flow Theory promotes a singular "correct" way to play games, marginalising other forms of engagement. He contends that the flow framework reduces games to delivery systems for continuous skill accumulation, suppressing valuable experiences including social connection, aesthetic appreciation, and reflective engagement. While Sweetser and Wyeth's GameFlow model explicitly includes social interaction as a flow component, contradicting Soderman's concern about social play, his critique of flow's blindness to aesthetic and contemplative experiences remains relevant.

The growth imperative embedded in Flow Theory presents particular problems. The classic flow diagram's upward trajectory assumes constant progression, reducing gameplay to endless challenge escalation. This growth-focused framework exemplifies how the aesthetic and contemplative experiences mentioned earlier become casualties of flow's focus. When players enter Soderman's described "tunnel vision," they lose not only self-awareness but also the ability to appreciate a game's artistic qualities, atmospheric details, or moments of quiet reflection. This loss of self, while central to flow, raises questions about letting players "zone out". Does completely zoning out allow for the formation of meaningful memories that are often integral to a memorable gaming experience?

Games benefit from deliberately designed breaks that interrupt flow, allowing players to reflect and form lasting memories of their experience. Alternating between flow-inducing challenges and other activities like exploration, storytelling, or social interaction creates richer, more varied experiences. Rather than treating flow as the sole design goal, developers can offer it as one experience among many.

Soderman does not advocate abandoning flow entirely but rather developing critical awareness of its limitations while applying it to game design. For adaptive game design, this means recognising that some players may not want or benefit from constant flow states. Systems truly responsive to player needs should accommodate those who prefer contemplative exploration, social interaction, or narrative engagement over challenge-based flow.

While Soderman's critique may overstate flow's rigidity, it raises valid questions about uncritical implementation. The theory's emphasis on continuous growth and challenge creates narrow experiential frameworks if applied dogmatically. However, Flow Theory's evolution from Csikszentmihalyi through GameFlow shows it can adapt to include diverse elements like social play. The key is implementing flow as a tool rather than a doctrine, creating games that offer flow experiences while respecting players who seek different forms of engagement. The broader ethical implications of these design choices, including concerns about addiction and player wellbeing, are examined in Subchapter 2.6.

2.2 Dynamic Difficulty Adjustment

> *Our mantra became help weaker players without changing the game for the better players. We called all this DDA, Dynamic Difficulty Adjustment... Good player, bad player, everyone loved Crash games. They never realized it is because they were all playing a slightly different game, balanced for their specific needs.*
>
> — Jason Rubin, Co-creator of *Crash Bandicoot*, in *All Things Andy Gavin* (2011)

Dynamic Difficulty Adjustment (**DDA**) represents game design's response to a fundamental problem: players have vastly different experience levels and skill sets, yet most games offer only fixed difficulty settings. DDA systems automatically modify game parameters based on player performance, aiming to eliminate both the frustration of excessive challenge and the boredom of insufficient difficulty. This pursuit of personalised gaming experiences promises to keep all players engaged, regardless of their individual abilities.

2.2.1 Understanding Dynamic Difficulty Adjustment

Traditional games handle difficulty through various systems, each with their own limitations. Games without difficulty options present a single challenge that may prove too easy for experienced players or impossibly hard for newcomers. Games that do offer difficulty selection often require players to choose before they even understand the game's mechanics. Even when games allow difficulty changes during play or gradually increase challenge as players progress, these systems cannot account for individual learning curves, temporary skill regression after breaks from playing, or the wide variance in player abilities. DDA systems, on the other hand, offer an alternative by continuously monitoring how players perform and adjusting the game's difficulty in real time. This adaptive approach aims to make games more accessible to less experienced players while maintaining appropriate challenge for skilled ones, particularly valuable in games that traditionally offer no difficulty options.

The need for DDA stems from how players actually develop skills during gameplay. According to *Game Feel: A Game Designer's Guide to Virtual Sensation*, skill improvement is not a smooth, predictable process. While players objectively get better at games over time, their subjective feeling of control varies dramatically based on whether the current challenge matches their ability. Designers can track player skill progress through levels, inventory, etc. to understand three key elements: what is currently being learned, what has already been mastered and what game elements are yet to be encountered? Players experience games differently when learning new mechanics compared to using skills they have already mastered (Swink, 2009). This insight reveals where static difficulty may fall short: it cannot respond to these natural fluctuations in player capability and focus.

In practice, DDA systems work by modifying specific game elements based on player performance. *Game Level Design* (2005) identifies several adjustment methods. These include changing how much damage enemies can take, adding or removing enemies from encounters, and adjusting how many health packs or ammunition boxes appear in levels (Byrne, 2005). In other words, the system follows simple logic: when players die repeatedly or struggle with a section, the game reduces difficulty accordingly. When players breeze through content, it increases the challenge. The crucial element is that these adjustments happen invisibly. Players

should never realise the game is helping them, as this would undermine their sense of achievement and make them feel inferior. The DDA system can do this by subtly changing the current encounter or adapting future encounters to preserve this invisibility of adaptation.

Understanding what makes DDA effective requires examining both theory and evidence. Hunicke's (2005) pioneering empirical research established three core principles for DDA design. First, the system must preserve game balance and prevent players from exploiting it for easy victories. Second, it must maintain consistent rules so players can still learn from their mistakes. Third, and perhaps most importantly, adjustments must be unobtrusive enough that players remain unaware of them, the same conclusion also reached by Byrne. These principles guided Hunicke's development of the Hamlet system, which used mathematical probability calculations to predict when players might die and adjusted weapon and health availability accordingly. They designed this research using a custom game created with the *Half-Life (1998)* engine.

Hunicke's research provided early validation that DDA systems improve player performance without conscious detection, establishing the viability of "change blindness" in game difficulty adjustment. The study, however, also revealed a crucial limitation: DDA effectiveness varied significantly with player skill level, with novice players showing little benefit from subtle adjustments compared to experienced players. This finding, though limited to first-person shooters, argues that successful DDA implementation requires different strategies for different player groups rather than a one-size-fits-all approach.

2.2.2 Approaching Adaptive Design

The implementation of DDA has evolved through various approaches, each attempting real-time difficulty adaptation. Mortazavi et al.'s (2024) systematic review categorises contemporary DDA systems into rule-based approaches, found in 34 papers, and data-driven approaches, documented in 42 papers. Rule-based systems use simple performance thresholds to trigger difficulty changes, while data-driven approaches employ machine learning techniques (Mortazavi et al., 2024). Zohaib's (2018) review further details these categories, identifying probabilistic methods, neural networks, dynamic scripting, and reinforcement learning among others. The trend, as both reviews note, has shifted from heuristic-based to data-driven approaches, particularly since 2011, with future systems likely to leverage advanced AI for sophisticated player modelling.

Regarding data collection, Mortazavi et al. found that 46 papers relied on gameplay data such as scores, completion times, and death counts, while only 19 papers used physiological data like EEG or heart rate monitoring. This preference for gameplay data stems from its efficiency and non-intrusiveness, as physiological approaches, while promising in research

settings, remain unrealistic for consumer applications where players cannot be expected to wear sensors during normal play. Many existing DDA systems use death counts as their primary metric for difficulty adjustment, though this simplistic approach fails to capture the complexity of player experience. Simply using specific death milestones for the design of an adaptive system does not take the varying and distinct player profiles into account, as discussed throughout Subchapter 2.1.

While these technical approaches vary in implementation, they share a common limitation: most focus on preventing failure through automatic adjustments rather than enhancing player experience. Mortazavi et al. document that DDA evaluation uses metrics including enjoyment, motivation, engagement, immersion, and notably, flow. This indicates that existing research primarily adjusts parameters like enemy strength, speed, and numbers, with systems designed to reduce player failure rates. The prominence of flow as an evaluation metric reflects its importance in DDA research, with some rule-based systems specifically using Flow Theory to guide their adjustments. Denisova and Cairns (2015) demonstrated this potential by using Flow Theory to guide their adaptive timing system. Their study found that players with flow-based adjustments scored more consistently around target goals and showed significantly higher immersion scores. Most importantly, players perceived no difference in challenge despite experiencing objectively different difficulty levels (Denisova and Cairns, 2015). This validates that maintaining challenge-skill balance through subtle adaptations can improve engagement without undermining the authentic game experience. While many rule-based systems automatically adjust difficulty based on failure rates, flow-oriented implementations like those by Denisova and Cairns or Chen seek to maintain optimal engagement while preserving player control.

Chen's analysis reveals why traditional DDA often fails to maintain flow. System-oriented DDA relies on flawed assumptions: it cannot access players' mental states through controllers, performance metrics like death counts do not equal flow experience, and repeated deaths may indicate experimentation rather than struggle. As Chen notes, one designer's adjustment preferences cannot suit a mass audience. His solution reconceptualises DDA as player-oriented design that provides choices rather than automatic adjustments. This represents a specific rule-based approach that uses flow maintenance as its guiding principle.

Chen's embedded choice system offers a specific method for implementing flow-based DDA. His approach solves a key paradox: players need options to maintain flow, but obvious choices disrupt the experience. By integrating difficulty selection into core gameplay mechanics, players unconsciously navigate toward their optimal challenge level. Chen argues this maintains the essential flow requirement of player control while avoiding the interruptions of menu-based

selections or the loss of agency in automated systems. His game *flOw (2006)*, specifically designed for his research, validated this approach, demonstrating that when difficulty choices become part of natural gameplay, players maintain better flow states. Chen visualised players' flow experiences with a graph (Figure 2.4) that illustrates how different choices produce different outcomes.

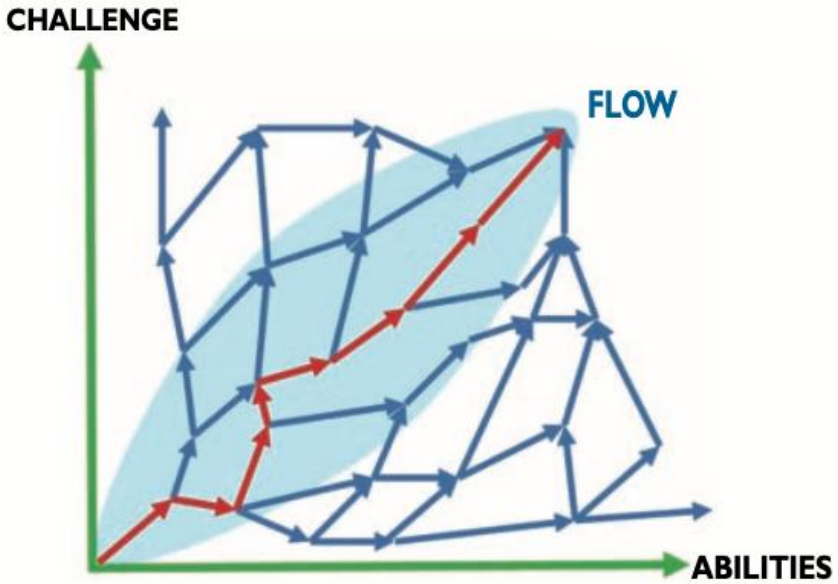

Figure 2.4 Chen's Graph Depicting Flow-Based Player Choices

Flow-based approaches demonstrate particular promise, as documented by Mortazavi et al. They treat difficulty as an experience to optimise rather than merely a problem to solve, focusing on maintaining the challenge-skill balance that characterises optimal player experience. Traditional approaches focus on maintaining specific failure rates or performance metrics. Flow-based approaches recognise that optimal experience requires player agency and control. Flow-based systems allow different players to progress at their own pace while maintaining engagement, whether through embedded choices, adaptive timing, or other mechanisms that preserve player agency. This shift from system-oriented to player-oriented DDA addresses the fundamental issue identified by both Mortazavi et al. and Zohaib: no single adjustment strategy works for all players or game types. By focusing on flow maintenance rather than simple failure prevention, these DDA approaches offer a more flexible and effective method for adaptive game design.

2.2.3 Critiques and Limitations

While DDA promises to create personalised gaming experiences, significant critiques have emerged regarding both its implementation and underlying assumptions. These concerns

range from fundamental design philosophy issues to practical limitations that undermine its effectiveness.

Schell (2008) identifies three fundamental problems with DDA systems. First, DDA spoils the reality of the game world. When challenges adjust relative to player performance, it damages the illusion of facing fixed, consistent opponents. Players may notice that enemies become weaker after they die repeatedly, breaking immersion and receiving the impression that the world exists solely to accommodate them rather than presenting authentic challenges. Second, these systems are inherently exploitable. Players can intentionally perform poorly to trigger easier gameplay in upcoming sections, turning difficulty adjustment into an unintended meta-game strategy. Third, DDA conflicts with how players naturally improve through practice. Many players derive satisfaction from mastering fixed challenges through repetition. When the game automatically reduces difficulty after failures, it can feel insulting or disappointing, robbing players of the satisfaction that comes from eventually overcoming a tough obstacle through persistence and skill development.

Rethinking dynamic difficulty adjustment for video game design by Guo et al. (2024) reveals a deeper problem: many current DDA systems rely too heavily on Flow Theory. Their literature review of 231 papers found that this flow-based approach creates significant limitations. Flow Theory itself presents ambiguous concepts with only subjective measurement methods, and research shows only moderate relationships between challenge-skill balance and actual flow states. This theoretical foundation provides limited guidance for diverse design goals beyond achieving "balanced challenge." As a result, many DDA implementations oversimplify their goal to maintaining a 50% success rate, which may not align with what players actually want from their gaming experience. However, this does not mean flow is useless for DDA design. To properly apply flow to DDA, designers should focus on maintaining player agency and control rather than just balancing challenge and skill (Guo et al., 2024).

The invisible nature of most DDA systems creates additional trust and transparency issues. Fisher and Kulshreshth's research highlights how hidden adjustments can feel manipulative or unfair to players. When difficulty changes occur without player awareness or consent, it creates what they term "invisible punishment" that can damage player confidence and agency (Fisher and Kulshreshth, 2025). While DDA should remain invisible during gameplay to preserve immersion, games should consider including disclaimers about the presence of adaptive systems, or better yet, provide options to disable DDA entirely.

Perhaps most importantly, DDA systems often fail to account for diverse player motivations. Not all players seek optimal challenge-skill balance. Some play games for relaxation, exploration, or narrative experiences rather than challenge. Others deliberately seek

extreme difficulty for the satisfaction of overcoming seemingly impossible obstacles. By assuming all players want dynamically balanced challenges, DDA systems may actually work against player preferences. Guo et al. propose that DDA should move beyond Flow Theory to consider both Objective Game Difficulty (measurable performance metrics) and Subjective Game Difficulty (player's perceived challenge), aligned with specific design goals rather than universal challenge balancing.

Despite these valid critiques, Flow Theory remains a useful framework for DDA when applied thoughtfully, acknowledging its limitations while leveraging its proven benefits for player engagement. Chen's thesis and Denisova and Cairns' research demonstrate that flow-based DDA can be effective when applied with appropriate nuance.

2.2.4 Notable DDA Applications

DDA has been successfully implemented in various forms across commercial games, each taking different approaches to accommodate player skill levels. Three notable implementations demonstrate the diversity of DDA: *Left 4 Dead*'s AI Director, *Ratchet & Clank*'s various parameter adjustments, and *Crash Bandicoot*'s additional checkpoints.

Left 4 Dead's AI Director (2008) represents one of the more sophisticated DDA implementations in commercial games. Rather than adjusting enemy health or damage, the Director orchestrates entire gameplay experiences through procedural pacing. The system cycles through Build Up, Peak, and Relax phases based on player stress and performance metrics, providing different quantities and types of enemies (see Figure 2.5). It maintains a consistent base difficulty to create uniquely different playthroughs without arbitrary stat modifications.

Figure 2.5 Common Infected During the Build Up Phase in Left 4 Dead (2008)

Ratchet & Clank's first three games (2002-2004) pioneered DDA through environmental and enemy parameter manipulation. The original game spawned extra nanotech crates (see Figure 2.6) for struggling players to restore more health, while *Going Commando* (2003) and *Up Your Arsenal* (2004) expanded this further, modifying enemy behaviour and amounts. The developer Insomniac Games referred to this approach as "additive difficulty", adjusting enemy counts and parameters so that both skilled and struggling players received tailored support rather than a straightforward reduction in challenge.

Figure 2.6 Nanotech Crate From Ratchet and Clank (2002)

Crash Bandicoot (1996) by video game developer Naughty Dog includes several features designed to assist players. You may find an Aku Aku mask, visible in Figure 2.7, that provides temporary protection against enemies when picked up. The game also has checkpoints that provide respawning in most of the levels (see Figure 2.8). Naughty Dog also introduced a notable multi-mechanism DDA. As Jason Rubin, developer and co-founder, explained in his development retrospective: "We were already learning. We had realized that if a novice player died a lot of times, we could give them an Aku Aku at the start of a round and they had a better chance to progress. And we figured out that if you died a lot when running from the boulder, we could just slow the boulder down a little each time. If you died too much a fruit crate would suddenly become a continue point. Eventually everyone succeeded at Crash" (cited in Gavin, 2011).

Figure 2.7 Aku Aku Mask from Crash Bandicoot (1996)

Figure 2.8 Checkpoint Crate from Crash Bandicoot (1996)

Among these three DDA implementations, *Crash Bandicoot*'s checkpoint-based system is particularly relevant for this research. By creating additional checkpoints when players died repeatedly, the game used checkpoint placement itself as a difficulty balancing tool. This use of checkpoints as a difficulty adjustment mechanism raises important questions: What exactly defines a checkpoint in game design? How do checkpoints affect player experience beyond their basic save function? The following chapter addresses these questions by examining the concept of checkpoints and exploring the limited academic research on this fundamental game mechanic.

2.3 Checkpoint Systems

BONFIRE LIT

— *Dark Souls* (FromSoftware, 2011)

Checkpoints are one of gaming's most common features, appearing across different genres of games. This chapter examines how checkpoints function in different game types, reviews existing research on their impact on player behaviour and performance, and considers potential areas for innovation.

2.3.1 Defining Checkpoints

Checkpoints are fundamental save mechanics that have evolved significantly since gaming's early days. Checkpoints can be defined as "points in a game level at which the game may be saved or at which the avatar will be reincarnated if he dies" (Adams and Rollings, 2007). This dual functionality of progress preservation and respawn location distinguishes checkpoints from other save systems. Unlike manual saves where players choose when and where to preserve progress, checkpoints operate automatically at predetermined locations, removing player agency in favour of streamlined gameplay flow. When players fail after reaching a checkpoint, the player's avatar reappears at the checkpoint location ready to attempt the challenge again. The game state typically resets to that moment: enemies may respawn and certain collected items can return.

Modern checkpoint design has evolved from the punishing systems of early gaming. Older titles often restarted players at the beginning of entire levels or even the complete game upon failure, a practice Adams and Rollings now consider poor design. Contemporary games handle failure more forgivingly through frequent checkpoints and minimal progress loss. As Byrne emphasises, being forced to replay sections multiple times represents "one of the biggest complaints about games," making thoughtful checkpoint placement crucial for player retention. Even the most enjoyable level becomes tedious with excessive forced repetition.

The positioning of checkpoints represents a critical balancing decision. Too few frustrate players through excessive repetition, while too many remove meaningful challenge and tension. Strategic placement can guide difficulty curves, provide breathing room after intense sequences, and create natural learning opportunities through controlled repetition. The following section examines what academic research has established about these effects.

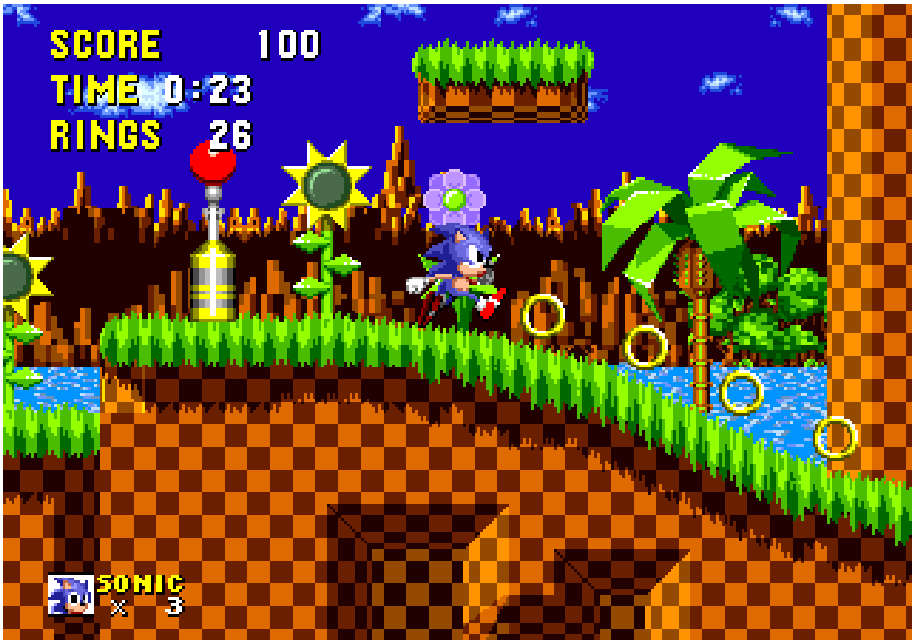

Figure 2.9 Lamp Post Checkpoint in Sonic the Hedgehog (1991)

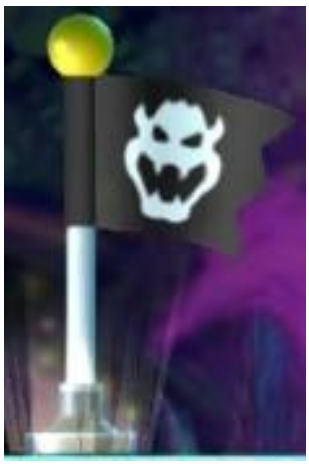

Figure 2.10 Midway Checkpoint Flag from New Super Mario Bros. U Deluxe (2019)

Checkpoint implementations vary considerably across games and genres. Some use visible markers like *Sonic the Hedgehog*'s (1991) lamp posts (Figure 2.9) that change colour when activated, clearly communicating save point locations to players. Others, like *Grand Theft Auto V* (2012), employ invisible mission checkpoints that save progress seamlessly without interrupting gameplay, at the cost of clarity. Another notable platformer implementation is *New Super Mario Bros.* (2006), which uses clearly marked checkpoint flags positioned at level midpoints (Figure 2.10). *Dark Souls*' (2011) bonfire system innovated on checkpoint design by making save points integral to gameplay, transforming them into multifunctional safe havens that serve strategic purposes beyond simple progress preservation (Figure 2.11). *Shovel Knight* (2014) took a different approach by allowing players to destroy checkpoints for extra currency

(Figure 2.12), creating a risk-reward system that lets players choose their own difficulty and stakes.

Figure 2.11 Player Resting by a Bonfire in Dark Souls (2011)

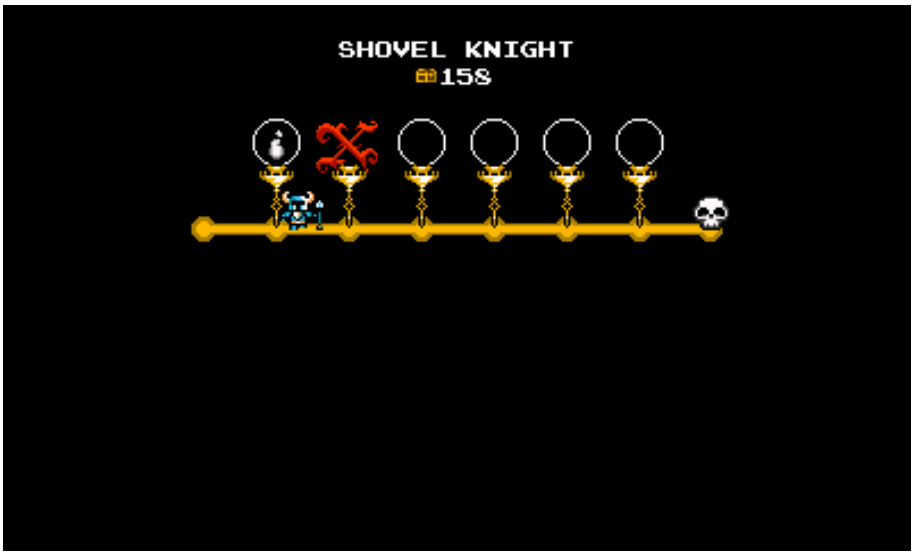

Figure 2.12 Tally Screen in Shovel Knight (2014) Displaying Active, Broken and Inactive Checkpoints

Beyond their basic save functionality, checkpoints shape player behaviour and experience in subtle ways. They create natural experimentation boundaries where players often take more risks immediately after checkpoints knowing failure consequences are minimal. They establish pacing rhythms as players push toward the next save point. The following chapter examines what academic research has established about these effects.

2.3.2 Checkpoints in Research

Despite checkpoints being ubiquitous in game design, academic research on their effects remains surprisingly limited. Johanson et al.'s (2023) study *If at First You Don't Succeed* represents one of the few investigations into how checkpoints affect player skill development

and progress. Their research tested 150 experienced gamers in a custom platformer, examining whether checkpoints help or hinder learning. The results revealed that checkpoints significantly improved player progress during training sessions, enabling players to complete more levels while facing fewer instances of stalled or regressed progress. When checkpoints were removed in transfer and retention tests, players maintained their performance levels, demonstrating that checkpoints supported progress without creating dependency or hindering genuine skill development (Johanson et al., 2023).

The study's findings reveal how checkpoints align with Flow Theory principles by creating clear short-term goals and reducing performance anxiety. Players with checkpoints reported feeling more comfortable taking risks and experimenting with different strategies, knowing they had safety nets at strategic points.

This psychological freedom enhanced rather than diminished learning, as players can focus on mastering new challenges without the frustration of repeating already-beaten sections. The research showed that checkpoints specifically reduced both "stalled progress time" where players got stuck at difficult points, and "regressed progress" where players performed worse than previous attempts. By maintaining players in their optimal challenge zone, checkpoints prevented the negative spiral of frustration and fatigue that often accompanies repeated failure.

Johanson et al. also tested the combined effects of checkpoints with forced breaks after death, finding near-additive benefits when both techniques were used together. While players overwhelmingly preferred checkpoints and disliked the forced pauses, the data revealed that both techniques independently contributed to skill development. This disconnect between player preference and actual performance benefits highlights an important consideration for checkpoint design: players may not always recognise what helps them improve, suggesting that checkpoint systems should balance player satisfaction with learning effectiveness.

This connection between checkpoints and skill development becomes particularly relevant when considering failure in games. As Johanson et al.'s research shows, player progress often stalls when repetition alone fails to produce improvement. Since checkpoints fundamentally shape how players experience and recover from failure, understanding failure's impact becomes essential for effective checkpoint design, and to a greater extent game design as a whole. The following chapter examines this relationship between failure and player experience.

2.4 The Paradox of Failure

> *Though we may dislike failure as such, failure is an integral element of the overall experience of playing a game, a motivator, something that helps us reconsider our strategies and see the strategic depth in a game, a clear proof that we have improved when*

we finally overcome it.

— Jesper Juul, *The Art of Failure: An Essay on the Pain of Playing Video Games* (2013)

Failure in games presents a paradox: players claim to dislike losing yet consistently seek experiences designed to make them fail. This chapter examines why players need failure for satisfaction and learning, exploring the psychological mechanisms that make defeat enjoyable when properly implemented. It then analyses how designers can harness failure as a tool for engagement, distinguishing between productive temporary setbacks and destructive perpetual failures. Understanding this relationship is essential for checkpoint design and the research framework of this thesis.

2.4.1 Why Players Need Failure

Video games present a unique contradiction: players explicitly state they want to win, yet consistently rate games as "too easy" when they never fail. Juul's (2009) research reveals that players who failed several times before succeeding rated games significantly higher than those who never failed at all. This counterintuitive result shows that failure serves essential psychological functions beyond simply contrasting with success.

The dual perspective model explains this contradiction. Players simultaneously operate from an internal goal-oriented perspective that desires victory, and an external aesthetic perspective that appreciates balanced challenge. This duality means players experience failure negatively in the moment while appreciating it retrospectively as part of a meaningful experience.

Juul's study found that players who felt personally responsible for failure rated games more positively than those who blamed the game itself. When players attributed failure internally, they experienced greater satisfaction than when attributing it externally. This contradicts assumptions about casual players preferring not to feel responsible for failure (Juul, 2009).

Failure adds content by forcing players to perceive game nuances they would otherwise miss. When asked what makes games "too easy," players most commonly cited lack of challenge and not having to rethink strategies. Without failure, games become forgettable experiences lacking substance. Juul's critique reveals a tension with Flow Theory: while flow requires removing friction for absorption, players actively want challenges that cause failure. The resolution lies in understanding that flow requires appropriate challenge, which necessarily includes failure. Difficulty should vary in waves, as stated by Chen, creating memorable moments of struggle and triumph rather than states of passive engagement.

Understanding failure types proves crucial for effective design. Juul identifies four punishment types but notes that all ultimately translate to setback punishment, with time as the fundamental currency. The question becomes not whether players fail, but how much time failure costs them. Checkpoints directly control this temporal cost, making them the primary tool for managing setback punishment. Modern games succeed not by eliminating failure but by reducing its temporal cost, respecting the player's time, while maintaining impact.

Frommel et al.'s (2021) research distinguishes between temporary failure that enables learning and perpetual failure that causes frustration. The key insight is that repeated temporary failures at the same point eventually become perpetual failure. Checkpoints prevent this transition by providing progressive anchors ensuring even repeated failure yields forward movement. Individual challenge orientation adds complexity, as some players enjoy failure while others are vulnerable to its negative effects. The solution is not eliminating failure but managing its presentation and cost, for example through checkpoint placement (Frommel et al., 2021).

Traditional DDA can easily misunderstand flow by eliminating failure, when flow actually requires appropriate challenge and the possibility of failure. What breaks flow is not failure itself but excessive temporal punishment which is the time cost of repeating content. Effective DDA should maintain failure frequency while modulating its temporal cost. Players need to fail to appreciate success, but failure must feel fair and low in temporal cost. Modern design succeeds through high failure rates with low temporal costs, enabled by checkpoints that decouple failure frequency from punishment severity. This approach preserves both flow and failure's psychological benefits.

2.4.2 Failure as a Design Tool

Frommel et al.'s framework provides practical guidance by identifying how failure transitions from productive to destructive. The failure-improvement cycle demonstrates that each failed attempt generates learning opportunities only when players can identify what went wrong and formulate new strategies. This cycle breaks when failures feel arbitrary or unsolvable, transforming temporary educational setbacks into perpetual frustration. The mechanics of creating meaningful failure centre on feedback clarity, as players must understand precisely why they failed in order to improve their approach. Vague deaths from unclear hitboxes or inconsistent enemy behaviour prevent learning, while clear failure communication enables strategy refinement. Frommel's participants consistently linked satisfaction to overcoming specific, understood challenges rather than random difficulty.

Player agency becomes crucial in failure design. Frommel discovered that players often create personal goals beyond designer intentions, as found in Flow Theory, such as completing

levels without damage or using self-imposed restrictions. These player-created challenges generate more meaningful failure experiences than designer-imposed ones. Checkpoint systems should accommodate this by offering optional use rather than forced activation. A player who may want more challenge through more setback should be able to bypass checkpoints, while struggling players can use them freely. By offering this choice, the game respects diverse player motivations.

The relationship between failure frequency and player learning follows a curve rather than linear progression. Initial failures provide maximum learning as players discover mechanics and strategies. Continued failure at the same point yields diminishing returns until learning stops and frustration dominates. Checkpoints serve as circuit breakers in this process, allowing players to solidify learning from one section before tackling the next.

The roguelike *Hades (2020)* exemplifies this philosophy by making death integral to both narrative and gameplay. Each failed escape attempt (Figure 2.13) advances character relationships and story, transforming failure from setback into a form of progression. Players can benefit from deaths to unlock new dialogue and upgrades, demonstrating how failure can function as progression rather than punishment. Understanding this, along with its interaction with checkpoints and flow, forms the basis for the synthesis.

Figure 2.13 Death Screen from Hades (2020)

2.5 Synthesis and Research Gap

Flow Theory and DDA research have produced systems for adapting various game elements, including enemy parameters, resource availability, and level generation. However,

checkpoints have received little attention as an adaptive element despite being present in most games. This chapter identifies this research gap and establishes how adaptive checkpoint systems can benefit theoretical understanding with applied game design research.

The literature review reveals substantial research on Flow Theory's application to games, progressing from Csikszentmihalyi's original psychological theory through Sweetser and Wyeth's GameFlow framework to Chen's embedded choice implementation. This progression demonstrates how maintaining challenge-skill balance is essential for flow states. Similarly, DDA research has evolved from rule-based systems that can incorporate sophisticated psychological theories to data-driven approaches, as documented by Mortazavi et al. and Zohaib. Importantly, these DDA studies not only demonstrate the value of flow in adaptive systems but frequently recommend flow-based approaches as optimal solutions. Studies on failure by Juul and Frommel have established its necessity for player satisfaction and learning. Yet remarkably, checkpoints receive minimal academic attention despite being the primary mechanism managing what Juul identifies as failure's fundamental currency: time. Van den Hoogen et al. (2012) found that while initial failure can feel challenging, repeated deaths without progress cause declining positive affect. Players transition from feeling challenged to feeling defeated. Checkpoints directly control this transition, yet remain understudied as adaptive elements.

Recent empirical work has examined checkpoints and respawn mechanics. Johanson et al. (2023) found that checkpoints support skill development without creating dependency. Cuerdo et al. (2021) compared static respawn conditions, including permadeath, level start, checkpoint, and savepoint, the point where the player may save their game progress. They found that respawn location significantly affected player autonomy and curiosity. However, Mortazavi et al.'s systematic review of 85 DDA papers identified no checkpoint-specific approaches. While DDA research has explored adapting enemy strength, spawn rates, and resource availability, checkpoint spacing as an adaptive element remains unexplored. Crash Bandicoot used additional checkpoint placement as a difficulty adjustment mechanism decades ago, yet this approach has seen no systematic academic investigation.

Current checkpoint implementations have mostly remained static across all players, ignoring individual differences in skill, learning speed, and frustration tolerance. Adams and Rollings, Byrne, and Schell all define checkpoints similarly, with all examples showcasing static uses of the game feature. With Chen's insight that different players possess individually unique Flow Zones, and Frommel's finding that players vary in challenge orientation, static checkpoints simply do not properly utilise these factors. While flow sources emphasise these individual differences in player experience and skill, checkpoint research continues to treat all players

uniformly, suggesting a disconnect between theoretical understanding and practical implementation.

Static checkpoint systems cannot respond to individual player needs. They may drive away novices through excessive repetition while boring experienced players with trivial consequences. This violates Flow Theory's central principle of dynamic challenge-skill balance. Checkpoints serve as the primary mediator between failure and progress, directly influencing whether players experience Frommel's productive "failure-improvement cycle" or destructive perpetual failure. When checkpoint spacing remains fixed regardless of player performance, the system cannot adapt to maintain optimal challenge levels for different experience groups.

Failure can naturally disrupt flow states, but well-designed checkpoints can help players recover more quickly by reducing the temporal cost of setbacks. While checkpoints serve this flow-supporting function, seemingly no research has examined how adaptive checkpoint activation could optimise this role for different player types. The potential for selectively activating predetermined checkpoint locations based on individual player performance remains largely unexplored, aside from *Crash Bandicoot*'s approach from the 1990s. Similarly, while both Chen's player-oriented DDA and checkpoint systems conceptually relate through player control and agency, this theoretical bridge has not been investigated. The industry's focus on adapting other game elements while keeping checkpoint systems static indicates an unexplored opportunity for enhancing player experience.

This research proposes adaptive checkpoint systems as a bridge between established DDA principles and an overlooked game mechanic. Applying Flow Theory to checkpoint design could produce systems that dynamically adjust to maintain individual players in their optimal challenge zones. Such systems would activate additional checkpoints at specific locations when flow-based rules indicate players are struggling, while maintaining wider spacing for those demonstrating mastery. Following Hunicke and Byrne's findings on invisibility, validated by *Crash Bandicoot*'s success, and supported by Denisova and Cairns' empirical evidence that subtle adjustments work better than obvious ones, adaptations must remain undetectable. Players should attribute progress to their own improvement rather than system assistance.

The connection between Juul's "failure adds content" principle and Soderman's critique that players need breaks from flow suggests that adaptive checkpoints must not eliminate failure but modulate its cost. The system should maintain challenge integrity while preventing frustration accumulation, recognising that both appropriate failure and strategic breaks are essential for maintaining flow states. Chen and Schell's shared critique that traditional DDA damages flow by removing player agency indicates adaptive checkpoints should preserve choice through optional activation or varying state preservation levels. Acknowledging that DDA can

adjust enemy parameters and player resources such as health or ammunition, adaptive checkpoints represent another resource that can be dynamically managed. Players must retain the choice to use or bypass checkpoints, as this agency is fundamental to flow experience.

Existing research has examined static checkpoint conditions (Cuerdo et al., 2021) and checkpoint effects on skill development (Johanson et al., 2023), but has not investigated adaptive checkpoint systems through the lens of Flow Theory. Given this gap in the literature and its implications for game design, this research addresses the following question:

To what extent do adaptive checkpoint systems affect the flow channel compared to fixed checkpoints across diverse player experience levels?

Based on the theoretical framework established through the literature review, the following hypotheses guide this investigation:

- The adaptive checkpoint system will yield higher flow intensity than the fixed checkpoints.
- Players using the adaptive checkpoints will show higher completion rates and lower quit rates compared to those using fixed checkpoints.
- Players using adaptive checkpoints will report different levels of frustration compared to those using fixed checkpoints.

The following chapter examines the ethical considerations that must guide the development and implementation of such adaptive systems, concluding this literature review.

2.6 Ethical Considerations

The implementation of flow mechanics and adaptive systems in games raises important ethical concerns about player wellbeing and designer responsibility. Building on the critiques discussed in previous chapters, this chapter establishes ethical guidelines for implementing these systems responsibly in both research and commercial contexts.

2.6.1 Ethical Implications of Flow-Based Game Design

The deliberate design of flow states presents a fundamental ethical question: when does engagement become exploitation? As Soderman's critique suggested, flow is a framework that shapes player behaviour, not just a psychological state. When designers create systems specifically to induce flow, they need to consider whether they are supporting genuine enjoyment or manipulating players.

The addiction potential of flow states requires careful consideration. Flow's characteristic features of lost time awareness and reduced self-consciousness, while creating enjoyable experiences, can lead to unhealthy gaming habits. Players in flow states lose the ability to self-regulate, which becomes particularly concerning when games use engagement metrics to maximise playtime. This "tunnel vision" blocks external awareness, but can also

prevent players from making conscious choices about their gaming behaviour (Soderman, 2021).

Ethical design should include natural break points that allow players to step back and reflect. Games need moments where players can form memories and process their experiences, as well as take physical or mental breaks. This is particularly relevant for checkpoint systems, which naturally create these pause points.

The growth focus embedded in Flow Theory raises another concern. Not all players want constant challenge escalation. Some prefer exploration, social interaction, or peaceful gameplay. By assuming everyone seeks flow states, games risk excluding players who engage with games differently. Ethical implementation means recognising flow as one option among many, not a universal standard applicable to all players. These same concerns extend specifically to DDA systems, which raise additional questions around transparency and player consent.

2.6.2 Transparency and Agency for DDA

DDA systems create specific ethical challenges around player control and informed consent. As discussed in Subchapter 2.2, research indicates that invisible DDA is generally more effective. However, hiding difficulty adjustments raises some ethical questions. Players cannot consent to something they do not know exists, turning gameplay somewhat deceptive.

The solution is not complete transparency, which can break immersion and reduce effectiveness. Instead, games should provide basic information about adaptive features in settings menus or tutorials, with clear options to disable them. Players should be informed when systems are adjusting their experience, even if they do not need to understand every detail.

Player agency extends beyond awareness of DDA to actual control over it. When games automatically lower difficulty after failures, players can feel patronised. Many players find satisfaction in overcoming fixed challenges through practice and persistence, as mentioned in Subchapter 2.4. Taking this away by forcing easier gameplay denies them earned victories. A more ethical approach to adaptive design preserves player choice through opt-in systems, sensitivity settings, or hybrid approaches that suggest rather than enforce these difficulty changes. The checkpoint DDA proposed in this research respects this principle by not forcing the use of checkpoints, but making them available for players if the system deems they will benefit from it.

On the other hand, the accessibility advantages of DDA provide strong ethical justification for the use of these systems. Adaptive difficulty allows players with disabilities, different skill or experience levels, or limited time to enjoy games that would otherwise exclude

them. This is not about lowering standards but recognising diverse player needs and scenarios. Players with disabilities do not necessarily want easier games, but may appreciate challenges that adapt to their specific needs.

Research using these systems must also follow certain ethical protocols. This includes clear communication about data collection, defined limits on how much adaptation occurs, and thorough debriefing about any deception involved. The goal is enhancing player experiences while respecting autonomy, diversity, and informed choice.

3 METHODOLOGY: DESIGNING AN ADAPTIVE CHECKPOINT SYSTEM

An effective approach to researching the value of adaptive checkpoints compared to fixed ones is by applying an adaptive system to an already existing and successful game and comparing two versions. This chapter describes how this was handled for this research and the design choices made to achieve this.

3.1 APPROACHING THE RESEARCH QUESTION

In order to find an answer to the research question, a suitable game must be identified. The game must have checkpoints as a game mechanic and be both challenging and successful at entertaining players by providing a solid experience. Then, an adaptive checkpoint system is designed that either adds or leaves out checkpoints throughout the game experience.

To research the effectiveness of adaptive checkpoints for flow stimulation, adaptive checkpoints must be compared with regular checkpoints. Regular checkpoints here are defined as checkpoints that are always present in their predetermined locations in the game levels. Adaptive checkpoints, on the other hand, are then defined as checkpoints that only appear in certain cases, based on the current game difficulty, which is determined by the success or struggle of the player.

A balance between internal and external validity is important for retrieving relevant data, while also yielding results that are applicable for the development of future commercial games. For this reason, a controlled between-subjects experiment was chosen to establish the causal relationship between the adaptive checkpoint system and a player's flow state. The adaptive checkpoint system is active for the experimental group, and the control group has no additional systems or other game-altering elements active. Otherwise, both groups have identical game levels with the same challenges, obstacles and conditions, based on the campaign of the original game, with only minor changes.

Participants are recruited using convenience sampling (fellow students, friends, family…), either in person or online, and are subsequently contacted digitally in order to receive the game experiment. For participants under the age of 18, participation requires the prior written consent of a parent or legal guardian.

After giving consent to participate, the participants are randomly assigned to one of the two groups, avoiding systematic bias, with each half of the total participant pool being assigned to only one of the two groups. The estimated total number of participants for this research is 40

at the minimum, given the recruitment constraints, with each group then having 20 participants, to ensure each group yields practical data for analysis. There are only two evident criteria for the participants: they must both want to participate and be capable of playing the game. Participants are unable to achieve a flow state if they play against their own volition. It would, of course, be unethical to have a person participate against their will. However, this is something that should be recognised when dealing with flow state measurements. It is also recommended that participants have at least some experience playing video games, so that the experiment is not their first time playing a video game, to avoid having participants struggle with the controls, for example. Retrospectively, it can be noted that no formal power analysis was conducted, and the study ultimately recruited 61 participants.

As this research handles the subject of Flow Theory, observing changes in a player's behaviour during the experiment is key for determining the flow state of a participant. Measuring this behaviour in people is challenging. It is important to acknowledge that a player may not enter a flow state at all, as the game's suitability for generating flow was not pre-validated thoroughly prior to the study. The observer must make sure to not interrupt the participant during the experiment, either. Such an interruption would disrupt the flow state, muddying the data quality and compromising it. A think-aloud protocol was initially considered, but was rejected for this reason.

For these reasons, in order to measure flow, the observational approaches mentioned above are incorporated into a telemetry system programmed in the experiment and a post-play questionnaire. This approach was influenced by the work of Rosa et al. (2021). Their DDA research used game telemetry, which would track parameters during participant playthroughs, and a questionnaire. Unlike this research, their questionnaire was given to be filled in after each stage of the experiment, testing flow at different points. This experiment limits the use of questionnaires to only one right after playing, as to avoid interrupting flow during the limited time of participation. The various facets of this methodology are further elaborated in the following subchapters, with each aspect being explained in detail.

3.2 THE GAME

The video game chosen for this experiment to be based on is the 2010 indie game *VVVVVV* by Terry Cavanagh (see Figure 3.1). The game was very well received for its challenging yet simple puzzle-platforming gameplay, which are ideal elements for a flow-stimulating game. There are also checkpoints present in *VVVVVV*, which allows the development of the DDA system for this experiment.

Figure 3.1 VVVVVV game banner

In this game, the player must save their stranded space crew. There are only a few simple controls: a button to move left, a button to move right, and an action button to flip their own gravity. The player cannot jump and must instead use this ability to navigate obstacles such as spikes. This flip only affects the player in this game version. Furthermore, VVVVVV has a simplistic art style that is not confusing or overly distracting. Following the tenth anniversary of the game's release, Cavanagh made the game open source, with all code available on GitHub in C++. These controls are easy to learn but hard to master, creating suitable conditions for flow across different experience levels, while it being open-source allows the implementation of the adaptive checkpoint system.

Minor changes have been made to *VVVVVV* to make it fully suitable for the experiment. Several options were disabled, such as speedrun modifiers and cheats, to prevent participants from circumventing the game's challenges. Fan-made levels are also not available in this version, leaving only the option to play the campaign.

This campaign is altered in several ways to form a focused experience that should on average take 15-20 minutes to avoid taking up too much of the participants' time, meant to be completed in one sitting. There are different areas the player must go through. These areas are referred to as levels. Each level consists of multiple rooms, being a segment of the level that fills the whole screen and usually offers a certain challenge. The experiment game mimics the demo of *VVVVVV*, containing the very first level "Space Station", followed by the level "The Laboratory". The player starts at the beginning of the Space Station as they would in the normal game. However, at the end of the Space Station, after using the teleporter and rescuing the first crewmate, the player is immediately teleported to the beginning of The Laboratory, with the entrance being blocked off. This way, the exploratory aspect of *VVVVVV* is removed.

The game ends in two ways: when the player manually quits or after reaching the end of The Laboratory. At this endpoint, after rescuing Victoria and using that level's teleporter, the game quits automatically. Any other teleporters have been removed from the game, leaving only one at the end of each level.

Figure 3.2 The last room of the tutorial for the experimental version. After this the DDA system activates.

To further create a more linear, focused experience, optional rooms with collectable items were blocked off, with one exception. In The Laboratory, one optional trinket room was repurposed to connect with the main path, adding one mandatory room. All level edits were made directly in the source code. With the trinket rooms removed, there is a total of 64 rooms the participants must go through to get to the end. The first eight serve as a tutorial of sorts to get the player accustomed to the controls. The tutorial ends after the eighth room (see Figure 3.2). The Space Station ends at room 23 (see Figure 3.3), and rooms 24 to 64 form The Laboratory, which introduces Gravity Walls as its main mechanic (see Figure 3.4). These force the player to automatically flip their gravity. Additional information about the design of the game experiment and how it functions is provided in Appendix A. The source code is available via the link provided in Appendix B.

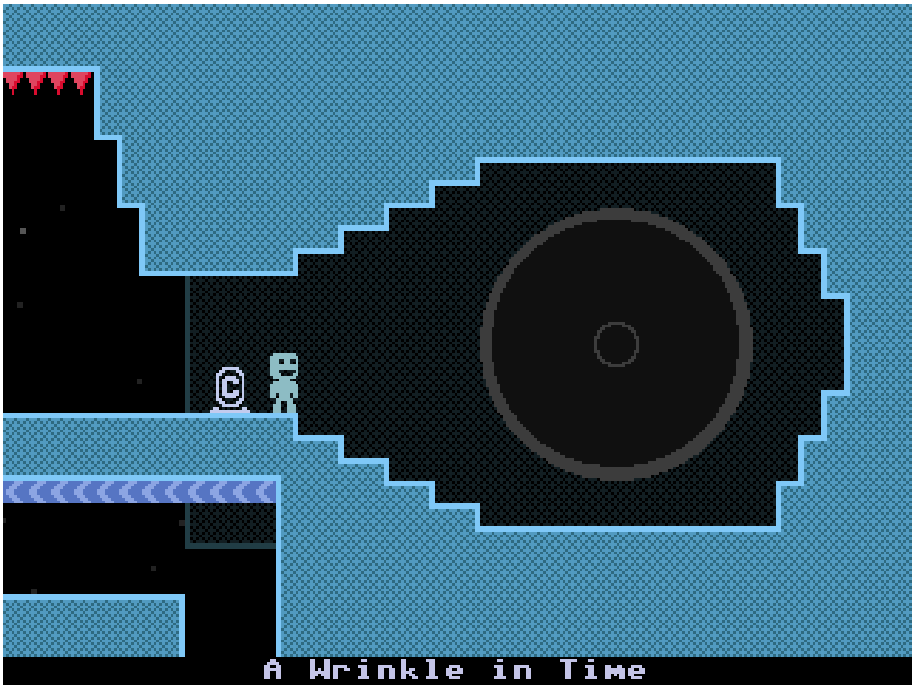

Figure 3.3 The teleporter at the end of the first level, "Space Station".

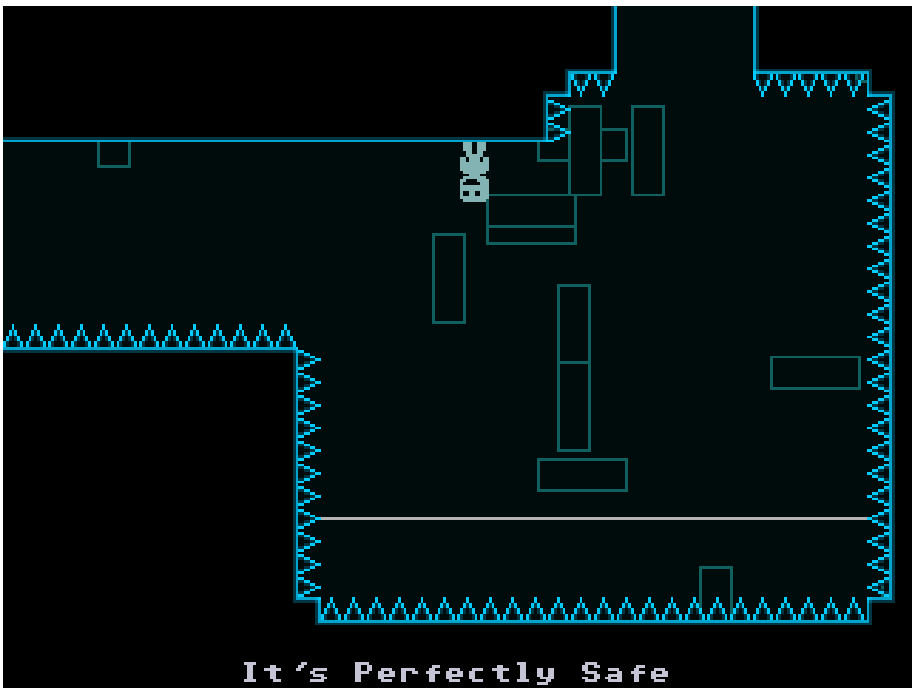

Figure 3.4 A Gravity Wall placed above spikes in "The Laboratory".

The final build of the experiment is available for participants both as a download and as a web browser game on Itch.io. Participants receive a link that leads them to the web page, where they can play the game using their browser of choice or download it to play it on their Windows device only. The browser version was compiled from C++ to WebAssembly using Emscripten. The following chapter describes the design of the adaptive checkpoint system and the telemetry implemented within this build.

3.3 DESIGNING THE SYSTEMS

3.3.1 Influential Methodologies

The literature review concluded that an optimal DDA system is invisible, continuously and automatically tracking the player's abilities. The system then uses this information to determine the changes that must be made to the experience to keep the player challenged without frustration. The two main academic works used to inspire the development of this system are those by Rosa et al. (2021) and Xue et al. (2017). The performance-based design of the former will be discussed first.

In their work, Rosa et al. similarly state three requirements for DDA, which agrees with the conclusions of the literature study: identify the player's skill, perceive and record changes in performance to then discretely maintain the game behaviour without being perceived by the player. They created a hybrid DDA system based on performance data and physiological data (skin conductance). The gameplay data gathered with telemetry consisted of the time to end a level, the quantity of successful jumps, and the distance travelled before dying. A physiological approach is out of scope for this experiment, and as *VVVVVV* uniquely does not have jumping but instead flipping, only the data for time and distance are interesting for this research.

Rosa et al. made use of a formula that analysed the components and obstacles in a level to derive the difficulty of that level, which is used to compare the participant's performance. A similar formula is not used for this experiment. This is out of scope, and a complex task for the game's interconnected use of rooms. Instead, each room has estimative thresholds for the unique challenge provided in that room. The threshold calibration is based on an observational analysis of six publicly available playthrough recordings of the *VVVVVV* demo on YouTube, using the deaths and time spent per room of all six players to assess substantiated thresholds for each room. This calibration is located in the main spreadsheet, available in Appendix B. The DDA system then compares the player's performance to these thresholds to conclude the player's experience level. Furthermore, this experiment opts to not only gather data for the adaptation of difficulty during participation, but also for analysis after participation to further assess the state of flow that participant was in.

Xue et al. proposed a framework to maximise a player's engagement throughout the entire game experience. Instead of focusing on performance, they took a probabilistic approach. Using a probabilistic graph-based model, they explain how DDA is a progression optimisation problem, maximising total engagement through the whole game. They model player progression through a game as transitions between states, where each state represents a difficulty level. The

goal is to find the optimal "path" through these difficulty states that maximises player engagement. They try to then figure out what difficulty for each state maximises total playtime across all players.

Their system looks at the current death count, how long they struggle, and historical data from similar players. A player can either win, try again, or quit. After data collection, depending on the quit probability, the difficulty optimiser determines whether the difficulty is increased or decreased, which then notifies the Dynamic Adjustment Service.

This experiment departs from Xue et al.'s approach in several ways. Rather than using probabilistic states, a set number of fixed difficulty levels is used. Since quitting the game ends the experiment entirely, quit probability is not a relevant factor. Instead, the two determinants of difficulty adjustment are whether the player struggled or succeeded in the most recently completed room.

Finally, the work of Chen is particularly applicable for this experiment, as the goal is to maintain a person's flow during the game. To briefly revisit, his work *Flow in Games* (2006) stated the three fundamental game design conditions:

1. The game is intrinsically rewarding, and the player wants to play the game
2. The game offers the right amount of challenge to create Flow Theory's skill-challenge balance
3. The player has a sense of personal control over the game activity

Chen also proposed three methodologies for game designers:

1. The game's Flow coverage can be expanded by including a wide spectrum of gameplay with different difficulties and elements
2. A Player-oriented Active DDA system allows different players to play at their own pace
3. Embedding DDA choices into the core gameplay mechanics will let the player make their choices through play

These rules are central for the design of the flow-based DDA system, for making the system more flexible. In *VVVVVV*, checkpoints are activated manually. This is done by walking past them, making them light up. Although in a limited fashion, this game incorporates embedded choices, which strengthens its potential for flow. The majority of the checkpoints in *VVVVVV* are placed in such a manner that the player is not forced to interact with them, allowing them to choose whether they want to activate a new one. A player can ignore a checkpoint in

case they want to add an additional challenge for themselves. This way, players have control over their point of return.

3.3.2 The Adaptive Checkpoint System Design

The DDA system uses a set number of variables that are tracked during gameplay to adapt the difficulty:

- Number of deaths in each room, with for each death:
    - The location of the death in that room
    - The room name
    - What number death this was
    - A timestamp when the player died
- Time spent in each room

After the player leaves the room, these variables are then compared to the predetermined thresholds of that room. There are three different thresholds: a death threshold, a short time threshold, and a long time threshold.

The "**death" threshold** represents the number of deaths required to indicate struggling, typically ranging from 1-2 for simpler rooms to 3-5 for more challenging ones, with exceptionally difficult rooms set higher. Play styles may vary, and some players could die frequently without getting frustrated, however this remains an acknowledged limitation of the approach. This threshold on its own is insufficient, too, as the player may die accidentally or while learning new mechanics.

The "**short time" threshold** helps distinguish between fast, accidental deaths and actual effort and struggling. If the player dies a few times but recovers quickly before the threshold is reached, then the game counts the room as a success. This threshold is generally set to 15-30 seconds. For very short rooms or falling sequences, where dying is enough of an indicator for struggle, this is set to 0-5 seconds.

The "**long time" threshold** often indicates apprehensive player behaviour, where they proceed extremely carefully due to anxiety about losing progress, or suggests they have taken breaks due to mounting frustration. This threshold is used to avoid the player getting overwhelmed or experiences too much anxiety or frustration. Set typically to 30-60 seconds for standard rooms, it extends to 600 seconds for rooms where this threshold is not applicable, such as falling sequences, rooms with bouncing mechanics, or other contexts where extended time is expected due to lack of control.

These thresholds work in tandem with each other and a **system that compares the closeness of death locations** in a room to determine struggling because of specific obstacles, visualised in Figure 3.5. There are four conditions that determine whether the player struggled in the room they left when entering a new, unexplored room:

- Death threshold and short time threshold are both exceeded (indicating genuine, repeated attempts)
- Long time threshold is exceeded (suggesting anxiety or frustration)
- Player died 3+ times within close proximity and the death threshold is exceeded (demonstrating a specific obstacle proves too challenging)
- Player died 3+ times within close proximity and the short time threshold is exceeded (typically only triggered in extremely difficult rooms where the death threshold is set to a high number)

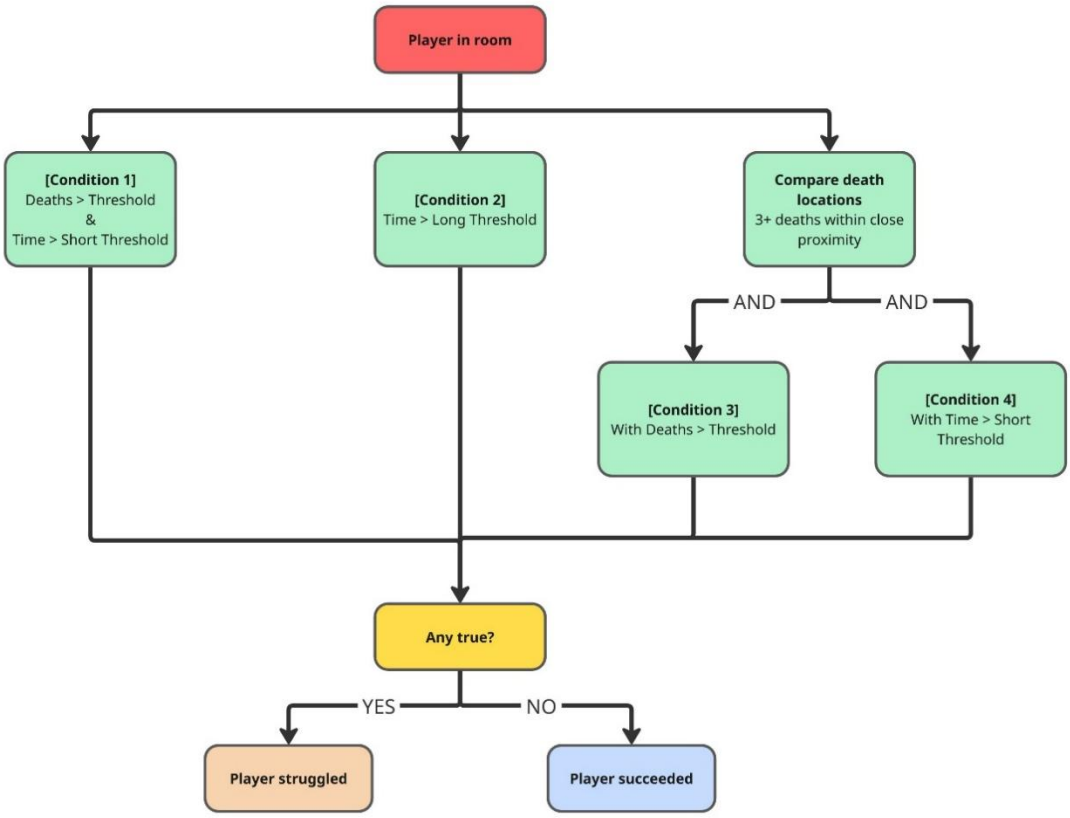

Figure 3.5 A chart showing the different conditions for struggling.

If any of these conditions are met, the system concludes that the player struggled in the previous room. If none of these conditions are met, it is concluded that the player succeeded well, but, for brevity's sake, this is simply referred to as succeeding.

After this, the difficulty level may be changed. The difficulty ranges from 1 to 7, where stage 1 is the easiest and stage 7 is the hardest, starting in the middle at stage 4. The system maintains consistent behaviour for struggles and successes, visualised in Figure 3.6.

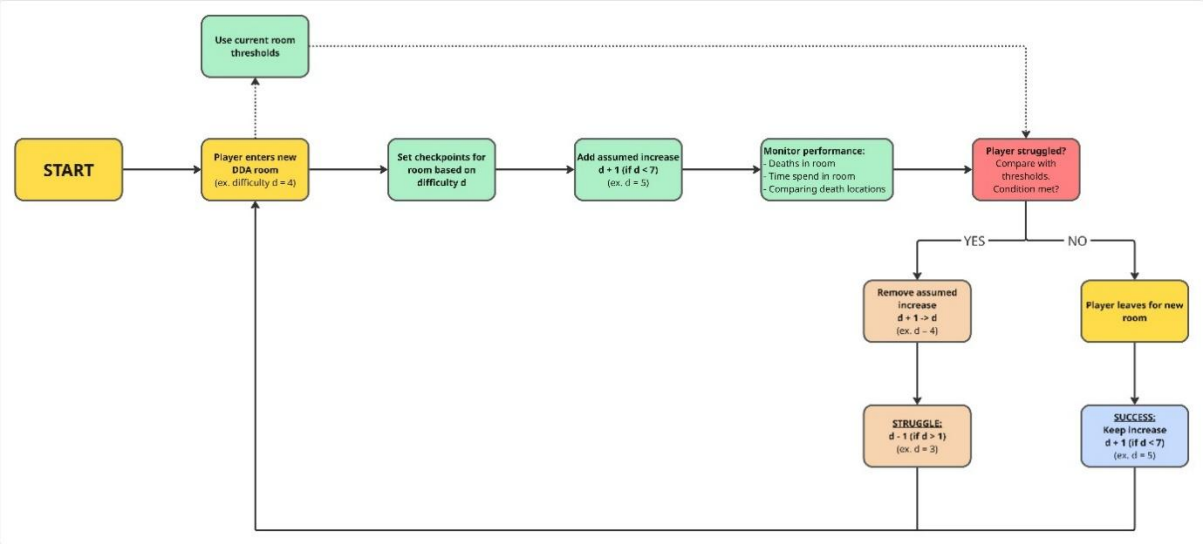

Figure 3.6 A chart visualising the process of the DDA checkpoint system.

Struggle detection occurs in real-time within the current room. When entering a new room, first the current difficulty d is applied. Then, the difficulty is automatically increased to d+1, except if it has already reached the maximum stage of 7. This increase is referred to as an "assumed increase" for this experiment. If the player succeeds in this room and never triggers any struggle conditions, this pre-emptive assumed increase was correct, and the difficulty has already been set to the correctly increased value, which allows the next room to have the correct amount of checkpoints to limit potential frustration for the player. Once struggle conditions are met, the system immediately identifies this. If the player struggles in this room by meeting one of these conditions, the assumed increase is removed and the difficulty is set back to d. This is then followed by another decrease of 1, resulting in a final difficulty of d-1, one lower than their starting difficulty when they first entered the current room, except if that starting difficulty was the minimum stage of 1. When entering the next room, this correctly lowered difficulty will be applied.

All checkpoint positions are predetermined during development. The DDA system does not generate new checkpoint locations at runtime but instead determines which of the available predetermined positions are active based on the current difficulty stage. Due to the small size of the rooms in *VVVVVV* and their lack of space for more than two checkpoints, it is superfluous to have a different change for each stage of difficulty. Instead, the difficulty stages are part of

distinct ranges. These ranges vary for each of the two levels. For the Space Station, there are two ranges. Stages 1-4 of difficulty will have one checkpoint present, while stages 5-7 have none.

The Laboratory is more challenging, which is why there are three different ranges for that level. An example where this is most noticeable is the third room of The Laboratory which normally has one checkpoint at the start, in the top-right corner (see Figure 3.7). For The Laboratory, stages 1-3 ensure that the most difficult rooms have two checkpoints present if possible for that room (depending on the space and size of that room). Then, if possible, stages 4 and 5 ensure that at least one checkpoint is present, and stages 6 and 7 have none present. There is one exception, however. Using this logic for setting checkpoints, if the player is experiencing the game on the hardest difficulty, they could encounter no checkpoints for multiple rooms in a row. To mitigate this, for every three or four rooms a fixed checkpoint will always be present to avoid the player being set back too far on death, which could lead to frustration.

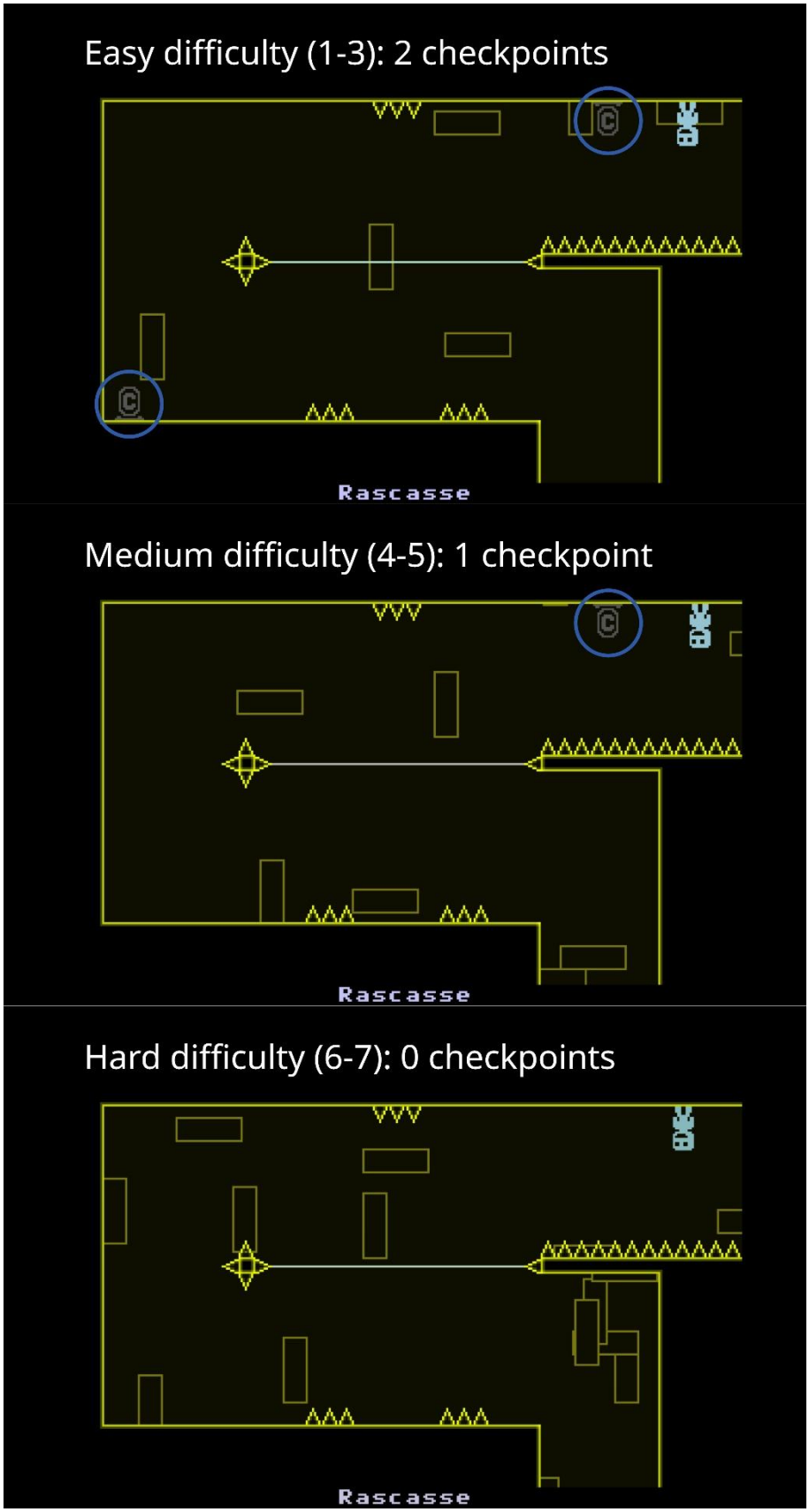

Figure 3.7 An example of the different amounts of checkpoints that can be present in a room based on current difficulty.

As the first eight rooms serve as a tutorial, the DDA is only activated afterwards, starting from the ninth room onward (see Figure 3.8). This allows the player to first get used to the controls, without making the game additionally easy or hard.

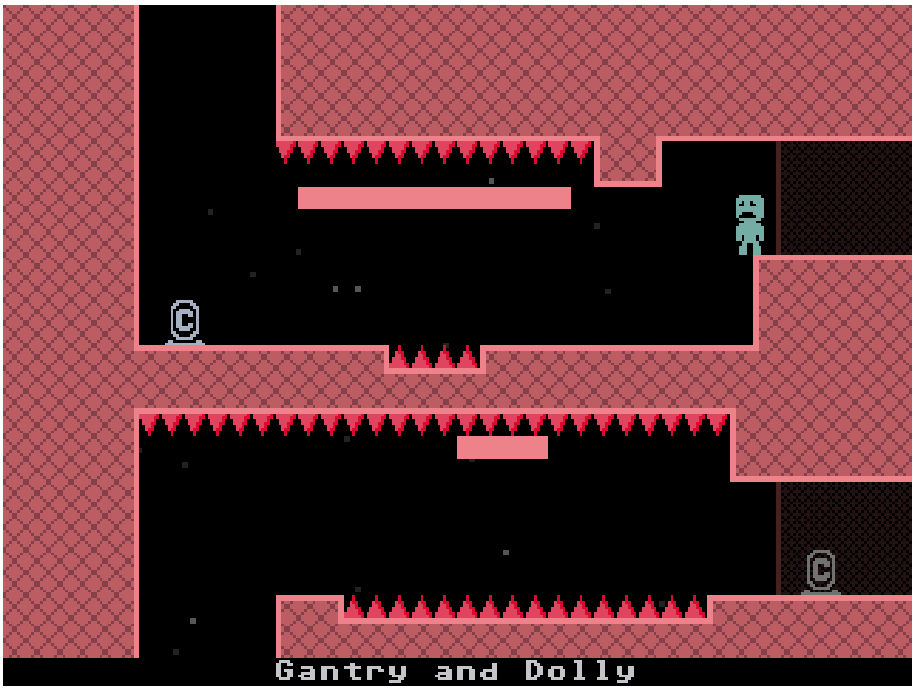

Figure 3.8 The first room with DDA active. The player enters from above and heads right to Comms Relay, entering again via the bottom-right where DDA is activated.

Beyond the tutorial, certain rooms throughout the game also have DDA disabled. These include rooms that function as transitions between areas or rooms where the challenge is minimal or absent. In these rooms, the original checkpoint configuration is preserved regardless of the current difficulty stage. This selective application ensures the DDA system only modifies rooms where difficulty adjustment is meaningful, not counting the player proceeding through what is essentially a hallway, as a success. This checkpoint configuration follows its own set of rules.

The predetermined checkpoint positions are mainly based on those of the original game. For the easiest difficulty level, more checkpoints were added to the more challenging rooms in The Laboratory for the system to activate. To avoid having the player replay large parts of the experiment after a long streak of success followed by sudden failure, some rooms always have a checkpoint to provide some safety. If those aren't present, the difficulty wouldn't lower fast enough to provide a checkpoint, since it only does so after struggling in several newly entered rooms.

3.3.3 Tracking Player Behaviour

The main goal of the telemetry system is to operationalise the potential flow of the participants during the experiment by looking at their progress, struggles and interactions. The use of such a system was inspired by the work of Rosa et al. and the logistical constraints of remote participation.

The telemetry system is an extension of the DDA system. It monitors data already used by the adaptive system, along with other variables that allow for a better understanding of the player's experience and to deduce a flow state with greater certainty. The exported data is as follows:

- If the player was part of the control or experimental group
- Timestamp of export
- The total number of deaths
- Total time spent
- Furthest room reached (number and name)
- Number of unique checkpoints activated
- Total key presses
- Average key presses per minute (**KPM**)
- When the difficulty changed, to what the difficulty changed, in what room and what condition caused a decrease
- For each room:
    - The room name
    - Timestamp of when the room was first entered
    - Total time spent in that room
    - Stage of difficulty when first entering the room (to determine the amount of available checkpoints in this room)
    - The total number of deaths in that room
    - What number death this was
    - The location of each death in that room
    - A timestamp when the player died
- KPM for every minute

This data is logged after the experiment has ended, which, as explained earlier, happens on quit or after the player completes the full game experiment by reaching the end of The Laboratory.

If downloaded, the local version of the game stores all data in a file which the participant sends over after playing. This file is located in the game's save folder. The game is also available to be played via a web browser, in which case the data is sent automatically to a Google Sheet (see Figure 3.9), using Google Apps Script. The data is then displayed in columns for analysis, with the final column also linking to a Google Document which contains the same elements as the local data file.

	A	B	C	D	E	F	G	H	I	
1	Timestamp	Group	Total Deaths	Total Time	Furthest Room Number	Furthest Room Name	Checkpoints	Avg. KPM	All KPM	Full Log
2	2026-01-10T12:03:22.074Z	EXPERIMENT (DDA)	14	9:09	63	Philadelphia Experiment	31	34.54	Min1: 61, Min2: 3, Min3: 0, Min4: 0, Min5: 75, Min6: 68, Min7: 42, Min8: 59, Min9: 7, Min10: 1	https://docs.google.com/do

Figure 3.9 Example data sent to the spreadsheet

Tracking the difficulty stage at the first entry into each room makes it possible to follow difficulty changes and determine how many checkpoints were available in each room. The number of activated checkpoints is mainly relevant for the experimental group, offering similar insight into the available checkpoints but also indicating whether participants made use of the provided checkpoints. KPM is recorded based on key release, where a significant drop in a given minute can indicate inactivity or a break during the participant's session. The spreadsheet and all telemetry logs are available in Appendix B. The following chapter describes the post-play questionnaire used to measure the flow experience of participants.

3.4 The Post-Play Questionnaire

An effective manner to measure a participant's flow is by means of a series of questions, known as a Flow Scale, which is completed during or after a flow experience. This research employs the Flow Short Scale (**FSS**) by Rheinberg et al. (2019). Even though other very effective alternative instruments exist, notably Jackson and Ecklund's (2002) 36-item Flow State Scale-2, the FSS was selected for its brevity and suitability for gaming research.

Participants must complete the questionnaire immediately after playing, before the memories of the experience starts to fade. The FSS's 10 flow items (scored on a 7-point Likert scale) can be completed in 30-40 seconds, reducing participant dropout risk due to disinterest, in case the participant got frustrated, while allowing participants to answer spontaneously.

The FSS was designed for the Experience Sampling Method, making it ideal for capturing flow states immediately after gameplay when the experience is still fresh.

Timing is crucial to ensure that the responses reflect the gameplay experience correctly. The FSS provides a single validated flow score (Cronbach's $\alpha = .90$) that can be directly compared between control and experimental groups using independent samples t-tests, making it suitable for this between-subjects design.

The items are as follows:

1. I feel just the right amount of challenge.
2. My thoughts/activities run fluidly and smoothly.
3. I don't notice time passing.
4. I have no difficulty concentrating.
5. My mind is completely clear.
6. I am totally absorbed in what I am doing.
7. The right thoughts/movements occur of their own accord.
8. I know what I have to do each step of the way.
9. I feel that I have everything under control.
10. I am completely lost in thought.

Five additional questions related to the experience are combined with the FSS, to get a clearer image of the participant's experience and frustration. These are also measured on a 7-point Likert scale:

11. I have a lot of general game experience.
12. I have lot of experience with these types of games (puzzle-platformers).
13. I felt frustrated at times.
14. I wanted more checkpoints.
15. I enjoyed this experience.

Finally, a text box is provided for any additional remarks the participant could have. The questionnaire is made using Google Forms. There is one Form for each group: one for the control group and one for the experimental group. The link to the Form is available via the Itch page that the participant receives, to make sure that each group completes the correct questionnaire. The anonymised data gathered using the FSS questionnaire is located in Appendix B. The ethical framework and limitations of this approach are addressed in the following subchapter.

3.5 Ethical Considerations and Methodological Limitations

Participants provide informed consent before beginning the experiment, acknowledging that their gameplay behaviour will be tracked and that they will complete a psychological assessment afterwards for measuring their potential flow.

This addresses what was mentioned in the literature review as well, being one of the potential ethical issues with using DDA in a game: effective DDA systems function best when

applied invisibly, where players remain unaware of any real-time adjustments. While this study only modifies checkpoint placement, which does not affect any skills of the player themselves, it is something that should be considered when designing such systems.

Participants may withdraw from the study at any time by closing the game and contacting the researcher via the same channel through which they received the game link. If a participant wishes to have their data deleted following withdrawal, this will happen manually upon request.

There is a specific methodological issue with designing such a checkpoint-based DDA system. This was stated in a previous chapter but should be repeated as it is imperative to avoid frustration. Players who die and respawn several rooms earlier may struggle with challenges they previously completed, yet the adaptive system cannot spawn checkpoints in already-visited rooms without compromising the invisibility of the DDA system. To mitigate this issue, the modified *VVVVVV* experiment game includes fixed checkpoints in specific rooms regardless of difficulty level, ensuring players never lose more than three or four rooms of progress, depending on the challenge provided in that set of rooms.

Since the experiment is played remotely by participants, it introduces limitations in environmental control. Participants may not maintain consistent focus throughout the session, or they may pause gameplay to take a break, for example, despite instructions to complete the experiment in one sitting. While attention checks could address this, it also risks interrupting players and disrupting the flow state the study aims to measure. Instead, participants are explicitly instructed to complete the session without taking breaks and to notify the researcher if any interruptions do occur. The telemetry tracks timestamps, progress, and key press activity. Participants who take extended or multiple breaks will be flagged during analysis and their data will be excluded.

No personally identifiable information is collected through the experiment. The telemetry system records only numerical gameplay data. Participant names are noted privately by the researcher solely to track contacts and differentiate datasets during collection, and are not included in the final thesis or any published materials. Participants are assigned a numerical ID during data cleaning. Data is stored across Google Sheets, Google Drive, and a local backup, with access limited to the researcher and supervisors. Following the conclusion of the research, anonymised datasets will be made publicly available via GitHub to support transparency and replicability, containing no personal information. No email addresses or contact details are captured at any point, including via the Google Form questionnaire.

3.6 Experiment Procedure and Analysing Data

This subchapter outlines the experimental procedure and subsequent data analysis. After contacting a participant, they must agree to the conditions of having their data tracked during the gameplay and to complete a psychological analysis of their experience in the form of the questionnaire. They are randomly assigned to either the control group (fixed checkpoints) or the experimental group (adaptive checkpoints), and they will receive a link to the Itch page of the appropriate version of the game. They will either receive the remote version for the control group (RC) or the remote version of the experimental group (RE), which is playable in their browser. If the remote version of the game does not work due to technical or connection issues, they will receive the local version (LC or LE), which is a downloadable .zip file. For the local version, the player must manually send the data file.

After the participant either quits or completes the game, they must fill out the FSS questionnaire as soon as possible after the game experience. They will fill in one of two versions: version C (control) or version E (experimental). The link to the questionnaire corresponding to their participant group is available on the same Itch page. Both include identical contents; however, this way the data for the two groups remain separate.

For each participant, the ten FSS item scores are summed to produce a total flow score ranging from 10 to 70, where higher values indicate stronger flow state experience. These scores are organised into a spreadsheet along with the responses to the additional questions about gameplay experience.

The primary analysis utilises an independent samples t-test to compare mean flow scores from the questionnaires between the two conditions. The FSS total score serves as the dependent variable, while the experimental condition serves as the grouping variable. Statistical significance is determined using $\alpha = 0.05$. This analysis produces group means, standard deviations, the t-statistic, degrees of freedom, and p-value to determine whether the difference in flow scores between groups is statistically significant.

To assess the practical significance of any observed difference, Cohen's d effect size is calculated using the difference between group means divided by the pooled standard deviation. This metric indicates whether the observed difference has practical importance for game design, and not just statistical significance.

Telemetry data, displayed either in the spreadsheet or in a file, serves three purposes. First, it shows how the DDA system reacted to and interacted with the player, tracking difficulty adjustments and spawning checkpoints. Second, it identifies behavioural patterns such as

struggles or time spent that may correlate with flow scores and can be used to further compare the two groups. Third, it can be used to create charts and heatmaps of where players struggled or succeeded, visualising where players experienced the most difficulty and how their performance evolved. The collected data is presented and analysed in the following chapter.

4 EXPERIMENT DATA AND RESULTS

This chapter presents the data collected from the control and experiment groups across all measures. Results are reported per hypothesis, preceded by an overview of the data collection, the analytical approach, and descriptive statistics.

4.1 DATA COLLECTION

63 participants were recruited via Discord, WhatsApp, and personal contact. Prior to accessing the game, each participant confirmed having read the instructions, which were also displayed on the game page itself. After doing so, they received access to one of the two versions of the game experiment.

Assignment to conditions was determined by a coin flip for a participant, with the following participant assigned to the opposite condition in order to maintain equal group sizes. When at equal sizes again, a new participant is once again assigned using a coin flip.

The datasets of two participants were excluded from the analysis for separate reasons. One participant did not follow the instructions properly, which led to their data being insufficient. Instead of first reading the instructions, they immediately opened the simultaneously received link, playing in short bursts instead of playing in one sitting as intended. This hampered potential flow, meaning that their data could not be used. The other participant experienced their device shutting down due to lack of power right before the end point of the game experiment. They were allowed to play again after a week's time passed, however their performance metrics showed results consistent with prior exposure to the game and high experience, and their dataset was therefore also excluded.

Oversights were identified in the telemetry system, however most were insignificant. Room 8 never provided any data, room 9 did not count time correctly since it had to be entered twice but only count for the second time and room 41 was missing for some participants since it was possible to enter room 42 before room 41, skipping room-specific data collection for that room. This did not form an issue since this room did not provide any challenge. The individual per-room death tracking capped at one hundred, however this turned out to be an underestimation. Due to an unexpectedly high number of deaths, death locations and timestamps were missing in rooms with high death numbers. None of these issues affected the variables used in the primary analysis, as totals were tracked separately too and per-room death data was not used. The final sample consisted of 61 participants, 30 assigned to the control condition and 31 to the experimental condition.

4.2 Analytical Approach

This research used Jamovi as statistical software. All analyses were conducted at $\alpha = .05$. Independent samples t-tests were used as the primary method for comparing group means for outcome variables, with Welch's t-test applied when Levene's test indicated a violation of the assumption of equal variances. Two-way ANOVA was used to further examine interaction effects between group condition and experience level. Fisher's exact test was used to compare completion rates between groups. A Pearson correlation was conducted to examine the relationship between total deaths and the total scores of the Flow Short Scale (FSS). Normality was assessed using the Shapiro-Wilk test. The Jamovi files used for analysis, along with the graphs it produced, are available via the link in Appendix B.

Using the self-reported game experience and platformer genre experience from the post-play questionnaire, participants were divided into three experience-based subgroups within each of the two conditions. The High Experience (High Exp) group consisted of participants who reported scores of 5-7 on both measures (control n = 12, experiment n = 12). The group referred to as "Game Experience Only" (Game Exp Only) contains participants who reported 5-7 for game experience but 1-4 for platformer experience (control n = 14, experiment n = 12). The Low Experience (Low Exp) group comprised all remaining participants, who reported 1-4 on both measures (control n = 4, experiment n = 7). These subgroups were created to examine whether the effect of the adaptive checkpoint system differed by prior gaming experience or varying experience levels. Due to their small sizes, all subgroup analyses should be considered exploratory as they are underpowered.

4.3 Descriptive Statistics

Table 4.1 Descriptive game and questionnaire data comparison between control and experiment groups

Variable	Control Group (n=30)		Experiment Group (n=31)		Cohen's D
	Mean	Standard Deviation	Mean	Standard Deviation	
FSS Total	52.43	7.71	53.16	8.28	-0.09
FSS Fluency	31.93	5.65	31.61	5.63	0.06
FSS Absorption	20.50	3.37	21.55	4.46	-0.26
Total Deaths	139.57	47.84	187.13	119.54	-0.52
Total Time (s)	1067.50	295.46	1394.50	665.40	-0.64
Checkpoints Activated	42.20	1.10	37.16	4.01	1.72
Frustration	4.00	1.93	4.90	1.83	-0.48
Enjoyment	6.07	0.98	6.16	0.93	-0.10
Game Experience	6.07	1.36	5.65	1.64	—

Platformer Genre Experience	4.07	1.95	3.65	1.85	—

Note. Frustration, Enjoyment, Game Experience and Platformer Genre Experience was measured on a 1-7 Likert scale. Negative d values indicate higher scores in the experiment group. Dashes indicate variables not subject to a primary group comparison t-test.

Flow-related data is comparable between groups. The game data, however, varies considerably. Reported frustration and enjoyment differs slightly, as does the self-reported game experience.

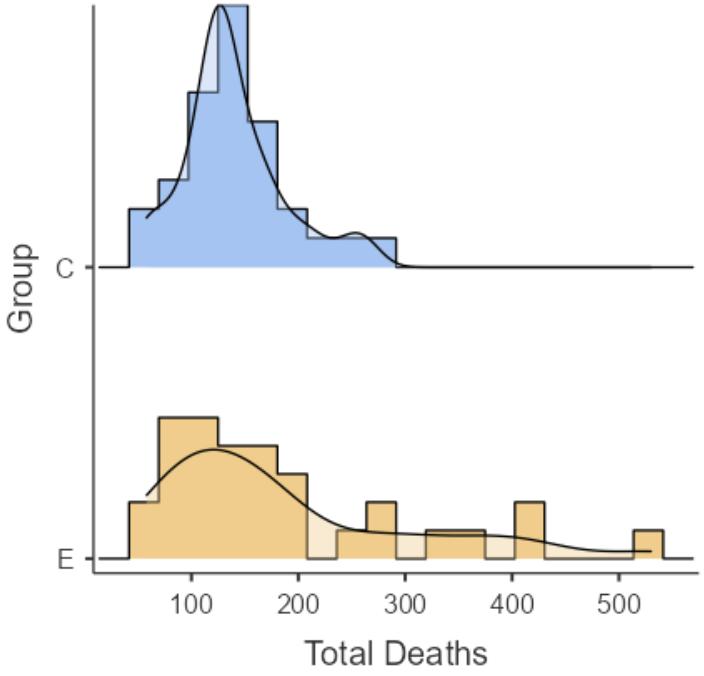

Figure 4.1 Histograms of total deaths between the two groups

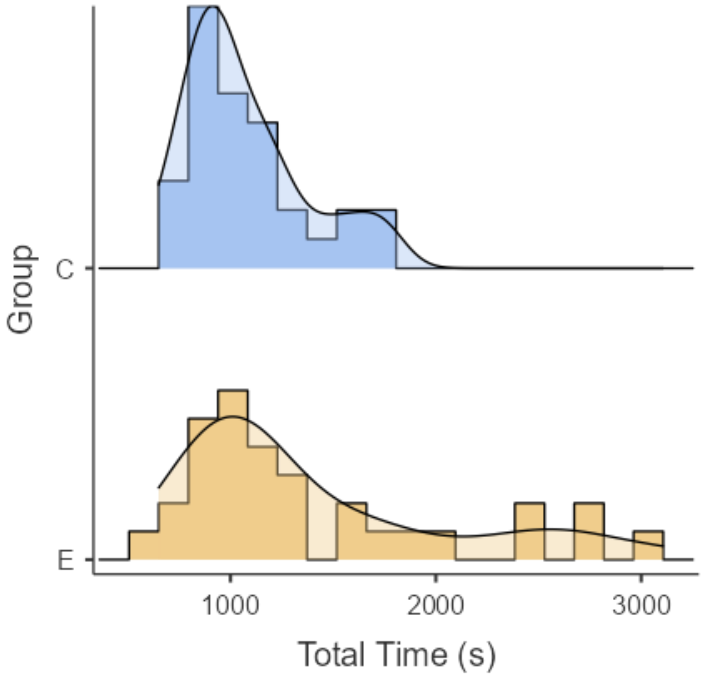

Figure 4.2 Histograms of total time between the two groups

These histograms show greater variance in total deaths (Figure 4.1) and total time (Figure 4.2) in the experiment (E) group compared to the control (C) group, while the control group shows a tighter distribution around lower values for both variables. The data from the participants of the experiment group is more spread out and reaches higher values, which points to these participants having had very different experiences.

4.4 Hypothesis 1: Flow

Hypothesis 1 predicted that participants using the adaptive checkpoint system would report higher flow intensity than those using the fixed checkpoints. This hypothesis was not supported by the results.

Table 4.2 Independent samples t-test results for FSS Total by group and experience level

	Overall (nC=30, nE=31)	**High Exp (nC=12, nE=12)**	**Game Exp Only (nC=14, nE=12)**	**Low Exp (nC=4, nE=7)**
Mean (C)	52.43	53.42	52.14	50.50
Mean (E)	53.16	58.00	48.75	52.43
t	-0.36	-1.28	1.14	-0.73
df	59	22	24	9
p	.724	.215	.266	.482
Cohen's d	-0.09	-0.52	0.45	-0.46

Note. All tests used Student's t unless otherwise noted. Means are included for subgroup comparisons only, as overall group means are reported in Table 4.1. Negative d values indicate

higher FSS Total scores in the experiment group. All subgroup comparisons are exploratory due to small sample sizes. No comparisons reached statistical significance (α = .05).

No significant difference in FSS Total score was found between the control and experiment groups overall, t(59) = -0.36, p = .724, d = -0.09. Subgroup analyses revealed no significant differences across any experience level. The High Exp subgroup showed the largest descriptive difference between groups (M = 58.00 vs M = 53.42, d = -0.52).

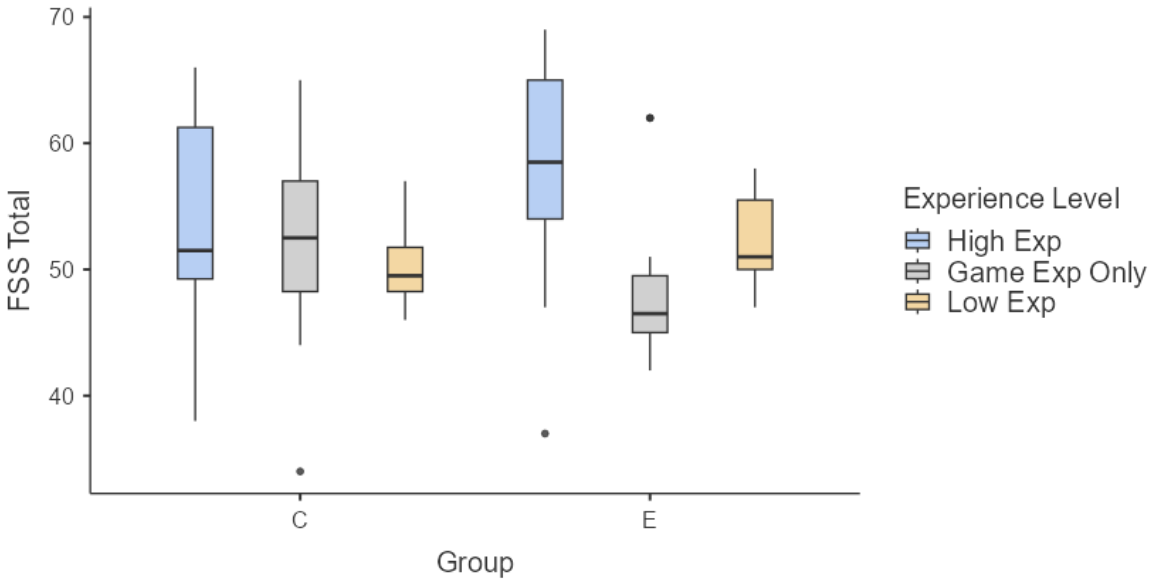

Figure 4.3 FSS Total score distributions by group and experience level displayed as a boxplot. Boxes represent the interquartile range and horizontal lines indicate the median.

FSS Total scores were further examined by experience level to assess whether the distribution of flow scores differed across subgroups. Figure 4.3 illustrates the distribution of FSS Total scores by experience level and group condition.

Table 4.3 Two-way ANOVA interaction effects (experience level × group) on FSS subscales

FSS Variable	F	df	p	η^2	η^2p
FSS Total	1.72	(2, 55)	.188	0.053	0.059
FSS Fluency	2.30	(2, 55)	.110	0.069	0.077
FSS Absorption	0.36	(2, 55)	.698	0.012	0.013

To examine whether the effect of the adaptive system on flow differed by experience level, a two-way ANOVA was conducted (Table 4.3). This revealed no significant interaction between experience level and group condition for FSS Total, F(2, 55) = 1.72, p = .188, η^2p = .059, or for either subscale. Subgroup analyses on FSS Fluency and Absorption showed no significant differences.

4.5 Hypothesis 2: Completion

Hypothesis 2 predicted that participants in the adaptive checkpoint condition would show higher completion rates compared to the fixed checkpoint condition. This hypothesis was not supported for completion rates, though significant differences were found across the behavioural variables. In addition to completion rates, the variables total deaths, total time, and checkpoints activated, were examined to assess whether the adaptive system produced other measurable differences in gameplay behaviour.

Table 4.4 Completion rates by group

Group	Completed	Did not complete	Completion rate	p (Fisher's exact)
C (n=30)	29	1	97%	.612
E (n=31)	28	3	90%	

Note. No significant difference in completion rate was found between groups (α = .05).

Table 4.5 Behavioural measure comparisons between groups

Variable	t	df	p	Cohen's d
Total Deaths	-2.05	39.6	.047	-0.52
Total Time (s)	-2.49	41.7	.017	-0.64
Checkpoints Activated	6.74	34.6	<.001	1.72

Note. All tests used Welch's t due to violation of the assumption of equal variances (Levene's p < .05). Negative d values indicate higher scores in the experiment group. Significant results (α = .05) are shown in bold.

Completion rates were high in both groups, with no significant difference between them, Fisher's exact p = .612. The experiment group also accumulated significantly more deaths, t(39.6) = -2.05, p = .047, d = -0.52, and spent significantly more time completing the game, t(41.7) = -2.49, p = .017, d = -0.64. The experiment group activated significantly fewer checkpoints than the control group, t(34.6) = 6.74, p < .001, d = 1.72.

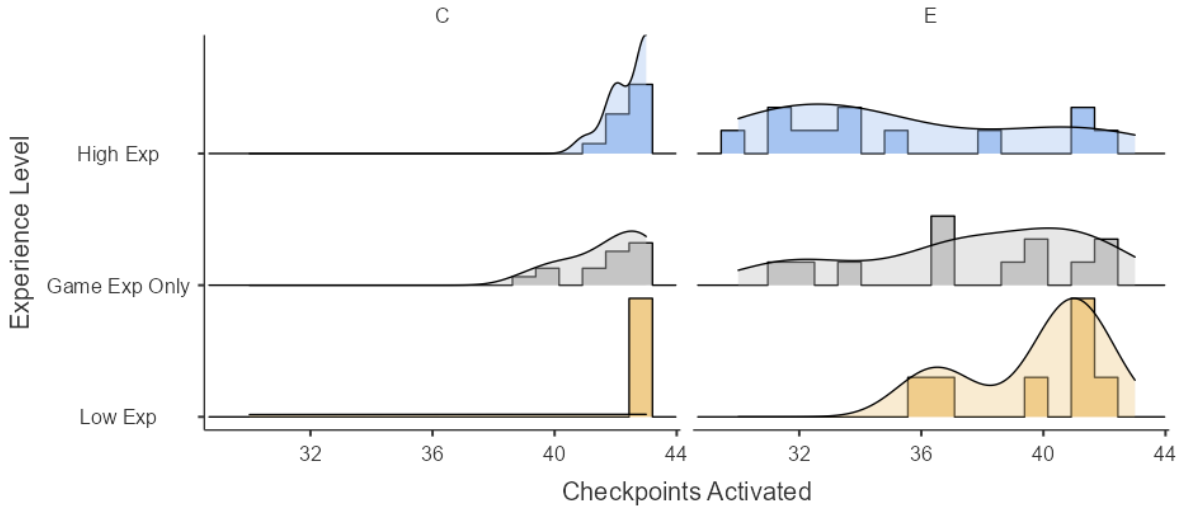

Figure 4.4 Checkpoints activated by group and experience level

Figure 4.4 shows the variance in checkpoints activated within the experiment group, which indicates the effect of the adaptive system: participants assessed as performing well received fewer checkpoints, while those who struggled had more checkpoints present.

4.6 Hypothesis 3: Frustration

The original formulation of Hypothesis 3 predicted that players using adaptive checkpoints would activate most of the checkpoints provided to them. This was revised prior to data collection to better align with the study's primary focus on player experience, as well as the data that was gathered. The original checkpoint activation behaviour is addressed as part of the behavioural analysis in Subchapter 4.5. The revised hypothesis 3 predicted that participants using the adaptive checkpoint system would report different levels of frustration compared to those using fixed checkpoints. This hypothesis was not supported overall, though subgroup analyses revealed significant differences across experience levels.

Table 4.6 Frustration and enjoyment comparisons between groups overall and by experience level

Comparison	Frustration				Enjoyment			
	t	df	p	Cohen's d	t	df	p	Cohen's d
Overall (nC=30, nE=31)	-1.88	59	.066	-0.48	-0.39	59	.701	-0.10
High Exp (nC=12, nE=12)	0.27	22	.790	0.11	0.00	22	1.000	0.00
Game Exp Only (nC=14, nE=12)	**-3.54**	24	**.002**	**-1.39**	0.91	24	.370	0.36
Low Exp (nC=4, nE=7)	-0.74	9	.479	-0.46	**-3.55**	9	**.006**	**-2.22**

Note. All tests used Student's t. Negative d values indicate higher scores in the experiment group. Overall frustration approached but did not reach significance (α = .05). All subgroup comparisons are exploratory due to small sample sizes. Significant results are shown in bold.

Observing Table 4.6, the overall frustration did not reach statistical significance, t(59) = -1.88, p = .066, d = -0.48, though the result approached the threshold. Subgroup analysis revealed a significant frustration difference in the Game Exp Only group, with experiment group participants reporting notably higher frustration (M = 5.83) than control group participants (M = 3.79), t(24) = -3.54, p = .002, d = -1.39. Overall enjoyment did not differ significantly between groups, t(59) = -0.39, p = .701, d = -0.10. Low Exp participants in the experiment group reported significantly higher enjoyment (M = 6.43) than their control group counterparts (M = 5.00), t(9) = -3.55, p = .006, d = -2.22. No other subgroup comparisons reached significance.

Table 4.7 Additional checkpoint desire comparison between groups and experience level

Comparison	**Additional Checkpoint Desire**			
	t	**df**	**p**	**Cohen's d**
Overall (nC=30, nE=31)	**-2.39**	59	**.020**	-0.61
High Exp (nC=12, nE=12)	-0.49	22	.628	-0.20
Game Exp Only (nC=14, nE=12)	-1.78	24	.088	-0.70
Low Exp (nC=4, nE=7)	-1.55	9	.156	-0.97

Note. All tests used Student's t. Negative d values indicate a higher desire for more checkpoints in the experiment group. Only the overall comparison reached statistical significance (α = .05). All subgroup comparisons are exploratory due to small sample sizes.

Experiment group participants reported a significantly higher desire for more checkpoints overall, t(59) = -2.39, p = .020, d = -0.61. No significant differences were found across any experience level subgroup, though the Game Exp Only subgroup showed the largest descriptive difference, t(24) = -1.78, p = .088, d = -0.70. All subgroup comparisons remained non-significant.

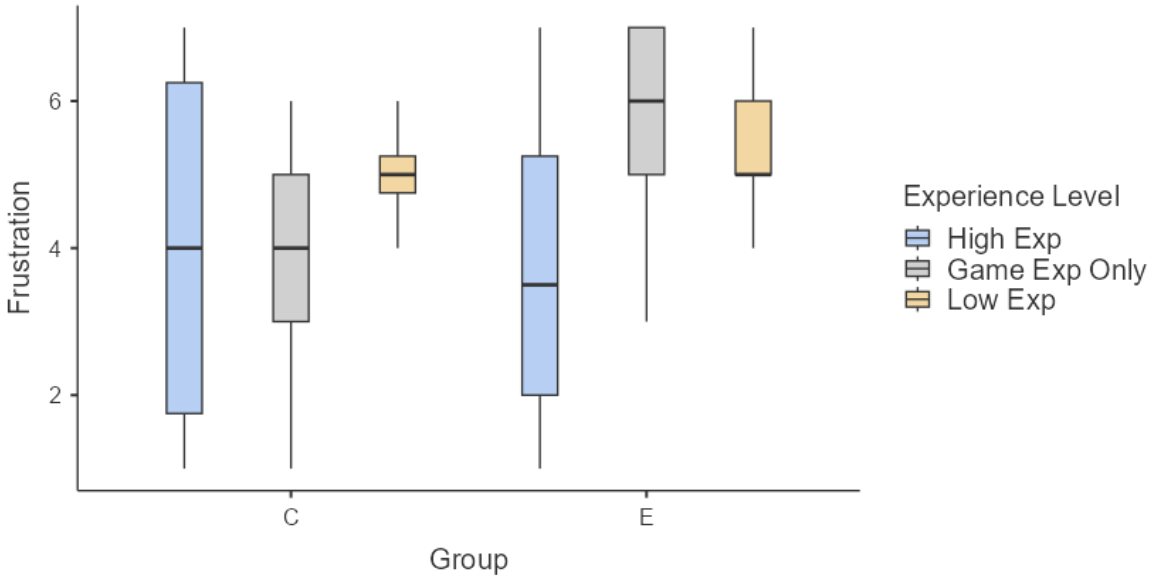

Figure 4.5 Frustration scores by group and experience level

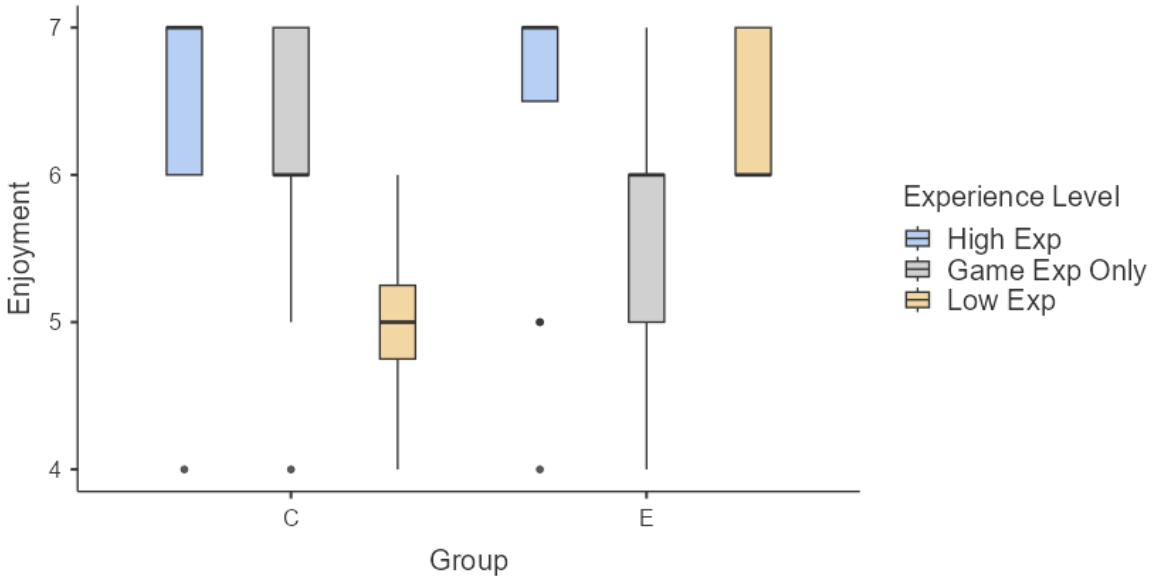

Figure 4.6 Enjoyment scores by group and experience level

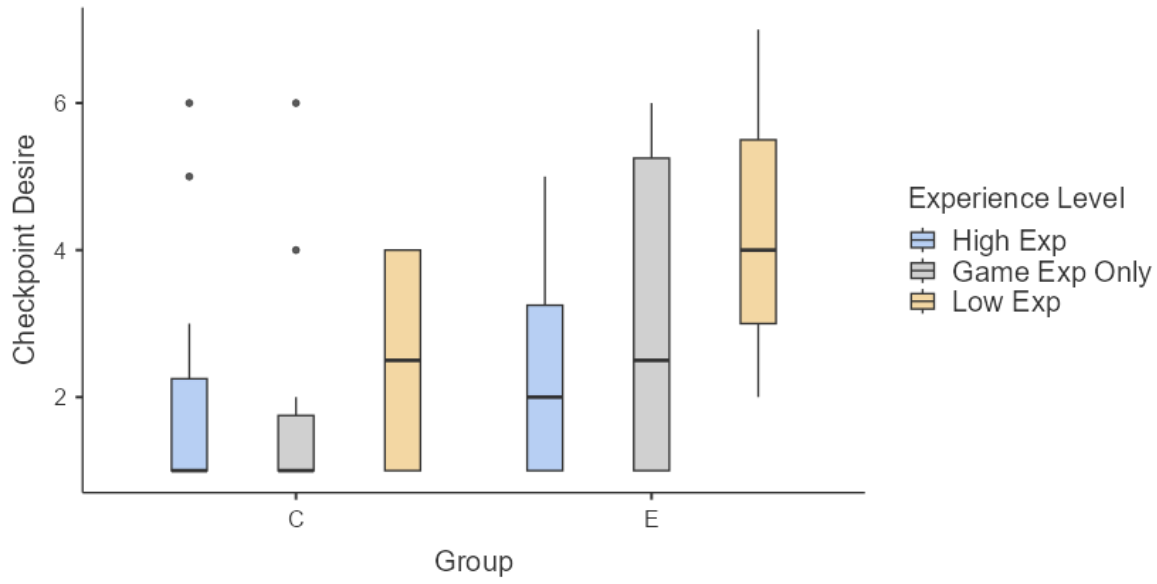

Figure 4.7 Additional checkpoint desire comparison between groups and experience level

The boxplot in Figure 4.5 illustrates that frustration scores were notably higher and more consistent in the Game Exp Only subgroup of the experiment group compared to the control group, while the High Exp subgroup showed greater spread in both conditions.

Figure 4.6 shows that High Exp and Game Exp Only participants showed comparable enjoyment scores across both groups, while Low Exp participants in the experiment group showed notably higher scores with less variability than their control group counterparts.

The boxplot in Figure 4.7 illustrates that most of the participants who wanted more checkpoints to be present in their experience were generally part of the experiment group. This was true for all experience levels, with Game Exp Only and Low Exp participants showing the most notable increases compared to their control group counterparts. High Exp participants showed relatively low checkpoint desire in both groups, with comparable medians, while the experiment group showed greater spread overall.

4.7 Exploratory Analysis

To examine whether increased deaths affected flow experience, a Pearson correlation was conducted between total deaths and FSS Total score. No significant relationship was found, $r(59) = -.055$, $p = .674$.

The number of difficulty increases and decreases experienced by experiment group participants was examined by experience level to assess how the adaptive system responded across groups.

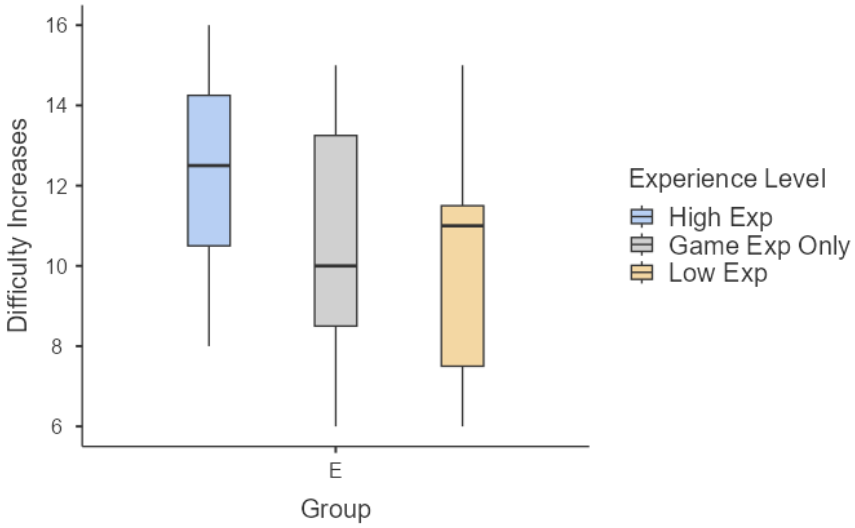

Figure 4.8 Difficulty increases by experience level in the experiment group

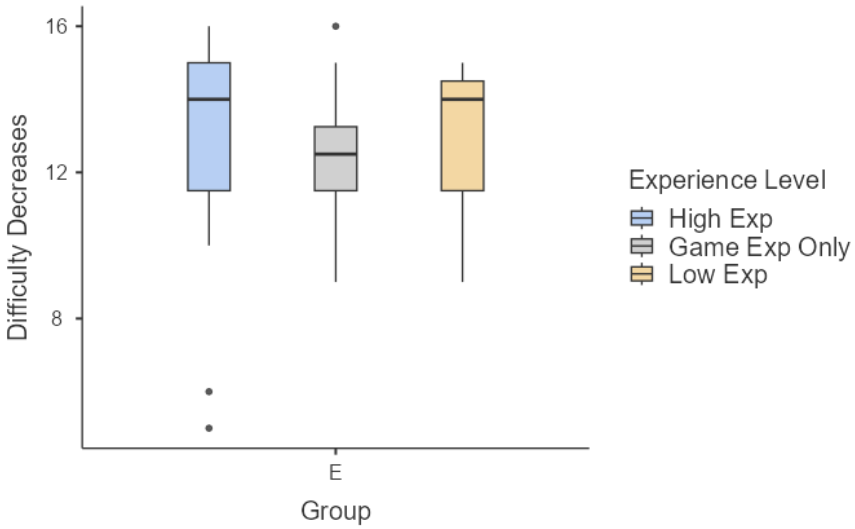

Figure 4.9 Difficulty decreases by experience level in the experiment group

High Exp participants experienced the greatest number of difficulty increases (Figure 4.8), with Low Exp participants showing the most variability in this measure. For difficulty decreases (Figure 4.9), High Exp and Low Exp participants received comparable numbers, while Game Exp Only participants received fewer decreases overall. These patterns and the hypotheses data are interpreted further in the discussion chapter that follows.

5 DISCUSSION AND INTERPRETATION

This chapter interprets the findings presented in Chapter 4 in relation to the research question and hypotheses. Each hypothesis is discussed individually, followed by a post-hoc validation of the approach of the adaptive checkpoint system and its threshold calibration, an assessment of the study's limitations, and the implications of the findings for game design.

5.1 HYPOTHESIS 1: FLOW

No significant difference in FSS Total score was found between groups, $t(59) = -0.36$, $p = .724$, indicating that H1 was not supported. However, subgroup analysis revealed a notable descriptive pattern. The High Exp subgroup of the experiment group reported noticeably higher flow scores than all other groups when looking at the boxplot in Figure 4.3. A comparison of means ($M = 58.00$ vs $M = 53.42$) revealed a medium effect size ($d = -0.52$), though this did not reach significance. This pattern hints at the possibility that fewer checkpoints provided a more difficult challenge which matches more appropriately with the skill of highly experienced players, aligning with Csikszentmihalyi's challenge-skill balance as a condition for flow. Chen similarly argues that the optimal level of challenge differs for each player, suggesting that systems offering less guidance to more skilled, or at least experienced, players may better support their Flow Zone.

This was reflected by some experiment group participants, with two Game Exp Only participants specifically noting that the challenge level felt perfect to them. In their remark, participant 35 stated that the game was well-designed and the challenges felt exactly right. Participant 54 wrote: “It was a fun game with just the right amount of challenge! I enjoyed it.” Notably, both participants belonged to the subgroup that reported the lowest average flow scores overall, and their individual FSS scores differed considerably (62 and 45 out of 70 respectively).

This suggests that even within a subgroup showing lower average flow, individual experiences varied substantially, and perceived challenge appropriateness did not necessarily translate into higher measured flow. The Game Exp Only group who used the adaptive system experienced the least amount of flow, even lower than the participants with lower experience. This further suggests that the DDA system made the game harder than optimal for certain participants who lacked experience with platformer games. The means of the Low Exp subgroup suggest that Low Exp participants in the experiment group experienced higher flow than Game Exp Only participants in the experiment group, although they had less experience overall and likely struggled more.

The ANOVA revealed no significant interaction with $F(2,55) = 1.72$, $p = .188$, but the partial $\eta^2 = .059$ suggests a small-to-medium effect the study was underpowered to detect. No significant interactions were found for any of the FSS subscales either, with FSS Fluency $F(2, 55) = 2.30$, $p = .110$, and FSS Absorption $F(2, 55) = 0.36$, $p = .698$, consistent with the overall null result.

Although no significant overall difference was found between groups, the descriptive patterns across experience levels suggest experience-dependent effects that the study was insufficiently powered to detect. Due to the small sample sizes, concrete conclusions cannot be drawn, so replication with adequately powered subgroups would be necessary to confirm them.

5.2 Hypothesis 2: Completion

Completion rates were comparable across groups, with 97% of control group participants and 90% of experiment group participants finishing the game. H2 was not supported, as no significant difference in completion rates was found between groups, with Fisher's exact test $p = .612$. Behavioural measures confirmed that the DDA system had a measurable effect on gameplay. These rates indicate that the game for the experiment group did not break motivation and maintained engagement.

Besides completion rates, three behavioural measures revealed significant differences between groups. The experiment group accumulated significantly more deaths than the control group, $t(39.6) = -2.05$, $p = .047$, $d = -0.52$, and spent significantly more time completing the game, $t(41.7) = -2.49$, $p = .017$, $d = -0.64$, indicating that the adaptive system produced a harder overall experience. The experiment group activated fewer checkpoints, $t(34.6) = 6.74$, $p < .001$, $d = 1.72$, confirming that the system functioned as intended by reducing checkpoint availability for participants performing well.

As shown in Figure 4.4, the distribution of checkpoints activated across experience levels further supports this. Low Exp participants in the experiment group activated the most checkpoints, Game Exp Only participants activated fewer, and High Exp participants activated the least. This pattern aligns with the intended behaviour of the system: providing more support to players who struggled and less to those who performed well, indicating that the DDA responded meaningfully to differences in player performance across experience levels.

The behavioural data suggests that the DDA system was functioning as intended but acted too harshly overall. While it successfully reduced checkpoints and increased challenge, the increased deaths and time indicate the difficulty was pushed beyond what was optimal for

certain participants. The comparable completion rates, however, suggest that motivation was preserved despite the harder experience. This is partly attributable to the strong game and level design of *VVVVVV*, likely softening the negative impact of checkpoint removal.

Several participants reflected on the checkpoint experience in their remarks. Participant 23 described completing a difficult section only to find no checkpoint immediately after, noting that the relief of finally getting through was quickly replaced by the realisation of having to repeat the section if they died again. Participant 12 mentioned repeatedly having to redo an easy section just to get to the actually challenging room because the checkpoint was placed at a room before the hard part, instead of right before the hard part. Even participant 7 suggested that fewer checkpoints in certain places would have made the game more enjoyable, indicating that the system's reductions were not always felt negatively and weren't always excessive. Participant 8, from the control group, nearly quit before a particularly relentless section but decided to push through regardless.

These remarks point to a fundamental challenge of this study's approach. The DDA system was added to a game that was already fully designed with fixed checkpoints in mind. This meant the main approach was to remove existing checkpoints, adding new ones only in some rooms. A system like this would be better suited to a game designed with adaptive checkpoints in mind, where checkpoint positions could be planned around the possibility of adjustment, rather than being constrained by a design built around fixed checkpoints from the start. The checkpoint experience of participants also had a direct effect on reported frustration levels, which is examined in the following subchapter.

5.3 Hypothesis 3: Frustration

H3 was not supported overall, as no significant difference in frustration was found between groups, $t(59) = -1.88$, $p = .066$, $d = -0.48$, though the result approached the threshold. At the subgroup level, Game Exp Only participants in the experiment group reported significantly higher frustration than their control group counterparts, $t(24) = -3.54$, $p = .002$, $d = -1.39$. Given that this group had game experience but less familiarity with platformer games specifically, the removal of checkpoints likely pushed the difficulty beyond what was optimal for them, moving them toward the anxiety side of the flow channel as described by Csikszentmihalyi.

The enjoyment boxplot in Figure 4.6 also shows a slight descriptive decrease in enjoyment for Game Exp Only participants in the experiment group compared to the control group, though this also did not reach significance, $t(24) = 0.91$, $p = .370$. This is consistent with

Figure 4.5, where Game Exp Only participants in the experiment group also showed higher frustration scores. Participants who experienced the most frustration also trended toward slightly lower enjoyment, though not significantly so. This was reflected by participant 23 who played the experiment version of the game. They noted that completing a difficult section without a checkpoint immediately following it meant getting frustrated by the thought of having to restart from further back than expected upon dying, illustrating how checkpoint placement directly affected the experience of challenge and level of frustration experienced in the experiment group.

Interestingly, Low Exp participants in the experiment group reported significantly higher enjoyment than their control counterparts, $t(9) = -3.55$, $p = .006$, $d = -2.22$, despite reporting comparable frustration. As shown in Figures 4.8 and 4.9**,** Low Exp participants mostly received difficulty decreases and remained at lower difficulty levels throughout, with only occasional increases when performance remained below the thresholds. This means their experience was generally closer to the base game's fixed checkpoint experience where more checkpoints were present, with slightly more challenge introduced at times if they performed well. This more gradual and supported experience may be what contributed to their higher enjoyment, as the adaptive system kept the difficulty within a range they could manage while still providing moments of added challenge.

Chen's concept of the Flow Zone is relevant here: challenge that sits just at the edge of a player's ability can feel more rewarding than a comfortable experience, and for Low Exp players the adaptive system appears to have landed closer to that point than the fixed checkpoint version did for most. This shift is clearly visible in Figure 4.6, where the Low Exp experiment group box moves substantially upward compared to the control group, while High Exp participants show nearly identical distributions across both conditions in both Figures 4.5 and 4.6.

It is also important to note that the lack of experience does not necessarily accurately reflect the skill of the player. This finding should, again, be treated as exploratory given the small subgroup sizes involved.

Experiment group participants reported a significantly higher desire for more checkpoints, $t(59) = -2.39$, $p = .020$, $d = -0.61$, which is expected given that their version specifically provided fewer checkpoints. This does not contradict the enjoyment finding, as wanting more checkpoints and enjoying the experience can both be the case. High Exp participants maintained comparable enjoyment levels despite the increased difficulty,

suggesting the adaptive system did not negatively affect their experience even when it pushed challenge higher.

5.4 Post-hoc Validation

To assess whether the DDA system's threshold calibration was appropriate, the 70th percentile (**P70**) death counts of the control group were compared to the set thresholds for each active room. The P70 value represents the point at which 70% of control group participants performed below, and was used as an approximation of average struggle. Thresholds set close to this value would trigger struggle detection for players performing below the average fixed-checkpoint player. This validation was conducted post-hoc, as no preliminary calibration testing was completed prior to the study. The full post-hoc validation spreadsheet is available in Appendix B.

Room	death_threshold	s-time_thresh	l-time_thresh	C_P70_deaths	C_P70_cleartime	E_P70_deaths	E_P70_cleartime	Difficulty	Verdict
9	2	10	20	2.00	18.0	3.00	23.0	**Moderate**	Well-calibrated
10	1	17	30	2.00	20.0	2.00	25.0	**Moderate**	Sensitive calibration. This was done on purpose to catch p
11	3	18	30	4.30	36.3	7.00	55.0	**Hard**	Also sensitive, meant to catch struggling players early.
12	INACTIVE	INACTIVE	INACTIVE	0.300	9.00	1.00	10.0	Easy	INACTIVE
13	1	0	10	1.00	8.00	1.00	9.00	**Easy**	Well-calibrated, simple room, having to dodge an enemy
14	INACTIVE	INACTIVE	INACTIVE	0.00	2.00	0.00	3.00	Trivial	INACTIVE
15	2	0	15	3.30	13.6	3.00	15.0	**Moderate**	Sensitive room thresholds again, meant to make the upco
16	2	0	15	16.0	9.00	21.0	10.0	**Very Hard**	Although deaths are high here, only the first attempt to re
17	2	0	600	7.00	11.3	9.00	13.0	**Hard**	Hard room where player falls through, so time threshold is
18	2	0	10	4.00	79.6	5.00	94.0	**Moderate**	Also a sensitively calibrated room. P70 for time heavily in
19	INACTIVE	INACTIVE	INACTIVE	0.00	10.3	0.00	12.0	Easy	INACTIVE
20	1	5	15	2.00	6.00	4.00	8.00	**Moderate**	Checkpoint before final section of level 1. Slightly sensitiv
21	3	10	20	3.00	33.2	5.00	40.0	**Moderate**	Well-calibrated, dividing more skilled players from the les
22	INACTIVE	INACTIVE	INACTIVE	0.00	26.6	0.00	31.0	No Challenge	INACTIVE (End of level 1)
23	INACTIVE	INACTIVE	INACTIVE	0.00	10.0	0.00	9.00	No Challenge	INACTIVE (Beginning of level 2)

Figure 5.1 Section of the post-hoc validation spreadsheet depicting the calibrated thresholds and the P70 values of both groups

Of the 32 active rooms, 10 were well-calibrated, 16 were intentionally sensitive and strict to focus on the first visit and to compensate for their non-linear design, 4 required adjustment, and 2 were not applicable due to thresholds being present but inactive. A section of the validation can be seen in Figure 5.1.

The difficulty change patterns shown in Figures 4.8 and 4.9 indicate how the system responded to different experience levels. High Exp participants experienced the greatest number of difficulty increases, meaning the system repeatedly identified their performance as above threshold and raised the difficulty accordingly. Low Exp participants showed the most variability in difficulty increases but accumulated comparable decreases to High Exp participants, suggesting the system detected their struggle but was not always able to raise difficulty again once reduced.

Game Exp Only participants received fewer changes in both directions, indicating less overall responsiveness of the system to this group. This is consistent with the interpretation

from Subchapter 5.1, where Game Exp Only participants in the experiment group also showed the lowest average flow scores.

A Pearson correlation between total deaths and FSS Total score found no significant relationship, $r(59) = -.055$, $p = .674$, meaning that accumulating more deaths did not reduce flow experience. This supports the interpretation that players tolerated the increased difficulty without a measurable negative effect on flow.

Four rooms presented calibration issues that warrant specific attention. Room 33, a moderately hard section requiring repeated gravity flipping from one side of a platform to the other, had a death threshold of 1 while the control group P70 was 3.0, meaning the threshold triggered for most players and produced unnecessary difficulty decreases. Room 47, a hard section repurposed from an optional challenge in the original game into a mandatory room, had a threshold of 5 while the control group P70 was 7, making it too strict. Room 55, positioned directly before the final few rooms, had a threshold of 4 while the P70 was 2 because it was interconnected with the previous room where most of the deaths happened, meaning the thresholds for this room should've been stricter. Room 56 is a straightforward section requiring only sustained forward movement, and clearing it easily triggered a difficulty increase that could remove a checkpoint before the hardest rooms in the game. This room should have been disabled from DDA evaluation entirely, making players always respawn in room 57 instead, which was the hardest room of the experience. Participant 7, who was part of the experiment group, noted that fewer checkpoints in certain places would have made the game more enjoyable. They suggested that the system's reductions were not always experienced negatively and that the checkpoint calibrations were generally reasonable for experienced players.

A further constraint on this validation is that several rooms, as mentioned earlier, can be revisited after first being cleared and are in some cases interconnected. Deaths accumulated during revisits inflate P70 values for those rooms, making thresholds appear stricter than they functionally were for first-attempt evaluation. This is a constraint of conducting the validation post-hoc on a non-linear game rather than a reflection of the threshold logic itself.

Overall, this post-hoc validation suggests that the threshold calibration for this approach was broadly appropriate for the scope of this study, though specific rooms and the non-linear structure of the game introduced constraints that affected the system's effectiveness in practice.

5.5 LIMITATIONS

The research was restricted by several limitations. These should be considered when interpreting the findings of this study. As explained in the methodology, the experiment was played in a browser using WebAssembly. This introduced input lag for a small number of participants and was brought up in the remarks by several participants. It affected precision movement and may have influenced FSS scores, frustration, or enjoyment.

The non-linear layout of *VVVVVV* created recurring challenges for the DDA system and its tracking. Several rooms can be revisited after first being cleared, meaning deaths accumulated during revisits inflated per-room death counts and contaminated P70 threshold values. This made it difficult to accurately assess performance in those rooms.

As the DDA system was retrofitted onto an already fully designed game with fixed checkpoints in mind, there were limiting factors. This constrained the approach to removing existing checkpoints rather than placing them with adaptive logic in mind.

Flow was measured exclusively through the Flow Short Scale, a self-report questionnaire completed after the session. Self-reported flow is inherently interpretive and subject to recall bias, and a single post-play measurement cannot fully capture how flow fluctuated throughout the session, given that flow is itself a complex concept. Complementing the FSS with in-session behavioural measures or repeated measurements would provide a more complete picture of each participant's flow experience.

Finally, the three experience-based subgroups were small, ranging from 4 to 14 participants per condition. No Bonferroni correction was applied to the multiple comparisons across subgroups, meaning the risk of Type 1 error is present. Notably, the two significant subgroup findings fall well below even a conservatively corrected threshold, suggesting they are unlikely to be Type 1 errors. As stated earlier, all subgroup analyses should be considered exploratory, and the patterns observed cannot be generalised.

5.6 IMPLICATIONS

The findings of this study suggest that adaptive checkpoint systems have potential as a tool for managing player experience, but that the implementation and approach significantly affect the outcome. The DDA system functioned as intended in many respects, successfully reducing the amount of checkpoints for more experienced, and likely more skilled, players and providing more for those who tended to struggle. However, it can be concluded that the system was calibrated too strictly overall.

A more linear game layout would allow more accurate threshold calibration and avoid issues with revisitable rooms. This type of system would be better suited to a game designed with these adaptive checkpoints in mind from the start of development, where checkpoint placement and threshold calibration can be planned around adaptive logic rather than constrained by an existing fixed design.

The results show divergent effects across experience levels, with High Exp, Game Exp Only, and Low Exp participants responding differently to the adaptive system. Descriptive observation pointed towards High Exp players reporting the highest observed flow scores in the experiment group, as visible in Figure 4.3, suggesting that a reduced checkpoint amount may have better matched their skill level. Low Exp players reported significantly higher enjoyment in the experiment group despite the frustration levels not differing significantly between groups. This indicates that the added challenge was experienced as satisfying rather than overwhelming for this group. It is important to note as well that low experience does not necessarily equate to low skill. Game Exp Only players, on the other hand, reported significantly more frustration and trended toward lower enjoyment, suggesting the system pushed the challenge beyond their optimal level.

A more sophisticated and extensive DDA approach that first accurately identifies a player's experience and skill before applying difficulty adjustments could address this. Rather than applying a uniform threshold-based approach to all players, a system that adapts its behaviour based on an early assessment of its player’s skill would be better positioned to provide effective tweaks and changes. Besides thresholds, alternative approaches tracking more behavioural data or implementing AI could also provide more nuanced and responsive detection. Additionally, applying larger difficulty adjustments in response to prolonged or continuous struggle, rather than single-stage changes, could make the system more responsive to struggling players. Further research with larger samples, adequately powered experience-based subgroups, and a game designed with adaptive checkpoints from the outset would be valuable to reach more definitive and comprehensive conclusions about the potential of this approach.

6 CONCLUSION AND IMPLICATIONS

6.1 FINDINGS AND CONCLUSION

This study investigated to what extent adaptive checkpoint systems affect the flow channel compared to fixed checkpoints across diverse player experience levels. To determine the effectiveness of adaptive checkpoints, a DDA system was applied to the checkpoints of the game *VVVVVV*. A between-subjects design was used, consisting of two groups, where one group had DDA present and the other did not. Participants' behaviour was observed using telemetry and the quality of their flow experience was measured using a flow scale, incorporated into a post-play questionnaire.

The experimental data suggests that this adaptive checkpoint system did not yield significantly higher overall flow on the Flow Short Scale. Patterns from experience-based subgroups, however, showed noteworthy differences, with the highly experienced subgroup (High Exp) in the experiment group reporting descriptively higher flow scores than their control counterpart (M = 58.00 vs. M = 53.42). The Game Exp Only group, where participants have general game experience but lack experience specific to the platformer genre, reported the lowest mean flow score of any subgroup (M = 48.75). A two-way ANOVA revealed no significant interaction between experience level and group condition on flow scores either, though effect sizes at the subgroup level suggest the system had differential impacts worth investigating further. The first hypothesis was not supported, though subgroup patterns suggest the potential for more meaningful conclusions with a larger sample (see Subchapter 5.1).

When comparing the completion rates of the two groups, they were very comparable, and the second hypothesis was not supported for completion rates. The experiences were, however, noticeably different. Total deaths and time spent were both significantly higher in the experiment group where DDA was active, indicating a more demanding experience. The number of checkpoints present was also much lower in this group. Participants of the High Exp subgroup received a lower number of checkpoints overall, while less experienced (Low Exp) participants had a higher number of checkpoints available, which indicates that the system was able to accommodate the needs of different players, despite making the overall experience more difficult (see Subchapter 5.2).

The final hypothesis was not fully supported, although the overall frustration comparison approached significance ($p = .066$, $d = -0.48$), with experiment group participants experiencing more frustration overall than those in the control group. The Game Exp Only subgroup showed the most pronounced frustration difference, reaching significance between

groups (M = 5.83 vs. M = 3.79, p = .002, d = -1.39). Low Exp participants in the experiment group reported significantly higher enjoyment than their control counterparts (M = 6.43 vs. M = 5.00, p = .006, d = -2.22), despite the overall tendency toward higher frustration in the experiment group. Additionally, experiment group participants overall reported a significantly higher desire for more checkpoints (p = .020, d = -0.61), further supporting the interpretation that the adaptive system produced a more demanding experience. These results suggest the system was beneficial to certain players while disadvantageous to others, depending on experience level (see Subchapter 5.3).

These results indicate that the adaptive checkpoint system designed for this study did not significantly extend the flow channel overall compared to fixed checkpoints. However, behavioural data and subgroup patterns suggest the system differentially affected players depending on experience level, with the most pronounced effects observed at the extremes. The extent of impact was therefore limited but not negligible, and remains inconclusive due to insufficient statistical power. These findings suggest the concept warrants further investigation with a larger, better-powered sample before conclusions about its practical value can be drawn.

6.2 Contribution

This research attempts the first empirical test of an adaptive checkpoint system, evaluated through Flow Theory. Prior to this study, no empirical research had examined adaptive checkpoint systems through the lens of Flow Theory, leaving a gap between DDA research and checkpoint design that this study begins to address by drawing attention to checkpoints as an underexplored adaptive element. The gameplay data and self-reported scores provide behavioural evidence that the adaptive checkpoints affect the play patterns and gameplay quality even when the subjective flow scores do not differ significantly. The subgroup analysis demonstrates that experience and skill level is key to adaptive systems and that it would be beneficial to focus more on determining the specific skill level of a player in order for the system to adapt to the needs of the player (see Subchapter 5.6).

6.3 Limitations

The sample size of 61 participants was insufficient for reliable subgroup conclusions, limiting the interpretive value of the experience-level analyses. No formal power analysis was conducted prior to data collection, meaning the study may have been underpowered to detect small to medium effects (see Subchapter 5.5). As participants played remotely, the conditions of the environments, such as distractions or hardware issues, were uncontrolled and therefore not

fully consistent, which potentially influenced the data (see Subchapter 3.5). One FSS item was reported as ambiguous by several participants, which may have introduced noise into the flow scores. Telemetry data was exported in a single instance at the end of the session, meaning any technical interruption prior to completion resulted in total data loss for that participant. A small number of participants inadvertently submitted duplicate datasets by re-entering the game after completion, requiring manual exclusion during data cleaning. The adaptive system was also only tested in the room-based, 2D platformer *VVVVVV*, and would need further research for other genres. Retrofitting a system onto an existing game constrains the design space, as checkpoint positions and room layouts cannot be fully optimised for the adaptive system from the ground up (see Subchapter 3.5). Finally, the measurement of flow was fully self-reported and, although it utilised a validated measurement tool and was combined with behavioural data, is subject to recall bias.

6.4 Future Directions

Future research should prioritise a larger participant pool to enable more reliable subgroup conclusions and better-powered statistical analyses. Threshold calibration would benefit from further refinement through extended playtesting, with room-specific thresholds adjusted per difficulty stage to take skill levels into account, rather than remaining fixed across all rooms. Greater use of death location data could enable more precise struggle detection, allowing the system to respond to specific obstacles rather than general room performance. While death location data was collected in this study, its application to struggle detection remained limited. The difficulty adjustment logic itself warrants further development: the current system risks stranding players at a higher difficulty level when the player encounters sudden difficulty after a streak of successful rooms. Future implementations should allow for multiple consecutive decreases or increases within a single room. When designing such systems, rooms without meaningful challenge should remain excluded from DDA tracking to prevent unintended difficulty shifts.

Telemetry reliability could be improved through periodic auto-saving throughout the session rather than a single export at the end. The post-play questionnaire could be expanded to include questions about specific games played previously, where participants felt they struggled most, and reasons for early quitting, which would provide richer context for interpreting both flow scores and behavioural data. A general assessment of the participant's experience or skill would also be beneficial, either through a questionnaire or via the game itself. Testing the system across different game genres and in games specifically designed and created with adaptive checkpoints from the ground up would be a valuable next step.

APPENDIX A

EXTENDED GAME DESCRIPTION

VVVVVV is structured around two levels for the purposes of this experiment, mirroring the length of the game's official demo. The first level, Space Station, serves primarily to introduce the controls and core gameplay mechanics. While most of its rooms are accessible to new players, several present genuine challenge, particularly "Stop and Reflect" and a three-room section beginning at "Driller." For clarity, this level is referred to as Space Station throughout, though it is technically named Space Station 1 in the original game. The second section of Space Station is not included in the experiment. The corresponding source file is *Spacestation2.cpp*.

The Laboratory does not represent a sharp difficulty increase at the outset, instead focusing on introducing Gravity Walls and platforming mechanics that make use of them. Challenge escalates toward the end of the level, with rooms such as "I'm Sorry", "Please Forgive Me!", and "A Difficult Chord" presenting the most demanding sections of the experiment.

All checkpoints in *VVVVVV* require manual activation by walking past them. Some are positioned such that they cannot be avoided, making them effectively mandatory. Teleporters function similarly to checkpoints and also cannot be bypassed, as they are required to progress between areas. Visually, an inactive checkpoint appears grey, while an activated checkpoint is brighter and flashes. Players do not receive any notification when a checkpoint is added or removed by the DDA system, though they may notice variation in difficulty across rooms.

Room numbering in this experiment follows the scheme used in the source code, where rooms are counted from 0. The tenth room encountered by the player is referred to as room 8, as "Gantry and Dolly" (room 9) is entered twice, once from above and once from below, and is counted as both the ninth and eleventh room in the sequence. The DDA system begins tracking from room 9 onward, covering rooms 9 through 63, for a total of 55 tracked rooms. Room 8 was not tracked due to an oversight in the telemetry system, though this had no impact on the data as the room contains no obstacles.

Of the 55 tracked rooms, 23 have DDA tracking disabled due to presenting no meaningful challenge, making difficulty adjustment unnecessary. These rooms are: 12, 14, 19, 22, 23, 28, 29, 30, 37, 38, 39, 43, 44, 45, 46, 48, 49, 52, 59, 60, 61, 62, and 63. The remaining 32 rooms have DDA tracking and adaptive checkpoints enabled: 9, 10, 11, 13, 15, 16, 17, 18, 20, 21, 24, 25, 26, 27, 31, 32, 33, 34, 35, 36, 40, 41, 42, 47, 50, 51, 53, 54, 55, 56, 57, and 58.

Several rooms were modified from the original game to create a linear, focused experience. Room 5, "Atmospheric Filtering Unit," had its opening to a trinket room below blocked off, and its second checkpoint was removed accordingly. Room 11, "Stop and Reflect," had its rightward exit to a trinket room sealed. Room 22 contains the teleporter marking the end of Space Station. After the rescue cutscene, the player is immediately teleported to the first room of The Laboratory rather than returning to the hub area, removing the exploratory aspect of the original game. The first room of The Laboratory, "Get Ready To Bounce" (room 23), has its entrance blocked to prevent backward navigation, and a checkpoint was added at the player's spawn point to prevent respawning to a save location in Space Station. Room 30, "Double-Slit Experiment," had its ceiling opening to a trinket room closed. Room 45, "Entanglement Generator," had its trinket room opening removed along with its teleporter, as the latter allows saving, which is incompatible with the single-session design of the experiment. Room 46, originally "The Tantalizing Trinket," was renamed "Here We Go Again" and reworked: the trinket was removed, the checkpoint was repositioned, and the room was connected directly to the drop leading into room 47, "The Bernoulli Principle," making this previously optional challenge mandatory. One trinket in the hidden room "Purest Unobtanium" remains in the experiment game but was never discovered by any participant. The final room, "Philadelphia Experiment" (room 63), ends the experiment: after the player rescues Victoria in the room before and enters the teleporter, the game exports the telemetry data and returns automatically to the main menu.

ADAPTIVE CHECKPOINT LOCATIONS

Space Station uses two difficulty ranges: stages 1-4 (easy) and stages 5-7 (hard). Rooms with adaptive checkpoints in this level are "The Yes Men" (room 10, 1 or 2 checkpoints), "Stop and Reflect" (room 11, 0 or 1), "The Sensible Room" (room 14, 0 or 1), and "Driller" (room 16, 1 or 2, with the adaptive one positioned after the drop section).

The Laboratory uses three difficulty ranges: stages 1-3 (easy), stages 4-5 (medium), and stages 6-7 (hard). Rooms with adaptive checkpoints are "Rascasse" (room 25, 0-2 checkpoints; normally 1, with an additional checkpoint available at the easiest difficulty), "Single-Slit Experiment" (room 27, 0 or 1; removed only at stages 6-7), "They Call Him Flipper" (room 31, 0-2; 2 checkpoints only at easiest difficulty), "Three's a Crowd" (room 32, 0-2; normally 1), "Brought to you by the letter G" (room 36, 0 or 1; present only at difficulty 5 or lower), "Barani, Barani" (room 41, 0 or 1; normally absent, added at difficulty 3 or lower), "Spike Strip Deployed" (room 50, 0-2; normally 2), and "A Difficult Chord" (room 57, 0 or 1; present only at difficulty 5 or lower).

APPENDIX B

ONLINE FILES

All data and analysis files associated with this study are publicly available via the following link: https://github.com/Zedero-git/MGT-DDA-VVVVVV.

This repository contains the altered source code of *VVVVVV* used for the game experiment. The "appendix" folder contains the following:

- Jamovi analysis files used for all statistical tests reported in this thesis, including graphs
- Post-hoc threshold validation spreadsheet (per room post-hoc validation.xlsx)
- General spreadsheet (MGT – Experiment Research Data.xlsx), which includes:
    - Key elements of the telemetry (blue refers to control group, pink refers to experiment group, orange refers to discarded data)
    - The questionnaire responses (either in English or Dutch based on the preferred language of the participant)
    - Player deaths per room (based on the full telemetry logs)
    - Player test data used for threshold calibration
- DDA threshold values exported separately as a plain text file (experiment dda thresholds.txt)
- A folder called "participant data logs" containing the individual telemetry logs for every participant

No personally identifiable information is included in any of these files. Participants are represented by anonymous numerical identifiers or the timestamp of when they completed the experiment.

REFERENCES

Adams, E., & Rollings, A. (2007). *Fundamentals of Game Design Game Design and Development Second Custom Edition (2007)*. Pearson Custom Publishing. http://archive.org/details/isbn_9780536045782

Byrne, E. (2005). *Game level design*. Hingham, Mass. : Charles River Media. http://archive.org/details/gameleveldesign0000byrn

Chen, J. (2006). *Flow in Games*. https://www.jenovachen.com/flowingames/Flow_in_games_final.pdf

Chen, J. (2006). Flow in games (and everything else). *Commun. ACM*, *50*(4), 31–34. https://doi.org/10.1145/1232743.1232769

Cowley, B., Charles, D., Black, M., & Hickey, R. (2008). Toward an understanding of flow in video games. *Comput. Entertain.*, *6*(2), 20:1-20:27. https://doi.org/10.1145/1371216.1371223

Csikszentmihalyi, M. (1990). *Flow: The Psychology of Optimal Experience*. Harper Perennial.

Csikszentmihalyi, M. (1996, September 1). *Go with the flow. Wired.* https://www.wired.com/1996/09/czik/

Cuerdo, M. A. M., Mahajan, A., & Melcer, E. F. (2021). Die-r Consequences: Player Experience and the Design of Failure through Respawning Mechanics. *2021 IEEE Conference on Games (CoG)*, 01–08. https://doi.org/10.1109/CoG52621.2021.9618894

Denisova, A., & Cairns, P. (2015). Adaptation in Digital Games: The Effect of Challenge Adjustment on Player Performance and Experience. *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play*, 97–101. https://doi.org/10.1145/2793107.2793141

Fandom. (n.d.). *Aku Aku*. In *Crash Bandicoot Wiki*. Retrieved October 28, 2025, from https://crashbandicoot.fandom.com/wiki/Aku_Aku

Fandom. (n.d.). *Checkpoint Crate*. In *Crash Bandicoot Wiki*. Retrieved October 28, 2025, from https://crashbandicoot.fandom.com/wiki/Checkpoint_Crate

Fandom. (n.d.). *Difficulty.* In *Ratchet & Clank Wiki.* Retrieved October 24, 2025, from https://ratchetandclank.fandom.com/wiki/Difficulty

Fandom. (n.d.). *Nanotech*. In *Ratchet & Clank Wiki*. Retrieved October 28, 2025, from https://ratchetandclank.fandom.com/wiki/Nanotech

Fandom. (n.d.). *Sonic the Hedgehog (1991)*. In *Sonic Wiki*. Retrieved October 28, 2025, from https://sonic.fandom.com/wiki/Sonic_the_Hedgehog_(1991)

Fandom. (n.d.). *The Director.* In *Left 4 Dead Wiki.* Retrieved October 24, 2025, from https://left4dead.fandom.com/wiki/The_Director

Fang, X., Zhang, J., & Chan, SusyS. (2013). Development of an Instrument for Studying Flow in Computer Game Play. *International Journal of Human-Computer Interaction*, *29*(7), 456–470. https://doi.org/10.1080/10447318.2012.715991

Fisher, N., & Kulshreshth, A. K. (2024). Exploring Dynamic Difficulty Adjustment Methods for Video Games. *Virtual Worlds*, *3*(2), 230–255. https://doi.org/10.3390/virtualworlds3020012

Fisher, N., & Kulshreshth, A. K. (2025). *Dynamic Difficulty Adjustment in Games: Concepts, Techniques, and Applications*. IntechOpen. https://doi.org/10.5772/intechopen.1011703

Fox, J., & Weisberg, S. (2023). *car: Companion to Applied Regression*. [R package]. Retrieved from https://cran.r-project.org/package=car.

Frommel, J., Klarkowski, M., & Mandryk, R. L. (2021). The Struggle is Spiel: On Failure and Success in Games. *Proceedings of the 16th International Conference on the Foundations of Digital Games*, 1–12. https://doi.org/10.1145/3472538.3472565

FromSoftware. (2011). *Dark Souls* [Video game]. Bandai Namco Entertainment.

Gavin, A. (2011, February 7). Making Crash Bandicoot – part 6. *All Things Andy Gavin*. https://all-things-andy-gavin.com/2011/02/07/making-crash-bandicoot-part-6/

Guo, Z., Thawonmas, R., & Ren, X. (2024). Rethinking dynamic difficulty adjustment for video game design. *Entertainment Computing*, *50*. https://doi.org/10.1016/j.entcom.2024.100663

Hamill, J. (2020, November 12). *There is no escape: How Hades connects game genre and Greek myth*. Nerds at Large. https://nerdsatlarge.wordpress.com/2020/11/12/there-is-no-escape-how-hades-connects-game-genre-and-greek-myth/

Hunicke, R. (2005). The case for dynamic difficulty adjustment in games. *Proceedings of the 2005 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology*, 429–433. https://doi.org/10.1145/1178477.1178573

Insomniac Games. (2002). *Ratchet & Clank* [Video game]. Sony Computer Entertainment.

Insomniac Games. (2003). *Ratchet & Clank: Going Commando* [Video game]. Sony Computer Entertainment.

Insomniac Games. (2004). *Ratchet & Clank: Up Your Arsenal* [Video game]. Sony Computer Entertainment.

Johanson, C., Piller, B., Gutwin, C., & Mandryk, R. (2023). If at First You Don't Succeed: Helping Players Make Progress in Games with Breaks and Checkpoints. *Proceedings of the ACM on Human-Computer Interaction*, *7*, 342–368. https://doi.org/10.1145/3611033

Juul, J. (2009). *Fear of Failing? The Many Meanings of Difficulty in Video Games*. 13.

Juul, J., Long, G., & Uricchio, W. (2013). *The Art of Failure: An Essay on the Pain of Playing Video Games*. MIT Press.

MarioWiki. (n.d.). *Checkpoint Flag*. Retrieved October 28, 2025, from https://www.mariowiki.com/Checkpoint_Flag

Mortazavi, F., Moradi, H., & Vahabie, A.-H. (2024). Dynamic difficulty adjustment approaches in video games: A systematic literature review. *Multimedia Tools and Applications*, *83*(35), 83227–83274. https://doi.org/10.1007/s11042-024-18768-x

Naughty Dog. (1996). *Crash Bandicoot* [Video game]. Sony Computer Entertainment.

Neuroscience of flow states. (n.d.). In *Encyclopedia.pub*. Retrieved October 28, 2025, from https://encyclopedia.pub/entry/2103

Nintendo EAD. (2006). *New Super Mario Bros.* [Video game]. Nintendo.

PracticalPie. (2023, October 31). *Flow (optimal experience in psychology)*. PracticalPie. https://practicalpie.com/flow-optimal-experience-in-psychology/

R Core Team (2024). *R: A Language and environment for statistical computing*. (Version 4.4) [Computer software]. Retrieved from https://cran.r-project.org. (R packages retrieved from CRAN snapshot 2024-08-07).

Rheinberg, F., Vollmeyer, R., Engster, S., & Sreeramoju, D. (2019). *FSS - Flow Short Scale (English Version)* [Measurement instrument]. https://www.researchgate.net/publication/373975862_FSS_-_Flow_Short_Scale_English_Version

Rockstar North. (2013). *Grand Theft Auto V* [Video game]. Rockstar Games.

Rosa, M. P. C., dos Santos, E. A., de Moraes, I. L. R., e Silva, T. B. P., Sarmet, M. M., Castanho, C. D., & Jacobi, R. P. (2021). Dynamic Difficulty Adjustment Using Performance and Affective Data in a Platform Game. In C. Stephanidis, M. M. Soares, E. Rosenzweig, A. Marcus, S. Yamamoto, H. Mori, P.-L. P. Rau, G. Meiselwitz, X. Fang, & A. Moallem (Eds.), *HCI International 2021—Late Breaking Papers: Design and User Experience* (pp. 367–386). Springer International Publishing. https://doi.org/10.1007/978-3-030-90238-4_26

Schell, J. (2008, August). *The Art of Game Design: A book of lenses*. ResearchGate. https://www.researchgate.net/publication/259532895_The_Art_of_Game_Design_A_book_of_lenses

Soderman, B. (2021). *Against flow: Video games and the flowing subject.* The MIT Press. https://research.ebsco.com/linkprocessor/plink?id=3e716154-e23b-31f8-92b3-7cdd2debf948

Sonic Team. (1991). *Sonic the Hedgehog* [Video game]. Sega.

Space Station 1. (n.d.). *VVVVVV Wiki.* https://vsix.dev/wiki/Space_Station_1

Supergiant Games. (2020). *Hades* [Video game]. Supergiant Games.

Sweetser, P., & Wyeth, P. (2005). GameFlow: A Model for Evaluating Player Enjoyment in Games. *ResearchGate*. https://doi.org/10.1145/1077246.1077253

Swink, S. (2009). *Game Feel: A Game Designer's Guide to Virtual Sensation*. https://gamifique.wordpress.com/wp-content/uploads/2011/11/2-game-feel.pdf

The jamovi project (2024). *jamovi*. (Version 2.6) [Computer Software]. Retrieved from https://www.jamovi.org.

The Laboratory. (n.d.). *VVVVVV Wiki.* https://vsix.dev/wiki/The_Laboratory

Valve Corporation. (2008a). *Left 4 Dead* [Video game]. Valve Corporation.

Valve Corporation. (2008b). *Left 4 Dead* [Video game]. Valve Corporation. https://store.steampowered.com/app/500/Left_4_Dead/

Valve Corporation. (2009). *Left 4 Dead 2* [Video game]. Valve Corporation.

van den Hoogen, W. M., Poels, K., IJsselsteijn, W. A., & De Kort, Y. (2012). Between Challenge and Defeat: Repeated Player-Death and Game Enjoyment. *Media Psychology*, 15(4), 443–459. https://doi.org/10.1080/15213269.2012.723117

VVVVVV. (n.d.). *VVVVVV Wiki.* https://vsix.dev/wiki/VVVVVV

Westhoff, N. (n.d.). *Literature basics*. IRP — Nils Westhoff. Retrieved October 28, 2025, from https://irp.nilswesthoff.com/chapters/literature-basics

Wikipedia. (n.d.). *Bonfire (Dark Souls)*. In *Wikipedia*. Retrieved October 28, 2025, from https://en.wikipedia.org/wiki/Bonfire_(Dark_Souls)

Xue, S., Wu, M., Kolen, J., Aghdaie, N., & Zaman, K. A. (n.d.). Dynamic difficulty adjustment for maximized engagement in digital games. *Proceedings of the 26th International Conference on World Wide Web Companion.* ACM. https://doi.org/10.1145/3041021.3054170

Yacht Club Games. (2014a). *Shovel Knight* [Video game]. Yacht Club Games.

Yacht Club Games. (2014b, June 19). *Check Point Design*. Yacht Club Games. https://www.yachtclubgames.com/blog/check-point-design/

Zohaib, M. (2018). Dynamic Difficulty Adjustment (DDA) in Computer Games: A Review. *Advances in Human-Computer Interaction*, *2018*(1), 5681652. https://doi.org/10.1155/2018/5681652