Towards Generating Surprising Content in 2D Platform Games

Chandranil Chakraborttii
nil.chakraborttii@trincoll.edu
Trinity College
Hartford, CT, USA

Lucas Ferreira
lucas.n.ferreira@ufv.br
Universidade Federal de Viçosa
Viçosa, MG, Brazil

ABSTRACT

Surprise is a key factor for driving engagement in video games. By incorporating unexpected elements, games can create a sense of excitement and curiosity, leading to a more immersive and enjoyable experience. In this paper, we propose a framework for generating levels in 2D platform games with embedded surprising content. Building upon the VCL (Violation of Expectation, Caught Off Guard, and Learning) model, we use it as a conceptual foundation to curate surprise by altering specific game elements (called metrics) during the level generation process. We developed a tile-based parametric level generator for Super Mario Bros., creating customized levels based on metrics such as linearity, enemy density, and pattern variation, among others. 393 participants in our study played 2 generated levels, with the first setting expectations and the second intentionally violating them by altering chosen metrics. By comparing player responses and gameplay data with the VCL model, we explore the phenomenon of surprise in games. Our findings reveal statistically significant correlations between certain metrics and player responses, teasing at the potential for automatically generating surprising levels in 2D platform games.

ACM Reference Format:
Chandranil Chakraborttii and Lucas Ferreira. 2024. Towards Generating Surprising Content in 2D Platform Games. In *Proceedings of the 19th International Conference on the Foundations of Digital Games (FDG 2024), May 21–24, 2024, Worcester, MA, USA.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3649921.3659848

1 INTRODUCTION

Games are a significant form of interactive entertainment, with 65% of Americans engaging across platforms like PC, console, and mobile [32]. Creating game content requires substantial manual effort, prompting interest in automated content generation. However, procedurally generated content often lacks diversity, leading to player disengagement as seen in games like No Man's Sky [9]. We assert that the key issue lies in the failure to establish emotional connections with players. Incorporating surprising elements in games enhances player engagement and demand for new content. Surprise enhances emotions, learning, and adaptation [21]. By infusing surprise into generated content, developers can foster deeper player connections and sustain engagement.

FDG 2024, May 21–24, 2024, Worcester, MA, USA

ACM ISBN 979-8-4007-0955-5/24/05
https://doi.org/10.1145/3649921.3659848

The VCL model [3] proposes that surprise in games is unique due to their interactive nature and player agency. It defines surprise as a sudden moment eliciting verbal, nonverbal, or physiological responses triggered by the violation of player expectations. According to the VCL model, surprise arises when an event challenges deeply ingrained beliefs, influenced by real-life and gaming experiences. The intensity of surprise is determined by the degree of belief change, quantified by the divergent factor (DF) - the difference between prior and post-event beliefs. However, the model faces challenges in quantifying qualitative information about DF, which limits its ability to predict player surprise. DF is influenced by factors such as past player experiences and personality traits. Improving our understanding of DF could enhance predictive capabilities regarding individual player states in games, facilitating the creation of personalized surprising content tailored to each player.

In this work, we address this limitation of VCL model by gaining insights about the DF of an event. More specifically, we empirically analyze what degree of change in beliefs (due to a change in level geometry) is needed to evoke a surprising reaction from a player. Additionally, we are interested in exploring why surprise is perceived differently among players. For example, why one game's level geometry might appear surprising to a player, but another player finds the same level normal. We developed a tile-based level generator of the popular game, Super Mario Bros, called Science Mario, as test bed for our experiments. Developing a new level generator tool for our research, allows us to alter chosen metrics (linearity, density of enemies, density of game elements, pattern variation, density of gaps, player speed, and gravity [16]) that affect level geometry or mechanics to suit our experimental needs. While existing level generators exist for games, they lacked the specific functionalities and capabilities required for our experiments. Our developed level generator accepts input in the form of an alphanumeric matrix, where each value corresponds to a particular element in the game (For example: pipe, goomba, etc.). As Super Mario Bros. is a popular game, players have often already played it, and are aware of the game's geometry and mechanics, thereby developing strong beliefs about the game. We hypothesize that a sudden change in the chosen level metrics affecting the game's geometry or mechanics will violate the prior beliefs, thereby inducing a surprising reaction from the player.

We generated a total of 49 levels, with 7 levels generated for each chosen metric. The first level, referred to as the base level, maintained a fixed value for the metric to emulate the player experience of the original Super Mario Bros. The subsequent 6 levels were generated with varying values of the metric, ranging from 0 to 1. Detailed information about our level generator, Science Mario, and the metrics used are provided in a later section (Section 4). For our study, participants were asked to play 2 generated levels of the game and complete an online form recording their experience

playing the game. Each player was randomly assigned a metric and played the base level first, followed by another randomly generated level with the altered metric. We ensured each generated level was playable, meaning it could be completed by a player, and balanced the study to have 8 participants per level combination. We analyzed player surprise perception during gameplay changes using our developed generator. Metrics were controlled using parameters that accepted values between 0 and 1. For e.g., in the 'density of enemies' metric, 0 represented no enemies in the generated level, while 1 represented an enemy at every possible platform tile in the level. We make the following contributions in this work:
(1) Exploration of VCL model of surprise, gathering insights about divergent factor of an event. In other words, what makes an event surprising to a player? (2) Framework for generating levels with embedded surprising content in 2D platforms. (3) Tile-based Super Mario Bros. level generator (Science Mario).

2 RELATED WORK

Surprise, considered one of the six basic human emotions, is universal and typically correlates positively with engagement [7]. Given the exponential growth of video games as a form of entertainment, understanding surprise in this context is crucial. Researchers have extensively analyzed the impact of video games on human emotions [4, 6, 22, 26], recognizing surprise as pivotal for enhancing player engagement, immersion, and enjoyment. Additionally, surprise boosts emotional responses and increases replayability in games [4, 19, 20]. Our review of prior research highlights a gap in understanding how individual players perceive surprising events, hindering the development of controlled and engaging levels, especially in the realm of 2D platformers.

Ely et al.[8] described surprise as a way of revealing information over a period of time, defining surprise as the deviation between current beliefs and previous expectations over time. A concept explored across various domains including casinos, politics, literature, and games. This concept was formalized into the Expectancy Violations Theory (EVT)[2], explaining how unexpected behaviors in social settings evoke surprise. EVT originated in the context of Proxemics, the study of human spatial requirements and population density's effects on behavior and communication [14]. Gross [13] proposed a related theory linking ignorance and surprise, emphasizing the benefits of incorporating surprise and ignorance into design processes. This theory suggests that awareness of ignorance startles individuals, highlighting the gap between sensory perceptions and environmental awareness [30]. The VCL [3] model of surprise integrates these theories, shaping a comprehensive understanding of surprise in games. Another model of surprise, surprise search [11, 36] uses an evolutionary algorithm that values unexpected behaviors over unseen ones. It uses a prediction model to anticipate outcomes and rewards deviations from expectations. This mirrors a self-surprise process, favoring individuals diverging from evolutionary trends to shape new ones. The algorithm outperforms other methods in deceptive problems and maze navigation tasks, but has limited applications in the context of 2D platform games. It consists of two modules: a predictive model based on past behaviors and a distance formula measuring deviation from expected outcomes. Similar to VCL model, the surprise search prediction model generates a speculative 'current' population based on previous generations, considering local or global behavioral information.

Generative methods [5] are functions that produce new artifacts by manipulating user-provided tuning values, static assets (like 3D models or audio samples), or higher-level inputs. These inputs are processed by a generator, a central component that combines and composites them to create a set of artifacts. Traditionally, the "generate and test" approach has been prevalent in computational creativity research. On an individual level, creators generate ideas and test them against domain knowledge until they find one that meets their personal criteria [34]. Hooshyar et al.[15] demonstrated that quantitative models of player experience can accurately predict player affective states using gameplay metrics and level parameters. They applied this to a procedurally generated version of Infinite Mario Bros[33]. Yannakakis [37] proposed camera control as crucial for adaptive interaction in games, correlating camera position with player responses and affective states using biometric inputs. Smith, Whitehead, et al.[29] analyzed platformer level components and structures to understand level design and challenge areas. Player experience modeling[25] employs predictive models to anticipate player behavior and preferences, often using machine learning algorithms [31]. Smith et al.[28] developed a taxonomy of player modeling techniques, while Yannakakis et al.[35, 38] discussed various approaches, including model-based and model-free methods. Quality-diversity algorithms [10] prioritize both quality and diversity in procedural content generation, diverging from standard evolutionary algorithms. Surprise-driven content creation [12] enhances player engagement by introducing surprise-based evolutionary search methods, as demonstrated in Unreal Tournament III for weapon generation. In contrast, our research delves into surprise as an emotion, specifically examining its correlation with level geometry and mechanics in 2D platform games. While existing studies shed light on surprise as an emotion experienced in media and its detection from various sensory inputs, many questions regarding surprise in video games remain unanswered. Further discussion and research are needed in this area.

3 VCL MODEL OF SURPRISE

Chakraborttii et al. [3] introduced the VCL framework for modeling surprise in games, building on two established theories: expectancy violations (EVT) [1] and the theory of ignorance and surprise [13]. In the context of video games, EVT explains how players experience surprise when their existing beliefs about the game world (their expectations) clash with what they actually encounter, resulting in a discrepancy. The second theory posits that surprise occurs when a previously unnoticed event suddenly enters the player's focus, often due to the limitations of human sensory and perceptual abilities. The VCL model integrates these two theories, proposing that a surprising event in video games occurs when one of the following conditions is met:

(Either) The player's expectations are suddenly violated.
(Or) The player is caught off guard.
(And) The event triggers learning.

Figure 1: A sample generated level in Science Mario

The occurrence of surprising events in games depends on the player's beliefs and prior experiences. When starting a new game, players draw on their existing knowledge and beliefs, acquired through past gaming experiences, to inform their actions. This reliance on prior beliefs is a key aspect of the VCL model. According to the VCL model, an event is considered surprising if it deviates significantly from the player's initial beliefs. The Divergent Factor (d) measures this deviation, representing the difference between the player's beliefs before and after an event (E). If d exceeds a player-specific threshold (Δ), the event is perceived as surprising. The significance of the event, quantified by the Divergent Factor ($d = DF(P, E)$), can vary between players (P1 and P2) due to individual differences in experience, adaptability, and personality [27]. Currently, the VCL model is theoretical, and defining how to calculate DF and Δ is a challenge. To address this, we conduct experiments with players to gather insights and inform the development of the VCL model.

4 TEST BED: SCIENCE MARIO

We created a level generator, Science Mario, inspired by Super Mario Bros, to generate levels for our user study. We chose Super Mario Bros as our testbed due to its established reputation in the research community, where it has been extensively studied [18, 24]. The availability of level generators [18] and simulation agents makes it an ideal choice. Moreover, the game's rich history ensures that most participants will have prior experience and expectations, providing a valuable foundation for our experiments. Figure 1 shows an example of a generated level. We built a parametric, tile-based level generator [16] utilizing a genetic algorithm (GA) [23], a class of evolutionary computation techniques that mimics real life evolution used in prior work to generate the levels. The approach is based on natural selection and is commonly used in search problems with exponential growth that leads to the impossibility of testing all potential solutions. This approach is flexible and can potentially generate a large number of diverse levels. The reason for building a new level generator was to generate parameterized custom levels based on the metrics chosen for the purpose of the user study.

The generator accepts an alphanumeric matrix as input where each matrix value corresponds to a particular element in the game (For example: pipe, goomba, etc). The GA begins with an initial population of alphanumeric matrices (levels) with chosen values to simulate a regular level. The matrix had a fixed number of rows of 30 (fixed width) and a variable number of columns between 120-300 (representing variable length). A fitness function is defined to evaluate each level based on desired metrics such as linearity, connectivity, and player experience with given constraints, assigning a score based on time to complete each level (by a bot) with lower scores indicating better levels. The fittest levels are selected to reproduce and form the next generation, simulating natural selection. Crossover (recombination) is performed between two selected levels to create a new level, and random mutations are applied to introduce diversity and escape local optima. The least fit levels are replaced with the new levels generated, and the process is repeated for a predetermined number of generations or until a satisfactory level is generated. This approach can generate diverse levels that satisfy the desired metrics, and the parameters of the generator (metrics) control certain characteristics of level geometry or mechanics, allowing the GA to search for optimal combinations to generate high-quality levels. The values of the metrics are normalized to accept values on a scale between 0-1. We chose the following metrics because they were used previously to evaluate procedural level generators, as in the Mario AI Framework [16]. The metric values were chosen in the range between 0.12 - 0.88 skipping the extreme cases with values of 0 and 1. The threshold difference was computed by dividing the number of levels + 1 by 1.0 and then using multiples of the threshold difference. We manually played through each of the generated levels to ensure the levels can be completed.

A. Linearity: This metric, derived from the R2 goodness-of-fit measure, evaluates the linearity of a level by fitting a line to the endpoints of each platform [16]. High linearity indicates levels where the player navigates primarily along a straight path towards the goal, while low linearity suggests more varied terrain, requiring the player to traverse curved paths between platforms. Altering the linearity impacts the level's structure and flow, shaping the player's perception of progression and exploration.

B. Density of enemies: Similar to linearity, as introduced by Horn et al. [16], this metric gauges a level's difficulty for players. It is computed by dividing the total number of enemies in the level by the level width. Modifying the density of enemies directly influences the challenge and intensity of the level, impacting player engagement and strategic choices.

C. Density of game elements: This metric quantifies the density of game elements in a level, with a higher value indicating more game elements present. Adjusting the density of various game elements, including power-ups and obstacles, can diversify gameplay experiences and affect how players interact with the environment.

D. Pattern variation: Pattern variation measures the diversity of game elements in a level, encompassing enemies and game elements. A value of 1 guarantees all game elements occur, while 0 signifies no variation. Incorporating varied patterns in levels ensures dynamic and unpredictable gameplay, sustaining player engagement and preventing monotony.

E. Gap density: This metric computes the gap frequency by dividing the number of gaps by the level width. A gap intensity of 1 signifies the maximum number of gaps necessary for completing the level. Higher values indicate more gaps or tiles within the level.

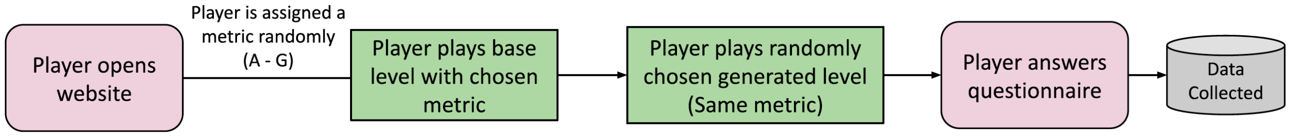

Figure 2: Methodology

Adjusting the density of gaps or obstacles influences the platforming dynamics, prompting players to navigate and jump strategically, thereby enhancing skill and tension during gameplay.

F. Player Speed: This metric reflects the player's movement speed, with lower values indicating slower movement. We set upper and lower threshold limits at $10x$ and $0.1x$ of the regular playing speed (normalized as 1 and 0 respectively). Player speed influences gameplay pace and intensity, affecting reaction time and decision-making for players.

G. Gravity: This metric signifies the gravity experienced by the player. Lower gravity values lead to higher jumps and longer jump times and distances. Like Player Speed, we set upper and lower threshold limits at $10x$ and $0.1x$ (normalized as 1 and 0). Adjusting gravity mechanics can substantially affect player movement and control dynamics, thereby shaping the overall feel and gameplay responsiveness.

5 USER STUDY

We recruited 393 participants (198 male, 183 female, and 14 others) for the user study through Amazon Mechanical Turk. Participants ranged in age from 18-65 and were compensated $2.02 on average for approximately 15 minutes of time. The interaction context was in English and the participants were recruited from English-speaking countries only (United Kingdom, Australia, Canada, USA, New Zealand). The study was held online and hosted on an Amazon EC2 server. As mentioned earlier, we generated a total of 49 game levels for our experiments, 7 levels for each chosen metric. Each metric was given an alphabet between A-G. The base level had a mean value (0.5 in a normalized scale between 0-1) and simulated a normal level from the original Super Mario Bros. game. There were 3 levels (L 1-3) generated with greater value than the mean for the metric. Similarly, the last 3 generated levels (L 5-7) for the metric, had a lower value than the mean with L4 representing the base level. Each player was randomly assigned a metric (say gap density E), and they first played the base level for the metric (E4), followed by another generated level (say E6) for that metric (gap density) assigned randomly. In this case, the gameplay sequence of Base Level (E4), followed by E6 level (E4-E6) for gap density represented a unique combination of levels. Similarly, E4-E1 represented another combination of levels.

To accurately capture the impact of change for a particular metric, all other metrics are kept fixed. For example, if a player is assigned a metric (say gap density), only the gaps in the level will change, while all other metrics (such as linearity, pattern variation, etc.) are fixed for both levels. We balanced the user study and ensured that there were 8 study participants per combination of levels for each metric. After the game, the participants completed the online form where they reported the intensity of perceived surprise for each of the levels played using a Likert scale [17] - where a surprise value of 1 referred to no surprise being experienced by participants and value of 7 – to the highest level of surprise. We also collected data regarding demography, prior gameplay experience, and genre experience (experience of playing similar 2D platform games). The end-to-end process is described in Figure 2. To test our hypotheses introduced earlier, we measure the difference in the absolute value of the chosen metric between the levels simulating the divergent factor between the levels (recorded as DF value corresponding to each level combination). The base level was meant to set the expectations of the game (representing a typical game) and the second level was meant to violate the expectations by changing the metrics and surprise the players. The goal of the user study is to explore the extent of change required in a level to curate surprise, i.e, divergent factor (DF) of an event. We are also interested in finding out how DF varies between players by relating to their past gameplay experience allowing us to consistently and precisely generate embedded surprising content in future games.

6 RESULTS

This section describes the data collected from the user study and our analysis to gain more insights into the applicability of the VCL model. Figure 3 shows how the intensity of perceived surprise varied for each metric using a box chart. The x-axis represents the average intensity of surprise reported by the participants for that particular combination of level and the y-axis shows the chosen metric (A-G). From the figure, it's evident that altering the level led to varying levels of perceived surprise intensity for different metrics among players, highlighting the individualized perception of surprise. Additionally, our analysis revealed that the intensity of surprise varied significantly between metrics, with some metrics eliciting higher surprise values. This suggests that certain game elements, such as level geometry or enemy density, have a greater impact on player surprise than others. Furthermore, the unique combinations of levels played by each player resulted in a range of surprise values, indicating that player experience and expectation play a significant role in perceiving surprise.

Figures 5-7 show how the intensity of reported surprise varies with changing values of a chosen metric. The x-axis corresponds to the ID of the generated levels and the y-axis corresponds to the average value of surprise reported. For each metric, the fourth level represented the base level, with levels 1-3 represented a decrease in the chosen metric, and levels 5-7 represented an increase in the chosen metric. Each figure displays the absolute change in the metric value between the two played levels (blue line), represented as the DF compared to the base level, along with the average difference in reported surprise (orange line). As the base level(Level 4), was designed to emulate a typical Super Mario Bros. level, it was characterized by predictability and a lack of surprises. This familiarity led to the lowest reported surprise values across all metrics, as players experienced a sense of normalcy and expectation confirmation.

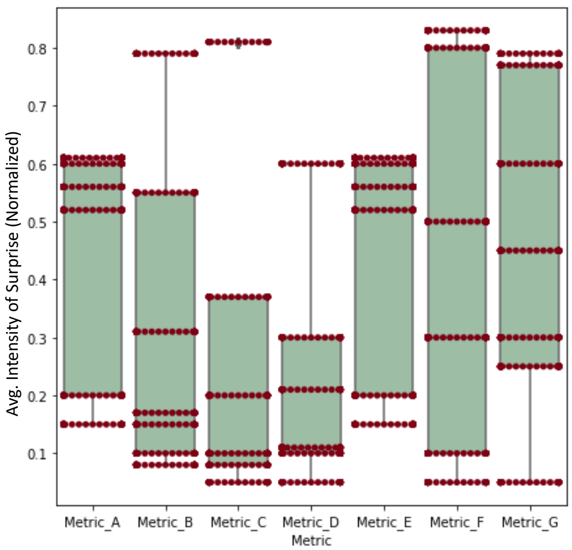

Figure 3: Intensity of surprise variation for each metric

Observation 1: Figure 4 shows a common pattern on how the average value of reported surprise increases with an increase in DF (absolute change) for 3 metrics - linearity, density of enemies, and density of game elements. As DF increases due to a change in the linearity of the level, or density of game elements/enemies, people get increasingly more surprised. However, this relationship is non-linear and varies based on the chosen metric. Peaks in surprise occur at specific levels of change, indicating sharp alterations evoke heightened reactions. Players feel more surprised if the level changes (DF) sharply from a certain level (base level in this case), showing peaks at A1 and A7 (for linearity), B1 and B7 (density of enemies), C1 and C7 (density of game elements). This is because the change in DF is highest for levels 1 and 7 (with base level 4) resulting in a highly altered level.

We also notice that lower linearity in the level surprised more players than the lower density of game elements and enemies indicating a change that results in an easier level for the players is less surprising than difficult levels. Players reported more surprise when they see a high number of different game elements or enemies, compared to the base level. This shows that a player has a belief about a level that will have some reasonable variation in the game elements. When this expectation gets violated, players get surprised. This happens when a player encounters a higher change in the appearance of these game elements. Upon running one way ANOVA test for the above-mentioned metrics, with Level ID as the independent variable and average surprise value as the dependent variable, we found there is a statistically significant relationship [$F(1, 764) = 4.18$, $p < 0.05$] between group combinations for linearity (i.e, between Base-A1, Base-A2, Base-A3 and so on) and density of enemies [$F(1, 764) = 5.27$, $p < 0.05$] but not for density of game elements [$F(4, 35) = 1.32$, $p > 0.05$]. Thus, we can argue that there is a statistical significance in the intensity of surprise between the levels for the metrics - linearity, and density of enemies. The findings support the VCL model's emphasis on expectation violation, as players reported more surprise when encountering a high number of different game elements or enemies, violating their belief about reasonable variation in the game elements. The statistical significance in the intensity of surprise between levels for linearity and density of enemies further reinforces the VCL model's applicability in understanding player surprise in video games.

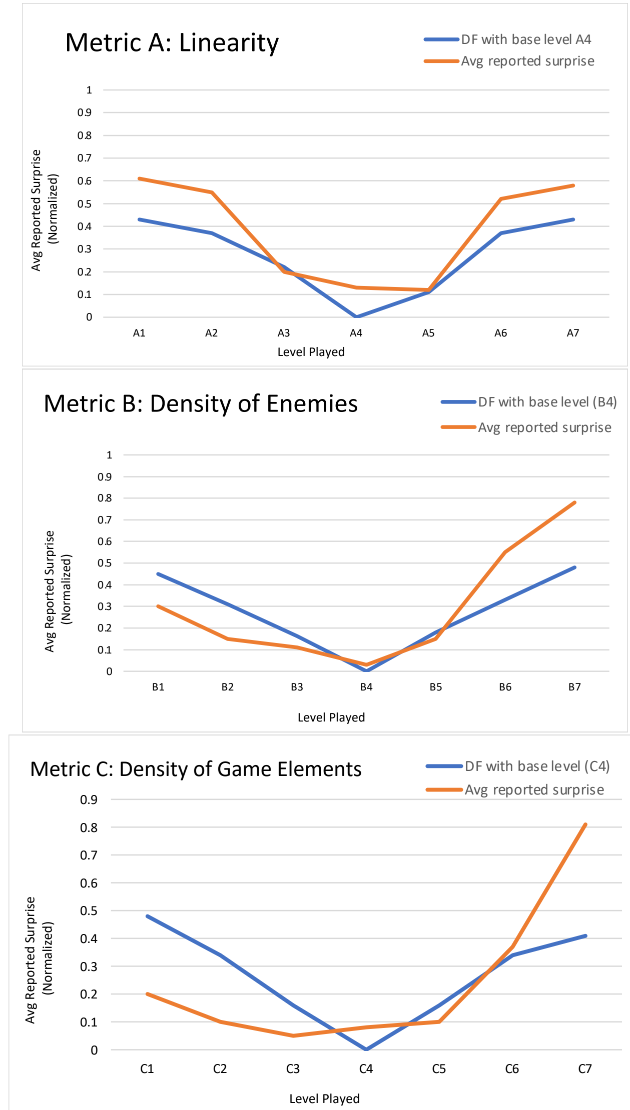

Figure 4: Average reported intensity of surprise variation with level linearity (top left), density of game elements (top right) and density of enemies (bottom)

Observation 2: For metrics pattern variation and density of gaps (Figure 5), we see a trend that shows players got more surprised as DF is increased, but there are exceptions to this in DF value for levels

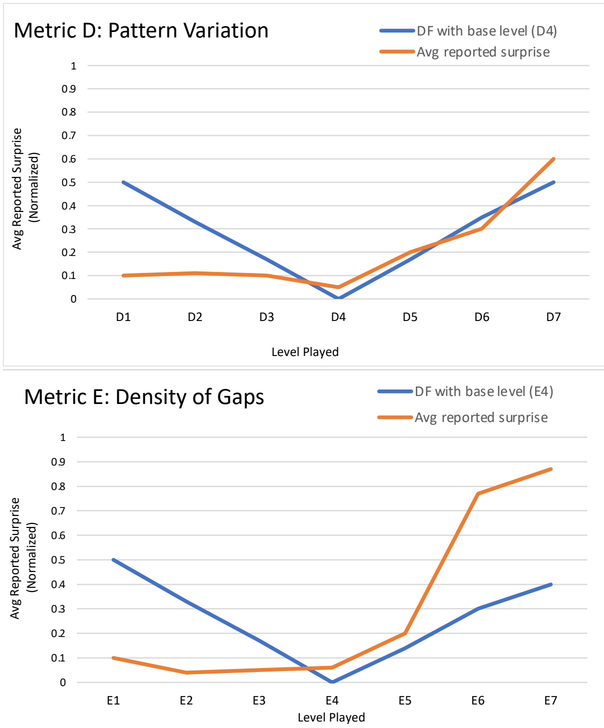

Figure 5: Comparing avg reported surprise intensity with (left) pattern variation & (right) density of gaps

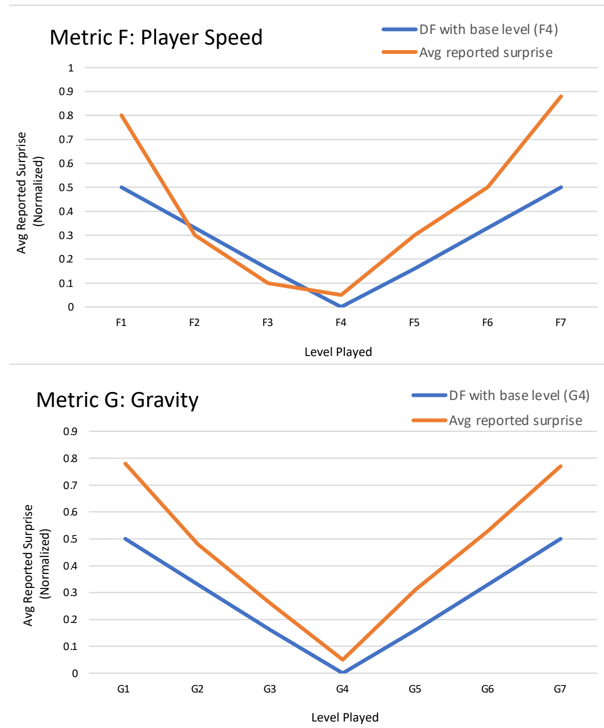

Figure 6: Comparing avg reported surprise intensity with (left) player-speed & (right) gravity

D1-D3 and C2-C4 respectively for these metrics. Thus, we can infer only a weak relationship between these metrics and DF. Players reported more intensity of surprise when the pattern variation and density of gaps is much higher than the base level (D7/ E7). This means that if the game elements are all the same, players get less surprised, but are more surprised when there exists a huge number of different game elements in a level. For density of gaps metric, we see that there is a high increase in the average value of surprise suddenly (after E5). So, as the gap frequency (DF) increases after a certain threshold, people tend to feel more surprised. Super Mario games generally have a lower number of gaps in the initial levels, violating which might be the reason why people got surprised more. Moreover, the finding that players reported more intensity of surprise when the pattern variation and density of gaps were much higher than the base level (D7/E7) reinforces the VCL model's proposal that surprise occurs when players are caught off guard. The sudden increase in surprise after E5 for the density of gaps metric also supports the VCL model's learning trigger condition, as players adapt to the changing game elements and experience surprise when their expectations are violated.

Observation 3: Figure 6 shows how a change in the physics metrics of the level affects surprise. Compared to the other metrics, we see a higher reported surprise with any change to the base level (representing normal physics). People reported being most surprised when DF is highest (G1-G7, F1-F7) with a consistent increase in surprise with increasing change. This shows players have a strong belief about the physics of the game which affects the timing of jumps, movement, the strategy of level completion, etc. Any change requires a strong change in prior beliefs resulting in a surprising reaction. Upon running one way ANOVA test for the above-mentioned metrics, with Level ID as the independent variable and average surprise value as the dependent variable, we found there is a statistically significant relationship between group combinations for both the metrics [$F(1, 764) = 8.54$, $p < 0.05$ for Player Speed, and $F(1, 764) = 13.16$, $p < 0.05$ for Gravity]. Thus, we can argue that there is a statistical significance in the intensity of surprise between the levels. From Figures 5-7, we can conclude that a little change to the levels geometry/physics is not perceived as surprising to players. However, significant changes in chosen metrics elicit a stronger reaction. We see that only some metrics reflect statistical correlation with chosen metrics, but variations are clear for all chosen metrics. Notably, level physics metrics, such as gravity and player speed, evoke a more pronounced surprise response compared to geometry features and are statistically significant ($p < 0.05$). These findings support the hypothesis that significant changes in chosen metrics lead to heightened surprise reactions among players, reinforcing the principles outlined in the VCL model. Based on our findings, we show that the hypotheses we proposed are valid for the chosen metrics (4 out of 7 chosen metrics reflect statistical correlation).

Observation 4: We found significant variability in the intensity of surprise among players, with prior gameplay experience and genre familiarity influencing surprise perception greatly. Players reported an average surprise intensity of 2.29, with a wide range of

1-7, indicating diverse reactions to surprising events. Prior video game experience was negatively correlated with intensity of surprise, [$r(764) = -2.61$, $p < 0.05$] suggesting that familiarity with game mechanics reduces surprise. Moreover, genre experience exhibits an even stronger negative correlation [$r(764) = -3.43$, $p < 0.05$] with surprise perception, particularly in the context of a 2D platform game, indicating that unfamiliarity with specific game genres heightens surprise reactions. These results align with the VCL model, highlighting the crucial role of player expectations and knowledge in shaping surprise experiences through expectancy violations and attention shifts.

7 CONCLUSION AND FUTURE WORK

In summary, we developed a tile-based level generator for Super Mario Bros. that creates custom levels based on various metrics. A user study was conducted where 393 participants played these generated levels, and their responses and gameplay data were correlated with the VCL model to understand surprise in games. Our hypothesis, that changing level metrics would elicit stronger surprise reactions, was tested and statistically significant correlations were found between level geometry metrics and player emotional responses to surprising events. In the future, we aim to use these insights to consistently generate controllable and customized surprising levels by applying changes based on chosen metrics. We also plan to consider players' past experiences and gameplay styles in level generation and explore how combinations of metric changes affect player surprise.

REFERENCES

[1] Judee K Burgoon. 2015. Expectancy violations theory. *The international encyclopedia of interpersonal communication* (2015), 1–9.
[2] Judee K Burgoon and Stephen B Jones. 1976. Toward a theory of personal space expectations and their violations. *Human Communication Research* 2, 2 (1976), 131–146.
[3] Chandranil Chakraborttii, Lucas N Ferreira, and Jim Whitehead. 2017. Towards generative emotions in games based on cognitive modeling, In Proceedings of the 12th International Conference on the Foundations of Digital Games. *Available at SSRN 4119230*, 1–2.
[4] Chandranil Chakraborttii and Anya Osborne. 2023. Characterizing player responses to surprising events in 2D platform games. *Entertainment Computing* 45 (2023), 100542.
[5] Kate Compton, Joseph C Osborn, and Michael Mateas. 2013. Generative methods. In *The Fourth Procedural Content Generation in Games workshop, PCG*, Vol. 1.
[6] Mihaly Csikszentmihalyi. 1997. Flow and the psychology of discovery and invention. *HarperPerennial, New York* 39 (1997).
[7] Paul Ekman. 1992. An argument for basic emotions. *Cognition & emotion* 6, 3-4 (1992), 169–200.
[8] Jeffrey Ely, Alexander Frankel, and Emir Kamenica. 2015. Suspense and surprise. *Journal of Political Economy* 123, 1 (2015), 215–260.
[9] Hello Games. 2016. No man's sky. *San Mateo: Sony Interactive Entertainment* (2016).
[10] Daniele Gravina, Ahmed Khalifa, Antonios Liapis, Julian Togelius, and Georgios N Yannakakis. 2019. Procedural content generation through quality diversity. In *2019 IEEE Conference on Games (CoG)*. IEEE, 1–8.
[11] Daniele Gravina, Antonios Liapis, and Georgios Yannakakis. 2016. Surprise search: Beyond objectives and novelty. In *Proceedings of the Genetic and Evolutionary Computation Conference 2016*. 677–684.
[12] Daniele Gravina, Antonios Liapis, and Georgios N Yannakakis. 2016. Constrained surprise search for content generation. In *2016 IEEE Conference on Computational Intelligence and Games (CIG)*. IEEE, 1–8.
[13] Matthias Gross. 2010. *Ignorance and surprise: Science, society, and ecological design.* MIT Press.
[14] Edward Twitchell Hall. 1962. *Proxemics: The study of man's spatial relations.*
[15] Danial Hooshyar, Moslem Yousefi, and Heuiseok Lim. 2018. Data-driven approaches to game player modeling: a systematic literature review. *ACM Computing Surveys (CSUR)* 50, 6 (2018), 1–19.
[16] Britton Horn, Steve Dahlskog, Noor Shaker, Gillian Smith, and Julian Togelius. 2014. A comparative evaluation of procedural level generators in the mario ai framework. (2014).
[17] Ankur Joshi, Saket Kale, Satish Chandel, and D Kumar Pal. 2015. Likert scale: Explored and explained. *British journal of applied science & technology* 7, 4 (2015), 396.
[18] Manuel Kerssemakers, Jeppe Tuxen, Julian Togelius, and Georgios N Yannakakis. 2012. A procedural procedural level generator generator. In *Computational Intelligence and Games (CIG), 2012 IEEE Conference on.* IEEE, 335–341.
[19] Daniel King, Paul Delfabbro, and Mark Griffiths. 2010. Video game structural characteristics: A new psychological taxonomy. *International journal of mental health and addiction* 8 (2010), 90–106.
[20] Nicole Lazzaro. 2009. Why we play: affect and the fun of games. *Human-computer interaction: Designing for diverse users and domains* 155 (2009), 679–700.
[21] Tania Luna and LeeAnn Renninger. 2015. *Surprise: Embrace the Unpredictable and Engineer the Unexpected.* TarcherPerigee.
[22] Johnathan Pagnutti Mobramaein, Husacar Sanchez, Jim Whitehead, and Brenda Laurel. 2016. Design Lessons From Binary Fission: A Crowd Sourced Game for Precondition Discovery. (2016).
[23] Fausto Mourato, Manuel Próspero dos Santos, and Fernando Birra. 2011. Automatic level generation for platform videogames using genetic algorithms. In *Proceedings of the 8th International Conference on Advances in Computer Entertainment Technology*. 1–8.
[24] Juan Ortega, Noor Shaker, Julian Togelius, and Georgios N Yannakakis. 2013. Imitating human playing styles in super mario bros. *Entertainment Computing* 4, 2 (2013), 93–104.
[25] Christopher Pedersen, Julian Togelius, and Georgios N Yannakakis. 2010. Modeling player experience for content creation. *IEEE Transactions on Computational Intelligence and AI in Games* 2, 1 (2010), 54–67.
[26] Niklas Ravaja, Timo Saari, Mikko Salminen, Jari Laarni, and Kari Kallinen. 2006. Phasic emotional reactions to video game events: A psychophysiological investigation. *Media psychology* 8, 4 (2006), 343–367.
[27] Rainer Reisenzein, Sandra Bördgen, Thomas Holtbernd, and Denise Matz. 2006. Evidence for strong dissociation between emotion and facial displays: The case of surprise. *Journal of personality and social psychology* 91, 2 (2006), 295.
[28] Adam M Smith, Chris Lewis, Kenneth Hullett, Gillian Smith, and Anne Sullivan. 2011. An inclusive taxonomy of player modeling. *University of California, Santa Cruz, Tech. Rep. UCSC-SOE-11-13* (2011).
[29] Gillian Smith, Mee Cha, and Jim Whitehead. 2008. A framework for analysis of 2D platformer levels. In *Proceedings of the 2008 ACM SIGGRAPH symposium on Video games*. 75–80.
[30] Darren Stanley. 2009. Complexity and the Phenomenological Structure of'Surprise'. *Emergence: Complexity and Organization* 11, 2 (2009), 46.
[31] Adam Summerville, Sam Snodgrass, Matthew Guzdial, Christoffer Holmgård, Amy K Hoover, Aaron Isaksen, Andy Nealen, and Julian Togelius. 2018. Procedural content generation via machine learning (PCGML). *IEEE Transactions on Games* 10, 3 (2018), 257–270.
[32] Sigmund Ed Tobias and John D Fletcher. 2011. *Computer games and instruction.* IAP Information Age Publishing.
[33] Julian Togelius, Sergey Karakovskiy, and Robin Baumgarten. 2010. The 2009 mario ai competition. In *IEEE Congress on Evolutionary Computation*. IEEE, 1–8.
[34] Julian Togelius, Georgios N Yannakakis, Kenneth O Stanley, and Cameron Browne. 2011. Search-based procedural content generation: A taxonomy and survey. *IEEE Transactions on Computational Intelligence and AI in Games* 3, 3 (2011), 172–186.
[35] Georgios N Yannakakis and John Hallam. 2008. Entertainment modeling through physiology in physical play. *International Journal of Human-Computer Studies* 66, 10 (2008), 741–755.
[36] Georgios N Yannakakis and Antonios Liapis. 2016. Searching for Surprise. In *Proceedings of the Seventh International Conference on Computational Creativity.*
[37] Georgios N Yannakakis, Héctor P Martínez, and Arnav Jhala. 2010. Towards affective camera control in games. *User Modeling and User-Adapted Interaction* 20, 4 (2010), 313–340.
[38] Georgios N Yannakakis, Pieter Spronck, Daniele Loiacono, and Elisabeth André. 2013. Player modeling. (2013).