Level Design Patterns That Invoke Curiosity-Driven Exploration: An Empirical Study Across Multiple Conditions

MARCELLO A. GÓMEZ-MAUREIRA, Leiden University, The Netherlands
ISABELLE KNIESTEDT, Delft University of Technology, The Netherlands
MAX VAN DUIJN, Leiden University, The Netherlands
CAROLIEN RIEFFE, Leiden University, The Netherlands
ASKE PLAAT, Leiden University, The Netherlands

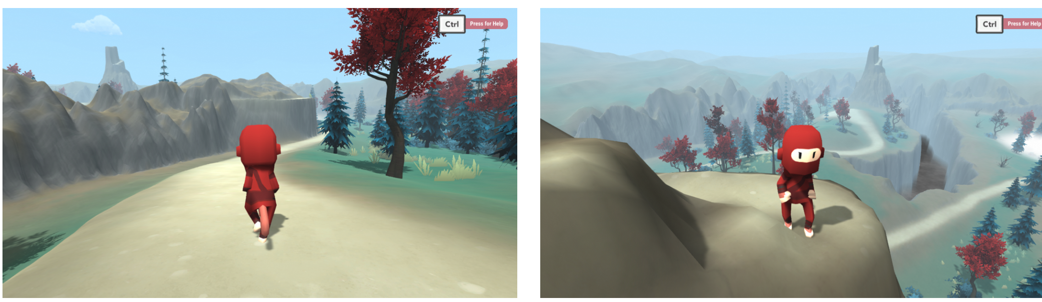

Fig. 1. Screenshots of the game in the nature aesthetic; one of two aesthetics in the experiment. Player character walking on the primary path (left), and looking down from 'Mountain B' (right).

Video games frequently feature 'open world' environments, designed to motivate exploration. Level design patterns are implemented to invoke curiosity and to guide player behavior. However, evidence of the efficacy of such patterns has remained theoretical. This study presents an empirical study of how level design patterns impact curiosity-driven exploration in a 3D open-world video game. 254 participants played a game in an empirical study using a between-subjects factorial design, testing 4 variables: presence or absence of patterns, goal or open-ended, nature and alien aesthetic, and assured or unassured compensation. Data collection consisted of in-game metrics and emotion word prompts as well as post-game questionnaires. Results show that design patterns invoke heightened exploration, but this effect is influenced by the presence of an explicit goal or monetary compensation. There appear to be many motivations behind exploratory behavior in games, with patterns raising expectations in players. A disposition for curiosity (i.e. 'trait curiosity') was not found to influence exploration. We interpret and discuss the impact of the conditions, individual patterns, and player motivations.

CCS Concepts: • **Software and its engineering** → **Interactive games**; • **Human-centered computing** → **Empirical studies in HCI**; **Empirical studies in interaction design**; *User studies.*

Authors' addresses: Marcello A. Gómez-Maureira, m.a.gomez.maureira@liacs.leidenuniv.nl, Leiden University, LIACS, Leiden, The Netherlands; Isabelle Kniestedt, i.kniestedt@tudelft.nl, Delft University of Technology, TPM, Delft, The Netherlands; Max van Duijn, m.j.van.duijn@liacs.leidenuniv.nl, Leiden University, LIACS, Leiden, The Netherlands; Carolien Rieffe, crieffe@fsw.leidenuniv.nl, Leiden University, FSW, Leiden, The Netherlands; Aske Plaat, a.plaat@liacs.leidenuniv.nl, Leiden University, LIACS, Leiden, The Netherlands.

2573-0142/2021/9-ART271
https://doi.org/10.1145/3474698

Additional Key Words and Phrases: curiosity, exploration, game user research, design patterns, video game design, player experience, player motivation

ACM Reference Format:
Marcello A. Gómez-Maureira, Isabelle Kniestedt, Max van Duijn, Carolien Rieffe, and Aske Plaat. 2021. Level Design Patterns That Invoke Curiosity-Driven Exploration: An Empirical Study Across Multiple Conditions. *Proc. ACM Hum.-Comput. Interact.* 5, CHI PLAY, Article 271 (September 2021), 32 pages. https://doi.org/10.1145/3474698

1 INTRODUCTION

Video games provide a wide range of experiences to those who play them [51, 54]. For some, playing games is about challenge, competition, and the thrill of proving their skills. Others play games to experience fantastical worlds and follow dramatic narratives. Players often play for a variety of reasons, with games frequently featuring multiple kinds of experiences to keep them engaged [4, 35, 67]. In many video games, developers combine narrative elements with an 'open world' environment that allows players to choose activities that they enjoy [22]. Letting players encounter game content as the result of their own curiosity promises to create a more memorable and enjoyable game experience, compared to using more authoritative directions [71]. However, players still need to be made aware of where purposefully designed game content can be encountered. Aside of outright directing players where to go, games structure their environment in ways that both invoke and direct the desire for exploration. In doing so, games can appear more interactive and 'free', as exploration is focused onto areas that are designed for it.

As a result, games benefit from invoking the desire to explore by design, and an increasing number of games feature environments that are created with this goal in mind. Games such as *Zelda: Breath of the Wild* [48] or *The Elder Scrolls V: Skyrim* [7] gained critical acclaim for instilling a sense of curiosity for exploration [50]. Although designers have an intuitive sense that curiosity is important to games [11, 33, 61], the way in which it can be purposefully invoked is not obvious and has seen limited empirical study. This is unfortunate, as a more evidence-based understanding of what invokes the desire to explore would provide not only a stronger foundation for the research of player experience, but also for the practice of game design, and even the development of engaging procedural environments.

With the study presented in this paper, we aim to perform fundamental work in filling this research gap. We assess the impact of four level design patterns that have been hypothesized to induce curiosity-driven exploration [18] by implementing them in a single-player, 3D game environment. These patterns should be understood as sets of parameters that, taken together, serve a specific design intention. The patterns are integrated through twelve individual implementations (three for each pattern) into a virtual game environment. The four implemented patterns are: (1) overcoming extreme points such as mountain peaks or other hard-to-reach structures, (2) resolving visual obstructions in the environment to discover what they hide, (3) out-of-place elements that appear to not 'fit' into the environment, and (4) understanding spatial connections between areas in the game environment. The expectation is that an environment purposefully designed to stimulate exploration will cause players to behave differently and regard the environment more positively.

In the experiment, exploration is measured by the combination of in-game actions (i.e., game metrics) and players' accompanying emotional investment (through a post-game questionnaire and self-reported 'emotion words' during gameplay). Gameplay is divided into a period of free exploration and one where participants need to wait before being able to complete the game (see Section 3.3). With this study, we aim to test the following **hypotheses**:

- **H1** (*H1a*) Level design patterns elicit more *exploratory behavior* and (*H1b*) positively affect *emotional experience* of the game.
- **H2** Having an explicit goal reduces *exploratory behavior*.
- **H3** Players with a higher predisposition for curiosity engage in more *exploratory behavior*.

The study of these level design patterns is the primary focus of the presented study. However, there are several aspects implicit to a digital game that may influence the behavior being studied here as well. For example, many games have a stated goal (e.g., a main quest) that directs players where to go. Game environments also do not simply consist of mere spatial volumes, but have a distinct environmental aesthetic to make the environment look enticing. To examine whether any differences between an environment with and without level design patterns are not due to these factors, we incorporated these variables into the experiment design. Finally, we also test for effects depending on whether participants are assured monetary compensation for their time or not. Compensation is common in research studies, and researchers of future studies building upon this work may consider to offer it to participants. In this study, we include compensation as a variable to examine whether it affects exploratory behavior in a video game. As such, this study follows a between-subjects 2x2x2x2 factorial design with the fixed factors: *patterns*, *goal*, *environment aesthetic*, and *compensation*.

We tested 254 participants, who were randomly sorted into different condition groups. Data collection consisted of questionnaires (both in- and post-game) and game metrics (e.g. distances from path and destination over time, play duration, position and rotation, and instances of going out of bounds; i.e. jumping into a chasm). In addition to direct measures, 'interpreted' emotion ratings were gathered through matching self-reported emotion words gathered during gameplay to affective components (valence, arousal, and dominance), based on the Glasgow Norms corpus [66].

The results of the study provide evidence for H1 and H2, but not for H3. Section 7 provides a full discussion on the interpretation of results. The contributions of this article lie in providing a first empirical study into level design patterns for exploration. Our results show that level design patterns clearly impact exploratory behavior, but their effects are influenced by other factors. Having an explicit goal severely reduces exploratory behavior until that goal is fulfilled and the game becomes more 'open ended'. Receiving monetary compensation reduces exploration as well, but patterns still induce players to explore more and perceive the experience more positively than when they play without. These findings can inform game design, but also the design of further studies in this area (e.g., variables to include, data analysis, and whether or not to offer compensation). Overall, this article aims to be a nuanced, practical example of the complexity in studying video games experimentally and analytically, and provides a valuable foundation for future work in the study of design patterns, curiosity, and player experience.

2 RELATED WORK

2.1 Curiosity

Curiosity, referred to as an intrinsic motivation to learn and explore, is considered to play an important role in many aspects of human life [38, 39, 58]. In education and research, it is credited as one of the most important factors for progress [21, 43, 63] and it has been considered a motivation behind creative efforts [65], technological advancements [23], and epistemic investigations [73, 75].

Most research efforts regarding curiosity have taken place in the fields of Philosophy [25, 63] and Psychology [5, 15]. The definition of curiosity varies for much of the past discussion, ranging from accounts of human aspirations to describing it as instigating stimulant for interaction with the environment[18]. Grossnickle [21] performed a meta-review of academic work in order to find commonalities in prior research. In the review, curiosity is discussed through various 'lenses' that focus on different aspects of curiosity. For example, some consider it a primal drive that requires satisfaction [5, 6], not unlike satisfying hunger [63]. Others posit it as a need to fill knowledge gaps [43]. In this interpretation, it is both necessary for a person to have existing knowledge that makes them aware of the gap, as well as to be capable of the evaluation that the gap is neither too large nor too insignificant to be filled.

Curiosity has been discussed both as a *trait* and as a *state*. As a trait, it describes a person's individual disposition to becoming curious, which is considered to be a fairly stable aspect of one's personality [42]. As a state, it is an 'in-the-moment' drive for exploratory behavior [43]. It has been separated into a feeling of interest (i.e., wanting to know for its own sake) and a feeling of deprivation (i.e., wanting to know because not knowing is frustrating or affects one negatively) [37]. Being in a state of curiosity can be both pleasant and uncomfortable, especially with growing intensity [39]. It follows that curiosity is a complex construct that involves both behavioral and emotional components [21] and that is interwoven with other emotions and motivations, such as uncertainty [3], surprise [49], and wonder [63]. As such, it is a challenging subject for empirical study.

Studies have shown an influential relationship between trait and state curiosity [29, 40, 57]. Most work on quantifying curiosity has been concerned with measuring trait curiosity [38, 41] or related personality traits, such as intrinsic motivation [13, 14, 45] or sensation seeking [78]. Although some researchers doubt the presence of stable curiosity traits [43], recent empirical work shows evidence for it [30]. Kashdan et al. [30] found that trait curiosity consists of five dimensions: Joyous Exploration (motivated by novelty), Deprivation Sensitivity (need of resolving), Stress Tolerance (ability to cope with uncertainty), Social Curiosity (wanting to know about others), and Thrill Seeking (enjoyment of anxiety). Each of these dimensions relates to a different type of motivation behind possible state curiosity and together they form the five-dimensional curiosity scale (5DC). Although dimensions of curiosity remain the subject of debate, the desire to engage in exploratory behavior is generally considered to be among them [5, 31, 69, 74].

In this study, we understand curiosity as **an intrinsic motivation for pursuing new knowledge and experiences** that is accompanied by **pleasure and excitement** [18].

2.2 Curiosity in Games

Video games are complex systems that give rise to a range of emotions and experiences [76]. They are a promising medium for the study of curiosity, as curiosity has been noted as an important, yet ill-defined, factor in their design. Costikyan [11] discusses the role of uncertainty in games, and describes curiosity as an important motivator to engage in gameplay. Klimmt [33] considers it part of a conceptual model for player engagement, i.e., a reason why people choose to play games. Schell [61] suggests that designers can stimulate curiosity by inspiring questions in players' minds through their game's design. Studies into player profiling seek to establish player types that involve personality traits and motivations, including curiosity [60]. Such player archetypes mirror aspects of Kashdan et al.'s curiosity model [30]. The BrainHex model [47], for example, features seven archetypes that match the characteristics of different dimensions of curiosity, such as the 'daredevil' archetype, which is defined by taking pleasure in overcoming risks, matching the 'thrill seeking' curiosity dimension. Further connections between different types of curiosity and gameplay have been established [18]. For example, social curiosity may relate to playing together or figuring out the stories behind non-player characters, deprivation sensitivity can indicate a preference for puzzles, and joyous exploration a desire to discover the ins and outs of a virtual environment.

Player types and curiosity dimensions thus give an indication of the types of gameplay that appeal to certain players. In the context of video games, exploration can happen in a variety of ways and for a number of reasons. Behaviorally speaking, exploration can happen *spatially* in a virtual space through the actions of a player avatar or *conceptually* by resolving knowledge gaps, e.g., through the solution of a logic puzzle [71]. Exploration in games is considered a form of creative expression and contributes to how satisfying players consider a game to be [54]. There are many ways to design a game, however, and not every approach may be equally successful in triggering state curiosity, even if the intended gameplay may align with a player's trait curiosity. Design research in- and outside of games has the challenge of both measuring interrelated idiosyncratic constructs, and informing future design implementations with insights gained by these measures [12]. Understanding why a specific design 'works' requires the analysis and creation of generalizable principles. To this end, design research is concerned with formulating knowledge that can inform the development of similar, but not identical, artifacts. One way of pursuing this, is by defining 'game design patterns' [8]; definitions of situations and elements that share structural similarities. These patterns can then serve as a guideline for the analysis and development of games. To et al. [71] investigated how game designers can elicit the curiosity of players, following a model of curiosity [34] that distinguishes between different triggers of curiosity. This approach is particularly useful for creating generalizable design guidelines, as it gives game designers a range of possible design interventions for invoking curiosity. Gómez Maureira et al. [18] analyzed the design of games that players associate with dimensions of the 5DC, and identified design patterns commonly used in games invoking spatial exploration (e.g., *Elder Scrolls: Skyrim* [7]) and social interaction (e.g., *World of Warcraft* [9]).

Level design patterns have been described in prior work [24, 32, 68], but are rarely empirically tested. This study aims to augment existing work by performing an empirical study of **level design patterns for spatial exploration** (described in Section 4). As such, we are concerned with studying player behavior and emotional responses to patterns that **shape the virtual environment**, in order to more closely assess which patterns are successful in motivating exploratory behavior and the emotional experiences that results from such behavior. Although games tend to use additional elements that induce players to explore the environment (e.g., providing objects to discover and collect [18]), this study purposefully limits such elements to study the design of the environment in isolation. We consider this an initial step in the further study of patterns for other types of curiosity. In itself, the results of this work (and other work that may build upon it) may provide insights into the design over game environments, which tend to become larger and more open-ended with each generation of games. So-called 'open world' games are based around invoking a desire for exploration. They give players freedom in how to progress in the game and what to interact with. Their success, therefore, is partially dependent on invoking this desire. Games will frequently strike a balance between 'expecting' exploration, and actively guiding player behavior. We pose that a better understanding how patterns impact player behavior and game experience can support designers in striking this balance.

2.3 Measuring Curiosity

As stated previously, curiosity is believed to have both behavioral and emotional components. Therefore, empirical studies examining curiosity should try to measure both these aspects. The measure of curiosity is informed by the measure of affective constructs in general [17, 62]. Previous work measured curiosity through questionnaires and psycho-physiological responses that are thought to relate to it [74]. Literature suggests that curiosity is experienced primarily as arousing [41], while the emotional experience (i.e., 'valence') of curiosity can be negative or positive. Whether or not curiosity leads to exploration can be measured by observing a person's actions. However, it can be challenging to 'correctly' attribute the motivations that elicited the behavior [63].

In video game research, emotional impact is generally measured retroactively through questionnaires once a play session has concluded (e.g., [10]). These instruments, however, are aimed at assessing the overall game experience, and lose valuable information regarding a player's moment-to-moment experience. As such, they are not enough to capture state curiosity as it occurs, and more immediate measures are required. Such measures could include the recording of physiological responses thought to correlate with an emotion [36], recording behavioral data through game metrics (e.g., position, rotation, actions), or incorporating ways of eliciting player feedback during the game. In general, a combination of measurements is desirable in order to comprehensively capture a player's experience [20]. There is no standardized method of eliciting feedback while playing, and adding measures also means that the measuring process and analysis rapidly form an increasing burden on both the player and the researcher. Ideally, measurements happen at the moment that they are happening, and capture relevant data (subjective and objective) from both the player and the game.

Developers have experimented with in-game measures for testing purposes. For example, the game *Subnautica* [72] implemented a form to make it easy for players to report issues during their game session [28]. The form contains a free-form text box to provide feedback, check-boxes for the type of feedback (general, gameplay, bug, frame rate), the option to include a screenshot, and a 4-point scale emotion rating (visualized through a range of angry to happy faces). The player's in-game coordinates, frame rate, and computer specifications are automatically logged with each submission. As such, the form combines elements of game metrics and subjective measures to form an in-the-moment picture of a player's experience.

For our study of level design patterns, we pose that players' movements through the areas where those patterns are implemented can be used as an indication of exploratory behavior. For our assessment of the motivations behind such exploration (i.e., whether or not that behavior is driven by curiosity) we rely on in-game prompts and post-game surveys. We describe the included measures in Section 5.3.

3 THE GAME DESIGN

The game created for this study is *Shinobi Valley*, a single-player video game inspired by action-adventure games like *The Legend of Zelda: Breath of the Wild* [48]. It was tested over multiple iterations during development and experimentally validated through a pilot study (see Section 5.5).

In *Shinobi Valley*, the player controls the character of a ninja (shinobi) monkey and explores a 3-dimensional virtual environment, which takes the shape of a valley between mountain ranges. A path snakes through the valley from one corner of the map to the other. At the end of this path, the monkey reunites with his ninja master and joins him in meditation.

There are no enemies in *Shinobi Valley* and no specific obstacles to overcome. There is no music in the game, but there are ambient noises such as wind and rustling leaves. The character itself also produces sound while walking and jumping. The atmosphere is generally quiet and calm. The game is played in browsers using either the mouse and keyboard or purely mouse-based controls. As such, players have to be comfortable with using either control scheme and be capable of moving a virtual character in 3D space.

Shinobi Valley was created for this study. Video games are often developed with the goal of providing as much entertainment as possible. They do this through a multitude of overlapping and interacting systems and mechanics, making it difficult to assess what each individual aspect contributes to the gameplay experience. Using a custom-made game, however, allows for detailed manipulation of the condition design, and closer examination of player behavior and experience (e.g., due to the ability to capture game metrics and intervening with other measurements) than using a commercial title would allow. The downside is that a game created solely for research purposes may be lacking in qualities that commercial games possess. We reflect on the impacts of this decision in Section 8.

3.1 Controls

The camera in *Shinobi Valley* has a third-person perspective and, as such, is positioned behind the player character. Players can move their character with the WASD or arrow keys, or by holding the left mouse-button down. Jumping is done by pressing SPACEBAR or clicking the right mouse-button. Running is done by holding SHIFT. If players are holding the left mouse-button to walk, the character will start running automatically after a few seconds. Players can turn the camera by moving the mouse. Players steer their character in the right direction by turning the camera. Camera turning direction and sensitivity can be controlled from a settings menu, which is accessed by pressing CTRL.

3.2 Tutorial

The tutorial is presented in sequence. Players are welcomed via a text message and introduced to their character. At this point, the player cannot yet control their character freely. The game explains the controls step by step, each time pausing for players to try them out before continuing. It first teaches the player to control the camera with the mouse. Then, it allows them to try and reverse the camera direction if they want, followed by the option to adjust camera sensitivity. Next, walking controls are explained. Finally, the jump button is explained and the player is told they can change settings at any time. From this point, the player is free to move around.

3.3 Progression

At the start of the game, players see their character in the virtual environment. After players complete the tutorial they are free to explore the environment as they wish. They will start on one side of the main path, which traverses the entire level. The ninja master is always on the opposite end of the path, sitting on a rock. Upon reaching the Master, the player is informed that he is meditating and to come back later. From this point on, a timer starts running. The timer is 5 minutes long, or 2.5 minutes if players have played for over 10 minutes before reaching the Master. Players can check back before the timer runs out, but the Master will still be meditating. The message will update to reflect the passing time, though it does not give an exact indication of when the Master will be done. Once the timer runs out, the Master will stand up and the player can join him. The game ends with the player and the Master meditating together.

The intention of the waiting period between reaching the Master and finishing the game is to encourage participants to engage with the environment. Players may choose to follow the path instead of responding to the design patterns. In such cases, patterns may have been either not noticed, or not provided a strong motivation for exploration, as compared to reaching the end of the path. The waiting period provides a separate occasion to evaluate the relative attraction of design patterns.

Interaction with design patterns in the waiting phase are evaluated separately, since exploration at that point might be more motivated by a desire to alleviate boredom. While players might not be in a general state of curiosity if they are bored, it is still instructive whether and which specific design patterns attract exploration under such circumstances. Players may still be curious about features in the environment while experiencing boredom.

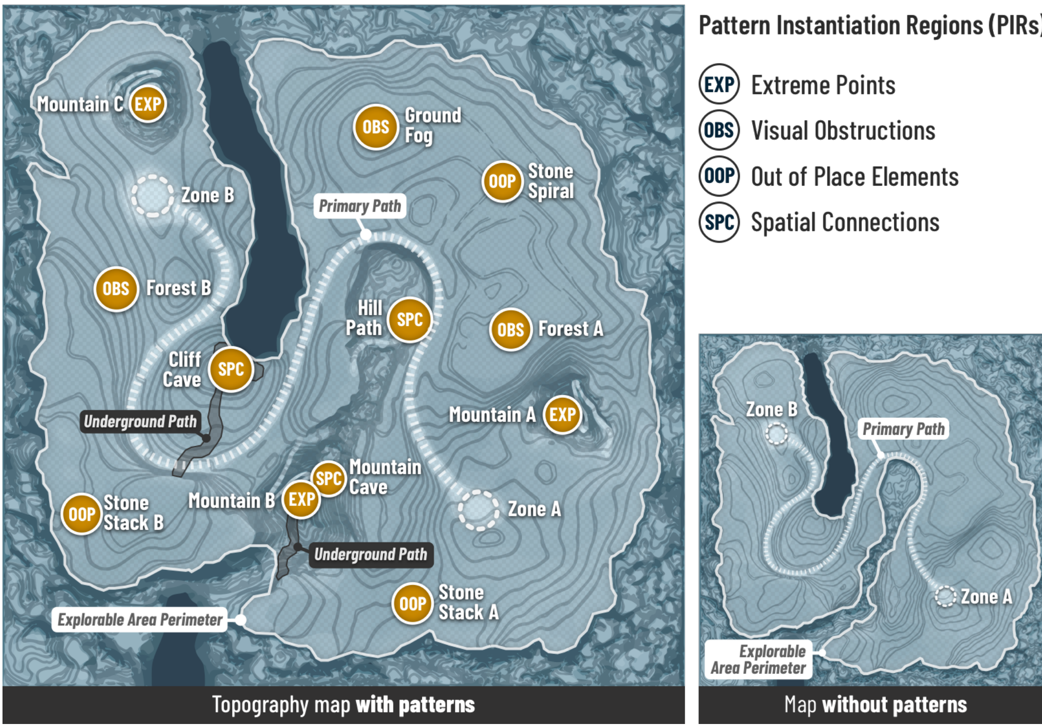

Fig. 2. Top-down schematic of the game environment with (left) and without (right) patterns present. The left schematic illustrates the location of pattern instantiation regions (PIRs).

4 INDEPENDENT GAME VARIABLES

To examine the research questions, the game's design can vary in three aspects; *presence of patterns*, *presence of a goal*, and *environment aesthetic*. Participants are randomly assigned a combination of variables. While the controls and general progression remain as explained above throughout the different versions, the environment differs depending on which condition (i.e., combination of variable states) a participant is assigned. Variables are tested pair-wise, meaning that different variable states overlap with others. Each participant plays only one possible version of the game. Each of the variables is described in the following sections.

In addition to game variables, it varied per participant whether or not they received financial compensation for participation. This variable is further discussed in Section 5.2.

4.1 Level Design Patterns for Spatial Exploration

The primary goal of this study is to examine the effects of level design patterns on exploratory behavior and emotional experience of a game. Hence, the primary focus of the study is the *presence or absence of level design patterns for spatial exploration in the game environment*. In both versions, the environment is bordered by steep cliffs the player cannot traverse. It also includes a chasm and a low mountain ridge to keep the player from cutting straight across to the other side. Finally, a *'primary' path* forms an S-curve from one side to the other (Figure 2).

Without patterns: The valley is purposely simple in design. It is fairly flat and sparsely populated with trees, with little to no outstanding features.

With patterns: All four level design patterns are implemented within the overall framework of the valley's design. Their presence alters the environment significantly, adding multiple high points, objects, and spatial connections that can draw the player's attention. Each pattern is implemented three times into the environment for a total of twelve unique implementations. In this paper, we refer to a design pattern implementation as a *pattern instantiation region* (PIR). Each PIR is situated off the path that crosses the valley (see Figure 2). The four design patterns and their resulting PIRs are discussed below.

4.1.1 Reaching Extreme Points (EXP). Games that encourage exploration often feature locations considerably higher than the rest of the game environment. In *Zelda: Breath of the Wild* [48] players climb mountains and towers to gain an overview of the environment. Alternatively, games may also include locations that are at extreme depths (e.g. *Subnautica* [72] and *Minecraft* [46]). In *Shinobi Valley*, this pattern is recreated through three tall mountains that players can climb via trails taking several hairpin turns. For two mountains (Mountain A and C), the trail is placed out of sight from the path and thus requires some effort to find. A third mountain (Mountain B) is placed near the middle of the path. The trail to its summit, while still out of the way, is more easily visible.

4.1.2 Resolving Visual Obstructions (OBS). Parts of a game environment can be deliberately obscured to motivate exploration. For example, strategy games often use a 'fog of war' [8] to hide areas in the game world. In our study, a version of this pattern is implemented through an area covered in fog (Ground Fog), as well as two areas consisting of dense bushes and trees (Forest A and B). Both implementations require players to explore an area closely if they want to see beyond the obstruction.

4.1.3 Out-of-Place Elements (OOP). Out of place elements are game objects that stand out in the context in which they are placed. For example, this could be a single tree within a field of stone statues, or a stone statue inside a forest. Players understand games as designed spaces and might therefore ascribe meaning to elements that appear to break a perceived pattern. In *Shinobi Valley* three regions are meant to represent this pattern: two areas with stone stacks (stones placed on top of each other to form columns; Stone Stack A and B), and a spiral of stones on the ground (Stone Spiral).

4.1.4 Understanding Spatial Connections (SPC). Games that allow players to navigate through an environment might feature complex, interconnected paths. Even when such paths are not designed to present a challenge in itself (as is the case in a labyrinth), exploration can be motivated by the desire to learn how spaces connect to each other. *Shinobi Valley* includes three regions that seek to motivate exploration in this manner: two cave systems (Cliff Cave and Mountain Cave), and a path leading to a hill plateau (Hill Path). These patterns are designed in such a way that a location is marked as reachable (e.g., by using the same visual appearance as the primary path), but without a clear indication as to how the player can get there. Players thus need to figure out how they can get to the location.

4.2 Goal

Many games are goal-driven activities with clear rules and win conditions. Even when they aim to provide the experience of 'free' exploration through seemingly open environments, they still have multiple ways of directing the player towards specific game elements (e.g., through an overarching story line, side quests, and 'collectibles'). These give structure to the game experience, even if many

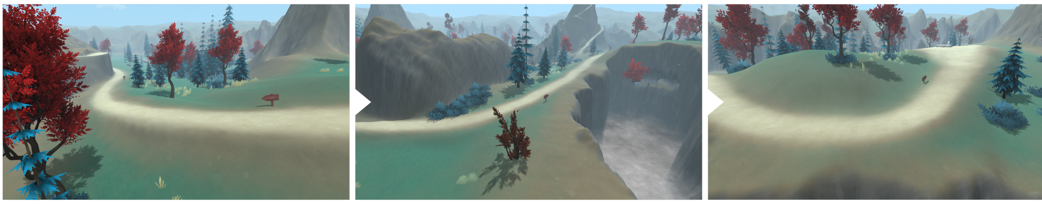

Fig. 3. Camera shot sequence at the beginning of the 'with goal' experiment factor, following the primary path. The sign posts shown in the sequence are not present without a goal.

players will stray off the intended path to explore. The inclusion of a goal (or context for the player's activities) in *Shinobi Valley* therefore makes sense. However, when a player is given particular instructions on what to do, the possibility exists that this takes precedence over everything else. In order to examine the effect of providing the player with an explicit goal or not, the second experiment condition concerns playing the game *with and without a goal.*

With goal: In this version, players are given a goal; *"This is you! You are a monkey ninja on a journey to meet your master. Your master awaits your arrival at the end of this path."* The message is followed by a 3-part camera sequence over 5 seconds that shows locations along the primary path towards the instructed goal (Figure 3). Along the path, several wooden signs point players towards their destination. After this introduction, players complete the tutorial to ensure they understand the controls. After completing the tutorial, a message states; *"Now go! Your master awaits."*

Without goal: The players do not receive any information about the player character or a goal. The primary path in the environment still acts as a suggestion, but there are no wooden signs pointing players to a specific direction. After completing the tutorial, players receive the message, *"You can now control your character."*

4.3 Environment Aesthetic

The main focus of this study is to explore the presence or absence of level design patterns. Our secondary interest is the presence of an explicit goal given to the player, and how this may influence exploratory behavior. Games, even those specifically designed for research studies, quickly become complex, and they are sensitive to confounding variables. *Shinobi Valley*'s design is purposefully limited in areas outside those directly related to the study goals to limit potential interference. It has, for example, very minimal sound design and lacks many elements common to the type of game it is based on (e.g., combat or obstacles).

However, video games generally feature a theme or environment aesthetic. The setting of a game can greatly influence how likely a particular player is to play a certain game (e.g., some people are attracted to science fiction settings, others to medieval fantasy). As such, the visual design of a research game may appeal differently to individual players, or have other unintended consequences. To examine whether any measured differences in behavior and emotional experience persist within a different aesthetic, we can change the game's aesthetic from a *nature* to an *alien* environment (see Figure 4). Both versions are created to be aesthetically pleasing. Neither setting changes the game environment structurally — they only alter its visual appearance.

Nature setting: The nature setting is lush and green, with a variety of green spruces and red maple trees. It invokes the feeling of being in the Japanese countryside.

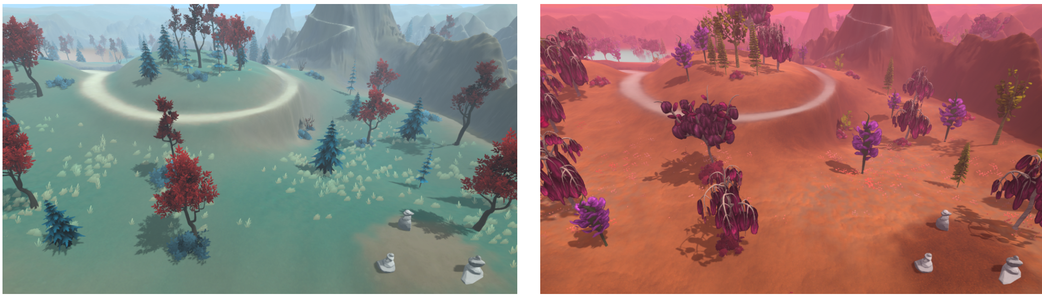

Fig. 4. Screenshots with the two aesthetics of the game: nature and alien.

Alien setting: The alien setting places the player character in a space suit on an alien planet. The environment consists of otherworldly colors and the vegetation has an insect-like quality.

5 RESEARCH DESIGN

The goal of this study is to examine the effect of level design patterns for spatial exploration. This study follows a between-subjects 2x2x2x2 factorial design, with the fixed factors of *patterns*, *goal*, *environment aesthetic*, and whether or not participants are *compensated*. Participants are randomly sorted into a condition group, given one of two options in each fixed factor. Participants play their randomized version of *Shinobi Valley*, during which data is collected in the form of game metrics and a periodical in-game survey. Additionally, participants answer a post-game questionnaire. The following section describes the *game randomization*, the *sampling methods*, the *measurements and data processing*, the *experiment procedure*, and the *pilot study*.

5.1 Game Randomization

Participants each played one possible game version, consisting of a combination of the independent variables described in Section 4. Whenever a participant begins the experiment, they are assigned a combination of the three game variables (i.e., patterns, goal, aesthetic). Each variable has two possible options. Goal and aesthetic each have a 50% chance for either option.

The 'pattern' variable has a 70% chance of playing *with* level design patterns present. Although this creates an imbalance in the sample sizes between the different conditions, it allows for more data to be collected on participants interacting with the (individual) level design patterns and a subsequent analysis of those interactions. Considering that the analysis of patterns is our primary goal in this study, we decided that this was a valuable trade-off to maximize available resources (i.e., time and budget for financial compensation). We reflect on the implications of this decision in Section 8.

The map of *Shinobi Valley* was designed to be reversible in spatial layout and PIR placement. Each participant's starting position is randomized between the two ends of the main path (see Figure 2). The Master is always positioned on the opposite end of the path. By randomizing the starting position, we aim to avoid uneven interaction with PIRs because they unintentionally stood out more or because participants encountered them first. This could also have been achieved by randomizing the PIRs themselves. However, this would have been far more complex to achieve as the PIRs (especially mountains and spatial connections) are integrated into the environment. The starting location, or *play direction*, is included in statistical testing as a nuisance variable.

5.2 Sampling

Participants were recruited through a combination of snowball sampling and crowd-sourcing. Snowball sampling included reaching participants through social media, the university environment, and word-of-mouth. Crowd-sourcing occurred over three platforms: Mechanical Turk [2], Prolific [55], and SurveyCircle [70]. Mechanical Turk and Prolific participants received an assured monetary compensation for participating of 3.00 EUR. All other participants could opt into a lottery for three 20.00 EUR vouchers. The only requirement for participating in the study was that participants have basic English comprehension and that their PC can run the game smoothly (Section 5.4). From Prolific, only female participants could participate; a decision made to counter-balance a lean towards male participants up to that point in the data collection.

Players are intrinsically motivated to engage with games for the experiences that they offer (e.g., enjoyment from overcoming challenges, interacting with the game world, social interaction) [59]. However, we consider it possible that playing a game for research purposes may influence how players engage with it, especially when they are compensated for their time. While some may be inclined to finish as fast as possible to maximize their gains, others might feel a sense of obligation to do well and 'earn' their reward. In this study, where we aim to study participants' exploratory behavior, the addition of an extrinsic reward may affect intrinsic motivation [16] and possibly, in turn, the behavior we are studying. To our knowledge, the effect of an extrinsic reward has not been tested for how it influences exploration in video games. As such, we pose that future studies may benefit from our examination of this effect. In our statistical analysis, **whether players did or did not receive assured monetary compensation is therefore treated as a fourth independent variable**.

Overall, 389 participants took part in the experiment, out of which 266 completed the game and the post-game survey. Incomplete measures, and participants accounting for the fastest 2.5% of survey completions compared to the median, were excluded. Of the remaining 254 participants, 48% were female (n=122), 50.8% male (n=129), and 1.2% (n=3) identified as non-binary. The mean age is 31.8 (SD=10.8, range=18-69). 35% of participants were recruited from MTurk (n=89), 31.1% through snowball-sampling and social media (n=79), 28.7% from Prolific (n=73), and 5.1% from SurveyCircle (n=13). The female participants recruited via Prolific balance out otherwise male dominant demographics (without Prolific, 69.6% of participants are male).

5.3 Measures

Data is collected during and right after gameplay. During gameplay, player behavior is logged through the use of game metrics and a repeating in-game survey. Once participants finish the game, they fill out a post-game survey. Each of the measures is described in the following sections.

Game Metrics: The game logs player parameters at 5 Hz. Each log-line includes position, camera rotation, avatar movement velocity, closest distance to the primary path, and distances to start and destination points. The player controls for jumping and running are logged at the time of input as timestamped events. Further timestamps are logged for arrival and leaving of PIRs, arrival and leaving of start and destination areas, instances of resetting the player character (when jumping into the chasm, or when getting stuck in the level geometry for longer than 2 seconds), and triggering and completion of the in-game survey.

Additional metrics are generated from the aforementioned measures; primarily through the use of spatio-temporal data created by participants in combination with the predefined locations of PIRs. These include measures for PIR visit counts, PIR stay duration, and spatial entropy of player movement (specifically, *Altieri's entropy* [1], which captures the impact of localized clustering in addition to overall heterogeneity of spatial data).

In-Game Survey: The in-game survey is a pop-up window that appears at predefined intervals during play, i.e., at minutes 1, 3, 5, 8, 12, 17, 23, and 30 from when the tutorial finishes. It is likely that players finish the game before having been exposed to all survey instances. The time between surveys increases over the play session to mitigate interruption of exploratory behavior. The timer for the next survey only starts when the previous one has been completed, ensuring that all participants experience the same amount of time between survey moments. The in-game survey also pauses the wait timer that runs after players have met the Master ninja character. The in-game survey asks participants 'how curious [they] feel at the moment' using a sliding scale between 0.0 and 1.0; players only see the visual representation of the slider position. Additionally, they are asked to describe their current emotional state in a single word using free text. Players need to enter text in order to continue.

Post-Game Survey: The post-game survey consists of four parts: demographics, questions about the game experience, selected modules from the Game User Experience Satisfaction Survey (GUESS [54]), and the Five-Dimensional Curiosity Scale (5DC [30]).

Data is collected on both participant age and gender. Participants are asked to rate what type of video game player they are on a five-point scale (ranging from 'Novice' to 'Core'), as well as their time spent playing games in the past year (ranging from 'Never' to 'More than 1 hour per day on average').

Additional questions ask participants to report about their game experience. They are asked whether they left the path (yes/no) and why (free text). Furthermore, they are asked whether any elements in the game stood out (positively or negatively), and whether they have any other comments about the experiment.

The GUESS is a validated questionnaire that examines game experience in a number of areas. Modules can be used independently, depending on the needs of the project. For this study, the modules 'Enjoyment', 'Creative Freedom', 'Play Engrossment', and 'Personal Gratification' were included to assess emotional investment in the game experience.

The 5DC is a validated survey assessing a person's trait curiosity (i.e., disposition to become curious) across five dimensions; Joyous Exploration, Deprivation Sensitivity, Stress Tolerance, Social Curiosity, and Thrill Seeking. It is included to examine whether certain types of trait curiosity correspond to participants' exploratory behavior or emotional investment in the game.

Data Processing: Game metrics are processed through custom Python scripts using the Pandas library [56]. Based on the in-game survey, 'emotion ratings' are formulated by matching free-text entries ('emotion words') to affective components (valence, arousal, and dominance) based on the Glasgow Norms corpus [66]. Qualitative data from the post-game survey is coded based on topic to discover patterns within the data (e.g., mentioning of specific PIRs, design patterns, and other game aspects). Survey results are compiled per category based on author instructions. Spatial entropy is calculated by re-sampling spatio-temporal data into a grid of 60 x 60 squares with player location increasing the value of each square by 1 for each 5 Hz interval, thus increasing entropy both for longer durations, and movement in the game environment. Calculations are created through the 'spatialentropy' Python package [77].

5.4 Procedure

Shinobi Valley is online and can be played in most modern browsers. Participants are directed to the game website, which opens with a description of the experiment. Participants are not informed that curiosity or exploration are the topics of the research. Instead, they are told that they will play a game for research and how long the study is scheduled to take (around 15-20 minutes).

The game's web-page checks whether the browser window is large enough for the game to be played in full resolution. If the window is smaller than HD dimensions (1920 by 1080 pixels), the participant can not proceed and instead sees a message to play the game on a larger display or resize their window. If the resolution is large enough, the participant can proceed to the game.

The game informs the participant that it is best played with headphones. The participant then proceeds with playing the game, starting with the tutorial (section 3.3). Participants are randomly sorted into a condition group and assigned a starting position (see Section 5.1).

During play, the game periodically checks the frame-rate at which it is running. When it registers less than 15 frames per second, the game stops playing and informs the participant that a performance issue has been detected. This check aims to ensure that participants experience the game at a minimum viable frame-rate, as bad performance can impact measured player behavior. As the participant moves around the virtual environment, the game logs their in-game actions. Periodically, the in-game survey appears in a 'pop-up' window that cannot be dismissed until it has been filled out.

The participant eventually finishes the game by finding the Master and waiting for the end of his meditation. Upon completing the game, the browser directs the participant to the post-survey website (section 5.3).

5.5 Pilot Study

A pilot study was conducted with 24 participants [19]. The goal of the pilot was to test the quality of the game, perform a preliminary measure of design pattern effects, and test the measurements and procedure described above. Participants responded positively to the game and exhibited exploratory behavior while playing without specifically being prompted to do so. This suggested that the game was of sufficient quality to be used in larger studies.

6 RESULTS

The statistical tests conducted in this article follow a Bayesian approach [52], and are calculated using JASP [26, 44]. The reported 'Bayes factor' (BF_{10}) indicates the probability of the presence of an effect versus the absence [64]. A BF value over 1 indicates that the tested hypothesis is more likely than the null-hypothesis. A BF value of 1 indicates that there is an equal chance of the hypothesis being different from the null-hypothesis as there is of them being similar. A value lower than 1 indicates that the null-hypothesis is more likely. Unlike classical hypothesis testing, a Bayesian test can therefore be used to indicate likeliness of the null-hypothesis, rather than only reject it [53]. This article uses BF synonymously with BF_{10} (indexes are provided when not testing against the null hypothesis). Following current common practices [27], we consider $BF > 3$ as moderate evidence for a hypothesized effect (i.e. at least three times higher likelihood of a hypothesized effect versus no effect), while $BF < 0.33$ indicates moderate evidence against the hypothesized effect (effectively, at least three times higher likelihood against a hypothesized effect). We conduct two-sided Bayesian T-tests to test whether observations are significantly different, and two-sided Bayesian Pearson correlations to assess relationships between measures. In the absence of well informed (and sourced) prior beliefs, the default values for uninformed priors provided by JASP are used. Data for statistical tests is prepared using the Pandas package in Python [56].

A report of statistical tests and the underlying data can be found in the supplementary material of this article[1], including settings that were used to calculate the results. The following sections describe noteworthy results derived from the experiment data. First, we describe the participant sample and general observations. We then present results relevant to the experiment factors and

[1]Support material and study data is also available from this OSF repository: https://doi.org/10.17605/OSF.IO/MVR37

	Arousal				Dominance				Valence			
	Indiv. Means		Indiv. SDs		Indiv. Means		Indiv. SDs		Indiv. Means		Indiv. SDs	
	M_M	SD_M	M_{SD}	SD_{SD}	M_M	SD_M	M_{SD}	SD_{SD}	M_M	SD_M	M_{SD}	SD_{SD}
Before Waiting	5.18	1.48	0.23	0.51	5.62	1.10	0.20	0.44	6.30	1.80	0.33	0.72
While Waiting	4.41	1.31	0.71	0.68	4.90	1.02	0.50	0.56	4.70	1.88	0.75	0.90

Table 1. 'Glasgow ratings' based on translated emotion words, split out over the three emotion dimensions arousal, dominance, and valence. For each, individual means and standard deviations are listed, which in turn provide means and standard deviations across all participants. Data points are split out to ratings 'before' and 'while' waiting.

the recorded nuisance variables. Finally, we present results relating to the performance of the four PIR sets (extreme points, visual obstructions, out-of-place elements, and spatial connections) and their interrelation with experiment factors.

6.1 Descriptive Statistics and General Observations

To recall, the study uses a between-subjects 2x2x2x2 factorial design, with play direction and gender recorded as nuisance variables. Each participant contributes data to each of the four factors, and is randomly assigned one of the two conditions of each factor (even probabilities for all experiment factors and play direction, but 70% probability for playing with patterns).

The participant breakdowns for the individual experiment factors are: With patterns 72.4% (n=184; n=70 without patterns), with goal statement 51.2% (n=130; n=124 without), alien environment 48.8% (n=124; n=130 in nature environment), with financial compensation 63.8% (n=162; n=93 without). 50.8% of participants played in A2B direction (n=129; n=125 in B2A direction). The participant breakdown closely matches the randomization percentages that were set as part of the experiment design. Participant procurement for the non-compensation group fell slightly short of the even-split target.

The mean frequency of playing games was 5.5, corresponding to an in-between of '1 hour per week on average' and '1 hour per day on average' (SD=1.5). The mean gamer type identified as between 'casual' and 'core' (M=2.6, SD=0.9). Mean play time was 10.3 minutes (SD=3.6, range=6.6-30.4). The mean play time before waiting was 3.7 minutes (SD=3.7, range=0.6-27.7), vs. 6.5 while waiting (SD=1.4, range 2.7-12.4). GUESS ratings for the game were M=4.2 (SD=1.2) for 'creative expression', M=3.9 (SD=1.3) for 'play engrossment', M=3.7 (SD=1.6) for 'enjoyment', and M=4.3 (SD=1.3) for 'personal gratification'. All are close to the scale midpoint of 4.

The mean in-game curiosity rating was 0.6 (SD=0.2, range=0.01-1.0); slightly above the scale mid-point of 0.5. Ratings steadily decreased over the play session, with M=0.71 (SD=0.2) for the 1st rating going down to M=0.49 (SD=0.3; n=34 due to differences in play length) for the 5th rating moment.

Players used 104 unique emotion words to rate their emotional state when providing curiosity ratings. Before waiting, the most frequent responses were 'curious' (12.6%), 'calm' (8%), and 'happy' (8%). While waiting, the most frequent responses were 'bored' (24.4%), 'annoyed' (7.5%), and 'curious' (6.4%). Translation of the words to emotion dimensions via the Glasgow Norms resulted in means and SDs for each dimension (i.e., arousal, dominance, and valence) for each participant. SDs indicate the 'emotional range' of words provided by the participant throughout the play session. From the individual participant results, we can calculate means and SDs for each emotion dimension across participants as well — resulting in a 'mean of individual means', 'SD of individual means', 'mean of individual SDs', and 'SD of individual SDs'. The resulting 'in-game Glasgow ratings' are listed

			Pattern split		Goal split	
Tag	Meaning	counts	with	w/o	with	w/o
explore	explore in general	142	105	37	67	75
wait	to pass time while waiting	80	51	29	61	19
landmark	explore a landmark (unspecified target)	45	40	5	25	20
boredom	to alleviate boredom	32	21	11	17	15
mountain	to go to specific PIR: mountains	26	25	1	15	11
boundaries	to test limits of environment	26	17	9	13	13
expect	to find an expected game element that is not implemented	23	15	8	11	12
rocks	to go to OOP PIRs	16	16	–	7	9
scenery	to look at aesthetic elements in the environment	10	7	3	6	4
fog	to go to specific PIR: ground fog	10	10	5	3	7
chasm	to go to explore chasm (possibly incl. cliff cave PIR)	10	6	4	7	3
landmark	game area: any landmark	68	63	5	30	38
noInteraction	lack of interactive elements in the game	60	43	17	35	25
scenery	aesthetic elements of the game environment	58	41	17	29	29
wait	negative experience of having to wait	32	19	13	16	16
mountain	game area: mountains	32	30	2	17	15
cave	game area: caves	27	27	–	13	14
relaxing	overall atmosphere is experienced as calming	20	15	5	13	7
fog	game area: fog	19	19	–	9	10
rocks	game area: stone stacks	13	13	–	3	10
noReward	lack of validation for actions by the player	12	11	1	5	7
noGoal	lack of purpose or goal	11	6	5	3	8
val-pos	Valence of comment: positive	57	44	13	31	26
val-neg	Valence of comment: negative	36	22	14	12	24
val-neutral	Valence of comment: neutral	30	23	7	13	17
val-mix	Valence of comment: mixed positive and negative	18	12	6	10	8

Table 2. Notable 'tagged' comments (coded by authors) with total counts, as well as split by 'with' or 'without' pattern and goal condition. Rows in the upper section are coded from reasons given for leaving the path, rows in the middle section are coded from elements that stood out to participants, and rows in the bottom section are valence of comments left for impression of the game as a whole. The table only contains comments that were given by at least 10 participants in total, and that are particularly relevant to the posed hypotheses. Rows are sorted by overall count within each section.

in Table 1. A paired-samples Bayesian T-Test of Glasgow ratings before and after waiting shows decisive evidence for measures differing between the two experiment phases (BF > 1k). Individual means are higher before waiting for all emotion dimensions, while individual SDs are lower before waiting.

Most measures of player behavior differ notably before waiting, compared to while waiting; such as play duration (shorter before waiting; BF > 1k), movement speed (BF > 1k) and camera motion (BF > 100), and spatial entropy (BF > 1k).

Results of qualitative data, gathered in the form of coding participant comments, are shown in Table 2.

6.2 Fixed Factor Results

In order to examine the impact of fixed factors on player behavior and emotional experience, we ran Bayesian ANOVA tests for a number of dependent variables. These include game metrics (i.e., distance traveled from path, distance traveled from destination, play duration, position, rotation,

	Before Waiting		While Waiting	
Dependent Variable	Best Model (in order of incl. probability)	BF_M	Best Model (in order of incl. probability)	BF_M
Spatial Entropy	$Goal^{1k}$ + Pat^{10} + [Goal × Pat]	25	$Comp^{10}$ + $Goal^{10}$ + [Comp × Goal]	33
Path Dist. (M)	$Goal^{1k}$ + Pat^{100} + [Goal × Pat] + $Comp^{10}$	38	Pat^{1k} + [Goal × Comp] + $Comp^{3}$ + Goal	48
Path Dist. (SD)	$Goal^{1k}$ + Pat^{10} + $Comp^{10}$ + [Goal × Pat]	22	[Comp × Goal] + $Comp^{10}$ + Goal	31
Destination Dist. (M)	$Goal^{10}$ + Pat^{3} + [Goal × Pat] + Env	12	$Goal^{100}$ + [Goal × Comp] + $Comp^{3}$	34
Destination Dist. (SD)	*Null*	50	$Goal^{10}$+[Goal × Pat]+$Comp^{3}$+[Goal × Comp]+Pat	12
Duration	$Goal^{1k}$ + Pat^{3} + [Goal × Pat]	38	*Null*	26
Position Delta (M)	Goal	14	$Comp^{100}$ + Goal + [Comp × Goal]	17
Position Delta (SD)	$Goal^{1k}$	13	$Comp^{1k}$ + Pat	32
Rotation Delta (M)	$Goal^{10}$	25	*Null*	18
Rotation Delta (SD)	$Goal^{3}$	39	*Null*	21
Out of Bounds	$Goal^{10}$ + Pat + [Goal × Pat]	19	$Comp^{100}$	33
Glasgow Arousal (M)	*Null*	48	Pat^{3}	25
Glasgow Arousal (SD)	$Goal^{100}$ + Pat	16	$Goal^{3}$	39
Glasgow Dom. (M)	*Null*	24	*Null*	13
Glasgow Dom. (SD)	$Goal^{3}$ + Pat	12	Pat^{10} + $Goal^{10}$	18
Glasgow Valence (M)	*Null*	22	Pat^{3} + $Comp^{3}$	20
Glasgow Valence (SD)	$Goal^{10}$ + Pat + Comp	7	$Goal^{10}$ + Pat^{10}	33

Table 3. Fixed factor ANOVA results: Best models of Bayesian ANOVA for dependent variables, if model Bayes factor $BF_M > 3$. Results are split out for 'before waiting' and 'while waiting'. Individual fixed factors in the model are sorted in descending order by probability of inclusion in the model (BF_{incl}, see supplementary material for values). Post Hoc T-Test results (BF_{10}) are included as superscripts if at least $BF_{10} > 3$, reported in steps: >3, >10, >100, >1k (1000). Interaction effects (e.g. [Goal × Pat]) do not have an associated BF_{10}. Fixed factor abbreviations are: Pattern (Pat), Compensation (Comp), and Environment (Env).

and instances of going out of bounds; i.e. jumping into the chasm) and the Glasgow emotion ratings described previously. Such a test results in a list of models (comprised of different combinations of the fixed factors), that have a likelihood of explaining differences in a specific measure (expressed as BF_M). For each fixed factor (and possible combination of fixed factors) a likelihood is calculated that they are part of a model that explains the difference (expressed as BF_{incl}). Finally, Post Hoc T-Tests show the likelihood a fixed factor contributes to differences in a particular measure (expressed as BF_{10}).

The Bayesian ANOVA tests were ran across the fixed factors *patterns*, *goal*, *environment*, and *compensation*. Two nuisance factors (gender and play direction) were added to the null model, and were thus included in all tested models. The exact settings are included in the supplementary files. Table 3 shows an overview of the tests, including the best model to explain differences in each variable and the BF_{10} value of each fixed factor when $BF_{10} > 3$.

6.3 Differences Between Design Patterns

To compare differences between the design of patterns, measures relating to individual pattern instantiation regions (PIRs) are grouped into PIR sets: extreme points (EXP), visual obstructions (OBS), out-of-place elements (OOP), and spatial connections (SPC). Each PIR set consists of 3 instances in the game environment. All results for differences between patterns are based on n=184 (about 70% of the overall sample) as participants without patterns obviously do not contribute data. PIR set measures are based on player activity in a predefined radius around individual PIRs of 8 game engine units (roughly equivalent to meters). For each PIR, three measures are considered: spatial entropy ('dispersion' of player movement), visit count (unique entries into a PIR lasting at least 1 second), and stay duration. Given that only movements within a confined radius are considered for each PIR, spatial entropy only gives an indication of player movement within the PIR radius.

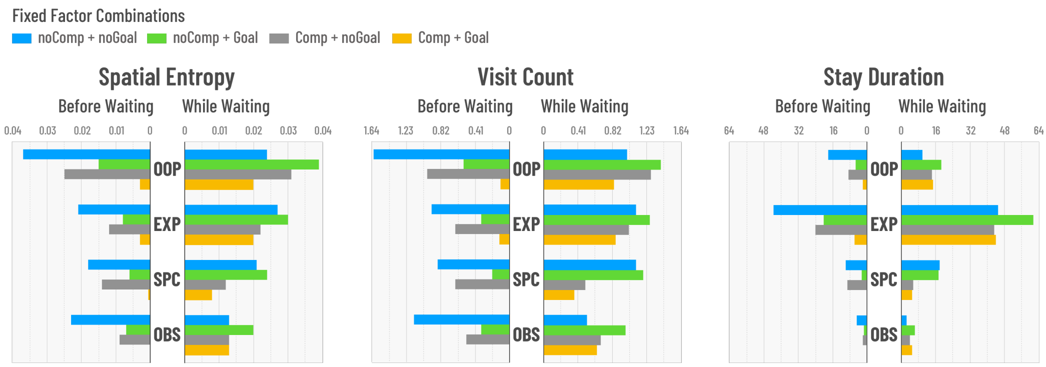

Fig. 5. Graphs showing mean measures of spatial entropy (left), visit count (middle), and stay duration (right) of the four PIR sets: out-of-place elements (OOP), extreme points (EXP), spatial connections (SPC), and visual obstructions (OBS). For each of the three graphs, bars extend left to illustrate measures before waiting, and right for while waiting. Color coding of the bars indicates the combination of compensation and goal condition for a measure.

Bayesian repeated measures ANOVA tests across PIR sets show decisive evidence ($BF_{10} > 1k$) for differences in PIR sets beyond the impact of subject factors (goal statement, environment aesthetic, or compensation). Figure 5 shows visual graphs of PIR set measures for different fixed factor combinations, as well as for 'before' and 'while' waiting. ANOVA tests show strong evidence that the 'environment' factor has no effect on these measures, and is thus not included in the figure. Table 4 shows the ANOVA results, with the last column showing the order of means of individual PIR sets. In some cases, PIR sets are considered statistically equal (e.g. PIR spatial entropy before waiting: OOP is higher, but EXP, OBS, and SPC are considered equal), even if differences seem to exist according to the means (as shown in Figure 5). In such cases, between-subject factors such as goal and compensation are a likely cause, according to the best model. For the aforementioned example, goal, compensation, as well as an interaction effect between PIR sets and goal statement are impacting the measure. For interpretation of differences between PIR sets irregardless of other factors, 'PIRs Post Hoc comparisons' in Table 4 displays the most probable results.

It should be noted that patterns do not 'perform' uniformly. Results of PIR sets should therefore be understood as including some 'performance' bias by individual pattern implementations. Table 5 lists stay durations and visit counts for individual PIRs of the four PIR sets. 'Ground Fog' stands out as having been visited 3 times more often than other patterns in OBS. 'Mountain A' stands out in EXP for fewer visits and shorter stay durations.

6.4 Notable Correlations

Bayesian Pearson correlations were calculated to provide context for measures for which a statistical correlation was not certain (e.g., path distance and spatial entropy are logically correlated) and useful for the study. PIR spatial entropy decisively correlates with PIR visits and stay durations ($BF > 1k$) and is thus used as indication to assess correlations between PIR sets and other metrics.

Before waiting: PIR sets. All PIR sets correlate positively with Glasgow SDs of all emotion dimensions ($BF > 1k$, highest correlation for OOP), but not with Glasgow means. They positively correlate with participants going 'outofbounds' (i.e., jumping into the chasm) for OOP, SPC, and OBS (all $BF > 1k$, highest correlation for OOP), but not EXP. Camera movements positively correlate with all PIR

Dependent Variable	Best Model (in order of inclusion probability)	BF_M	PIRs in order of means
Before Waiting			
PIR Spatial Entropy	PIRs + Goal^{1k} + Comp^{1k} + [PIRs × Goal]	40	OOP > EXP = OBS = SPC
Stay Duration	PIRs + Goal^{1k} + [PIRs × Goal] + Comp^{10} + [PIRs × Comp]	115	EXP > OOP ≈ SPC > OBS
Visit Counts	Goal^{1k} + PIRs + Comp^{1k}	27	OOP > EXP = OBS = SPC
While Waiting			
PIR Spatial Entropy	PIRs + Comp^{1k}	41	OOP ≈ EXP > OBS = SPC
Stay Duration	PIRs	18	EXP > OOP ≈ SPC > OBS
Visit Counts	PIRs + Comp^{100} + [PIRs × Env] + [PIRs × Comp] + Env	27	OOP = EXP > OBS = SPC

Table 4. PIR Sets repeated measures ANOVA results: Best models of Bayesian ANOVA for dependent variables, if model Bayes factor $BF_M > 3$. The top half of the table lists results for 'before waiting', the lower half for 'while waiting'. Individual fixed factors in the model are sorted in descending order by probability of inclusion in the model (BF_{incl}, see supplementary material for values). Post Hoc T-Test results (BF_{10}) are included as superscripts if at least $BF_{10} > 3$, reported in steps: >3, >10, >100, >1k (1000). Interaction effects (e.g. [Goal × Pat]) do not have an associated BF_{10}. Post Hoc comparison for individual PIR sets are listed in the last column and sorted by means. Fixed factor abbreviations are: Pattern (Pat), Compensation (Comp), and Environment (Env).

Design pattern	visited by (n)	$mean_{visit}$	SD_{visit}	$mean_{\%stay}$	$SD_{\%stay}$
(OOP) Stone Stack A	42.9% (79)	1.5	0.8	1.74	1.99
(OOP) Stone Stack B	44.6% (82)	1.4	0.7	1.45	2.03
(OOP) Stone Spiral	45.7% (84)	1.3	0.6	1.76	1.80
(EXP) Mountain B	48.4% (89)	1.3	0.6	7.08	4.10
(EXP) Mountain C	44.0% (81)	1.2	0.5	4.94	4.71
(EXP) Mountain A	31.0% (57)	1.1	0.3	3.81	3.06
(SPC) Cliff Cave	28.8% (53)	1.6	0.8	2.83	2.01
(SPC) Hill Path	29.3% (54)	1.2	0.5	1.14	0.91
(SPC) Mountain Cave	24.5% (45)	1.1	0.3	1.29	1.57
(OBS) Ground Fog	49.5% (91)	1.4	0.6	0.73	0.60
(OBS) Forest B	18.5% (34)	1.2	0.5	0.79	0.90
(OBS) Forest A	19.6% (36)	1.0	0.2	0.65	0.61

Table 5. Visit counts and overall stay duration for individual PIRs, sorted by PIR sets and in order of mean visit count across participants. Visits are individual instances of at least > 1 sec time spent in a PIR. Stays are listed in percent of a player's session length.

sets entropies (BF > 1k for all but SPC, which had BF > 30); here the highest correlation was with EXP. PIR sets do not correlate with game frequency or player type.

While waiting: PIR sets. Positive correlation with game frequency and player type for EXP (BF > 30), OBS (BF > 30), and SPC (BF > 3); but not OOP. Glasgow measures do not correlate, except for EXP and arousal SD (BF > 3). Positive correlation with 'outofbounds' events for OBS (BF > 3) and EXP (BF > 1k). Camera movements positively correlate for EXP and SPC (both BF > 1k).

Non-correlations. Some correlation results are notable for their *lack* of correlation with other measures. 5DC measures showed evidence for a lack of correlation with game metrics, in-game curiosity ratings, or Glasgow ratings. The only exception is the 'thrill seeking' dimension, which

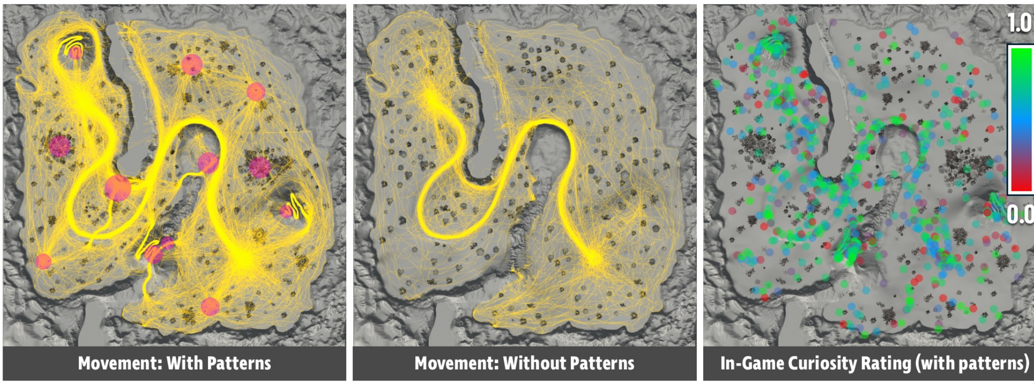

Fig. 6. Player movement paths for the 'with patterns' (left, includes pattern locations) and 'without patterns' (middle) experiment factors; and distribution of in-game curiosity ratings (right).

correlated with variation in camera rotation before waiting (BF > 10). PIR set measures did not correlate with age, GUESS ratings, or in-game curiosity ratings.

6.5 Nuisance Variables

Play direction was recorded as a nuisance variable, but evaluated through a Bayesian Student T-Test to track its impact on measures. Most measures did not differ by play direction. Of note are path distance mean (BF > 100) and SD (BF > 1k), both especially 'while waiting' (BF ≈ 1k), and differences in the OOP 'while waiting' measures visit count (BF> 1k), stay duration (BF > 3), and PIR spatial entropy (BF > 1k). Here, the proximity of 'Stone Stack A' to Zone A likely provided a stronger attractor than an equivalent OOP pattern for players heading towards Zone B.

Participant gender was reduced to female and male participants to assess the impact on measures through a T-Test. Measures impacted by gender differences are likely affected by a difference in 'gamer type' (BF > 1k) and 'play frequency' (BF > 3), with female players having lower measures in both due to differences in sampling distributions. GUESS measures Engrossment, Enjoyment, and Personal Gratification differ (all BF > 30, all higher for female). In-game curiosity ratings are increased for female players (BF > 100), as are Glasgow ratings for arousal (BF > 3) and valence (BF > 3) while waiting. Female players had shorter (BF > 30) and fewer (BF > 10) visits to SPC PIRs while waiting, as well as fewer visits (BF > 3) to EXP PIRs while waiting. Finally, female players moved slower (BF > 30) and had fewer camera movements (BF > 100).

7 DISCUSSION

The primary goal of this study is to examine the effect of level design patterns for spatial exploration on player behavior and experience. Across multiple measures, results show that the presence of patterns indeed influenced how players interacted with the environment, and that patterns had an emotional impact. However, how exactly players were influenced, depended on other factors. While results suggest that *environment aesthetic* has little to no impact, having a *a goal* and being *compensated* evidently affects exploration. In some circumstances, the relative impact indeed exceeds that of pattern presence.

It is important to discuss player behavior over two phases of the game: *before waiting* for the Master, and *while waiting* for the Master to stop meditating. It can be hypothesized that a player's motivation for exploration and their emotional experience shifted at this point, although how

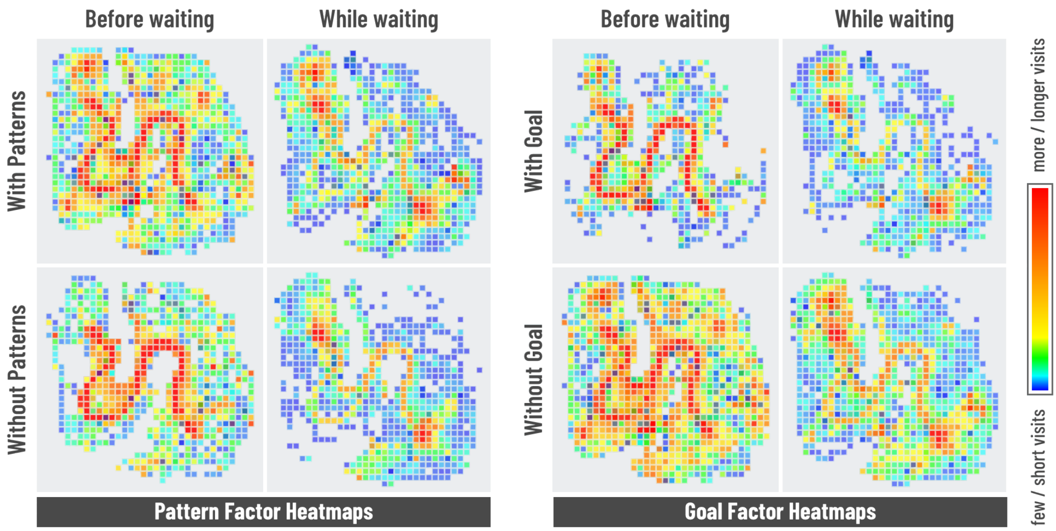

Fig. 7. Heat-maps of player presence split out by the experimental factors of goal statement and pattern presence.

exactly depends on the condition group. As such, the following sections discuss differences between these two phases of the game.

We controlled for two nuisance variables: play direction and gender. Play direction had some impact on player behavior, as players generally seemed to explore more while waiting when they ended in Zone A. Although the map was designed to be reversible, Zone A has an 'out-of-place' element in close proximity, whereas Zone B features a 'visual obstruction' element. Based on the popularity of patterns from this set, 'Stone Stack A' likely provided a stronger motivation for exploration, whereas Zone B had fewer interesting PIRs in close range. Aside of a preference for visiting 'Stone Stack A', however, the overall impact appears limited. In addition to play direction, the demographic of players had some impact as well. Female players generally had higher emotional investment in the game, and also scored their game experience higher (GUESS). Neither of these impacted the results in a way that make them specific to players based on their gender.

7.1 Impact of Patterns

In general, level design patterns caused participants to venture further away from the path and further away from their destination (i.e., the Master), resulting in movement across the environment that was overall more dispersed. These differences in exploratory behavior are visible in the visualizations of player presence (see Figure 6 and Figure 7) and confirmed by the statistical analysis. Although these results come with certain caveats (discussed in the following sections), patterns clearly had an effect on exploratory behavior. As such, we consider **H1a to be supported**. The impact of individual patterns is discussed in more detail later in this section.

In addition to affecting behavior, we hypothesized that the patterns would also positively affect players' emotional experience. This was measured in two ways: through the in-game use of 'emotion words' and the post-game measures. Unfortunately, the GUESS showed little to no differences between conditions. We do not necessarily consider this a fault of the GUESS, but possibly a side effect of our study design. Due to the decision to make participants wait for five minutes in an attempt to gather data under different circumstances, it is likely that players grew bored. This is supported by Glasgow ratings of the in-game reported ‘emotion words’, which saw a decrease in valence, dominance, and arousal while waiting as compared to before waiting. We suggest that the last five minutes colored participants’ overall experience, eliminating any differences the GUESS might have uncovered. However, we do see a difference in comments that participants made — those who played with patterns seem more likely to comment positively about the game overall.

We also see an impact of patterns on the in-game Glasgow ratings derived from emotion words. This effect, however, is primarily visible in the spread (i.e., standard deviation) of the ratings. Before waiting, Glasgow means are not affected. Impacts on spread, on the other hand, were primarily due to the goal condition and only to a lesser extent due to patterns. While waiting, patterns had a small impact on Glasgow means, which happened in combination with compensation. The spread in dominance and valence was impacted by patterns as well, in combination with the goal condition.

Overall, the results suggest that patterns had an impact on the range of emotions expressed through the in-game ratings, with the emotional range increasing in the presence of patterns. Before waiting, we argue that a lack of patterns did not necessarily result in a more subdued emotional experience. Participants that cared to explore likely did so, while those that did not care to explore were driven by finding out what they were supposed to do, rather than being negatively impacted by a lack of patterns. However, the presence of patterns likely leads participants to explore, or at least makes emotional impact.

While waiting, the presence of patterns is responsible for higher arousal and valence. This happens in combination with compensation, suggesting that participants that rushed to the stated goal (more likely driven by extrinsic, financial motivation) found elements that interested them while waiting as compared to those playing without patterns. It is not clear, however, whether the impact was due to exploratory curiosity, or the desire to alleviate boredom. Participant comments suggest a mix of motivations, e.g., left path to explore (55.9%), because having to wait (31.1%), to explore a landmark (17.7%), or due to boredom (12.6%). The spread of emotional values was increased due to patterns while waiting as well, suggesting more ‘ups and downs’ in the play experience.

Having patterns elicited more comments from participants, suggesting that participants with patterns felt more strongly about their experience and the effort they put into the study. Although they more commonly commented on leaving the path to explore, they were also more likely to mention the lack of interaction in the game. We hypothesize that the presence of patterns created expectations. The relative novelty of encountering PIRs was not perceived as a reward in itself. This is likely why emotional impact was more tied to fluctuations rather than an overall increase — players probably enjoyed the moments when they were engaging in exploration, but may have experienced disappointment when their efforts went unrecognized by the game system (either through an explicit reward, or by encountering interactive content).

Based on these findings, we consider that **support for H1b exists**, albeit with some annotations. The presence of PIRs alone is not sufficient for increasing emotional investment. **Instead, they afford a possibility for exploration that, when realized, increases emotional investment.**

7.2 Impact of Goal Statement

Whether or not players were given a goal had a substantial impact on the effects of patterns. Before waiting for the Master, among participants without a goal, those with patterns ventured further from the path than those without. They were also further away from the destination point (i.e., Master), i.e., they moved around more as they closed the distance to the destination. Entropy measures also confirm this behavior.

These differences were severely reduced, however, when a goal was introduced. When given a goal, the presence of patterns had a much lower impact on any of the measures. Although we hypothesized that having a goal would reduce exploration, the magnitude of the difference is surprising. The formulation of the goal is not very specific and relatively subtle — it only informs the player they are in search of their master and that he awaits them. In addition to this, there are a few signs along the path pointing in the Master's direction. The path itself, however, is present in all conditions. We argue that the path always hints at a goal, as it is generally understood that paths lead to somewhere of interest, so it is not unreasonable to suspect they would follow the path as one of their first actions. Yet it is likely that players not provided with a goal spent the first part of the game more in search of what to do. Overall, in the presence of patterns, players without a goal were still more likely to leave the path behind and explore the boundaries of the environment compared to having no patterns present. A possibility is that PIRs are of interest to players irrespective of whether or not a goal was provided, but that part of the difference is masked by participants without a provided goal trying to figure out what to do.

While waiting, the data shows quite different behavior. Generally, players with patterns showed more exploratory behavior than those without. Particularly interesting, however, is that participants with a goal generally explored more than those without. We suggest that this is because participants with a goal explored *less* before finding the Master. As such, they were focused on accomplishing their goal when the game started. Once the game provided them with a new goal (i.e., wait for the Master to finish meditating), they felt free to explore. At this point, players with patterns explored more than those without. Players that had already explored before because they did not have a goal, did not feel the need to explore as much once they had to wait. Based on these findings, we consider **H2 to be supported**.

When we consider emotional impact, we can see that players without a goal had higher fluctuations in their emotional experience. However, this is only true before waiting for the Master. While waiting, the opposite is true: affect fluctuates more with a goal. We propose that being uncertain about the goal of the game creates more potential for emotional investment, as players take it upon themselves to find out what the game is about. As the other data suggests as well, players with a goal were likely more focused on achieving it. Once they are asked to wait, they are presented with a new situation. Meanwhile, players that had no goal are actually given one while waiting. While fluctuations in emotional investment do not necessarily indicate that players enjoy the experience more overall, the concept of designing for interest curves or experiential fluctuations in a game (e.g. [61]) is understood as a strategy to increase overall interest. The results show that having a goal had little impact on the affect means.

Despite our findings, it is not necessarily the case that a goal reduces exploration in all situations. In part, our results may be due to the nature of the experiment. While players may prioritize a goal or 'quest' when playing a commercial game, it is also not unlikely for them to abandon it in favor of freely exploring an (open) environment when presented with one — likely, this depends a lot on the player and their play style. We suggest the possibility that, due to participants knowing they were participating in a study, they were expecting to receive instructions on what they were supposed to accomplish. As such, those who received a goal prioritized it, thinking it would be necessary for them to complete it in order to successfully finish the study. Those who did not receive a goal, could have been motivated by a search of what they were supposed to do. As such, it could be said that their motivation to explore was not strictly one of curiosity, but rather one of seeking a purpose in the context of the study. If this was their only motivation, however, we consider it more likely that players would have stuck to the path, as it gave an implicit indication of where such a purpose might be found. Since this is not the case, we can argue that patterns still had an impact and that players were not solely motivated by their desire to complete the experiment.

7.3 Impact of Compensation

As stated before, we considered it possible that the addition of an extrinsic reward (i.e., monetary compensation) could influence the intrinsically motivated behavior of exploration. In addition to players looking for what they were supposed to do to finish the study, the addition of an (assured) monetary reward could have motivated players to finish the study as quickly as possible. If this were the case, we would expect them to put in minimal effort, spending as little time as possible and exploring only to a minimal extent.

Before waiting, whether or not people were compensated had limited effect. We see some differences in how far participants ventured from the path, although patterns and goal have a stronger influence on this behavior. While waiting, however, compensation was more likely to have an influence on measured behavior. It was the most likely measure for several dependent variables, including spatial entropy, distance from path, and distance from destination. Participants were also less likely to see what would happen if they jump into the chasm. Overall, this shows that participants who were assured compensation were less likely to travel away from the end point once they had to wait, suggesting they wanted to be able to get to the ending as quickly as they could.

Regarding emotional experience, compensation had little influence of note. However, it was the largest contributor to differences in valence means while waiting. Interestingly, valence was higher on average for people who were being compensated (note that this is the only measure that was notably higher among people who were being compensated). We interpret this to mean that participants who were being compensated were more pleased knowing that they would progress the experiment within a certain time. We suspect that those who were not compensated (and especially those without a goal who had already spent more time exploring) were more annoyed at being made to wait. The presence of patterns, in turn, mitigated this somewhat.

7.4 Analysis of Patterns

Besides examining the overall impacts of patterns, we were interested in the effects of the individual patterns themselves. To this end, we analyzed the data from only those participants who had patterns in their game environment, which was also the rationale for randomizing with a 70% chance for having patterns present.

Before waiting, we can observe that participants without a goal or compensation interacted with patterns the most; they tended to visit PIRs and stayed for a while. This is contrasted with participants that have both a goal and are being compensated, who barely visited any PIRs and did not stay long at those they did visit, suggesting they were trying to get to the Master as quickly as possible. Out of the remainder, participants *without a goal that were compensated* visited more PIRs and stayed longer than those *with a goal who were not compensated*. We take this to mean that players without a goal were searching for one, while the compensation drove them forward. Those who were not being compensated took more time to explore the PIRs, even if they had a sense of where to go.

While waiting, we see a lot more interaction with patterns than before. Overall, participants visited more PIRs and stayed longer while waiting. Since they knew that they had time to spend at this point, they took more time to look around. Participants with a goal who were not compensated visited the most PIRs and stayed the longest. We consider this in line with previously discussed findings that those with a goal spent less time to get to the Master and then used the waiting time to explore. Similarly, interaction with the patterns remained lowest among participants with a goal who were compensated — although they were also likely to get to the Master early, they were less willing to venture away from the destination to explore the patterns.

Out of Place (OOP). PIRs of this pattern attracted the most visitors, with resulting visits being rather short. Spatial entropy, however, was generally high. This suggests that this pattern causes local exploration to occur, where an object of interest is being examined from multiple angles. We suggest that this could either be to examine the aesthetic of the object itself, or with the intention of finding out what function it could serve. In many games, OOP elements are used to guide players to rewards (e.g., collectibles and upgrades for the player character), opportunities for interaction, or progress in the game narrative. In such cases, finding out what can be done at these elements is a game mechanic in itself (e.g., requiring players to perform specific actions to make progress). Continued engagement could be the desire to gather more 'clues' as to what these elements are about. In addition to this, all patterns except OOP were more attractive to players with higher game experience. We take this to suggest that PIRs of this pattern have universal appeal, regardless of gaming experience, while others attract only more experienced players.

Extreme Points (EXP). EXP PIRs were visited by many players, and caused them to stay longer than other patterns. One reason for this could be that players use the higher vantage points to get the 'lay of the land'. It allows players to 'visually explore' the environment, i.e., gain an understanding of it without having to travel there. This is perhaps especially the case for players without a goal, who stay the longest at EXP PIRs before waiting. While waiting, stay durations are fairly similar regardless of fixed factors. It could be that reaching these patterns is challenging and, as a result, succeeding in that challenge can feel rewarding in itself, causing players to take a moment to appreciate the result of their effort. The EXP pattern correlating with the spread of arousal ratings suggests that interaction with the pattern has moments of varying excitement. Reaching a high place also allows players to see their surroundings from a new and interesting perspective. As such, participants may stay there for aesthetic reasons (i.e., enjoy the view), as is indicated by a correlation with increased camera rotation. Even if the game does not provide any specific interactivity, surveying the environment from a vantage point can be experienced as an engaging activity. Out of all patterns, EXP PIRs were most often mentioned by participants as reasons to leave the path or as outstanding features.

Spatial Connections (SPC). PIRs in this pattern were not visited as often as EXP or OOP patterns, and players stayed for short time periods. Players did stay longer at SPC PIRs than they did at OBS PIRs. The 'Cliff Cave' had the longest stay duration. Compared to the others, it offered a unique vantage point of the environment (i.e., from within the chasm). We theorize that providing an interesting view contributes to the appreciation of a PIR, as indicated by participants looking around more — neither of the other SPC PIRs provided a 'better' view than the mountains did. It is possible that understanding how spaces connect can offer an intrinsic reward similar to that of exploring EXPs. Out of all PIRs, caves were commented on fairly frequently (nearly as much as mountains). This suggests that they stood out to participants and left an impression. However, this doesn't translate into as many visits. It could be that less people noticed the SPC PIRs to begin with. Additionally, it may be because their entrances are more hidden (while mountains were easily available), and therefore the effort to figure them out was too high for some. Games tend to implement this pattern not as a way to guide players, but to present them with a challenge that often involves a reward. The fact that players looked around more when at these PIRs may also suggest they were looking for something. Since there was nothing to find (a fact commented on often by participants), it is possible players did not feel the need to bother with figuring out these PIRs once the lack of reward had become apparent.

Visual Obstructions (OBS). OBS PIRs were visited the least of all patterns (with one exception), and players stayed the shortest. Considering the measure of spatial entropy, it would also appear that players primarily ran through these PIRs, possibly without even registering them. 'Ground Fog' stands out as an exception, as it was visited by more players than any other individual PIR. It was also fairly often commented on by participants after the game. It is possible that the two dense forests were not understood as places for potential exploration, but rather natural boundaries. Of course, it is also possible that the potential was understood but simply was not appealing. A challenge in implementing this pattern successfully stems from the fact that visual obstructions must still appear to be surmountable. Games often enclose the interactive space with environmental obstructions that signal that they cannot be overcome (which is also the case for the game in this study). OBS patterns can easily be misread as areas that cannot be explored. As a result, games tend to implement this pattern more often for 'secrets' (i.e., additional content that is designed to be encountered by a small selection of players), and not to guide player progress. We consider this the reason that this pattern was more likely to be visited by players with higher game experience. Based on having played other games, they were possibly better equipped to recognize the PIRs as potentially interesting. Alternatively, it could also be that less experienced players avoided them for the possibility of danger as their vision was obscured.

7.5 Impact of Trait Curiosity

For **H3**, we stated that the general disposition for curiosity (i.e. 'trait curiosity') would have an impact on exploratory behavior. Our results show that curiosity dimensions did not correlate with exploratory behavior measures (with the exception of camera rotation with 'thrill seeking') or emotional experience. Whether or not a player experienced curiosity for exploration in the game did not seem to be impacted by their general disposition. It is possible that the threshold for engaging in exploratory behavior in a game like this is relatively low; or that measures of trait curiosity in the physical world do not correspond to game environments. Based on these results, we **reject H3.**

7.6 Measures of Exploration

Measures in this study involved validated psychometric instruments (i.e., GUESS, 5DC), game metrics, and exploratory in-game measures that have yet to prove their viability in game user research studies. Especially the use of 'in the moment' measures of emotional states through a combination of a curiosity scale and interpreting the affect of emotion words has, to our knowledge, not been described to measure exploration in games before. Our results suggest that in-game curiosity ratings showed meaningful correlations with how players assess their game experience. However, these ratings also showed a wide variance across players, suggesting that unexplored factors contribute to being in a state of curiosity. One possible weakness of the measure is that it was gathered at fixed points in time instead of taking measure at specific player actions or locations in the game.

The interpretation of emotion words provided a stronger basis for evaluating affective states that accompany exploration than the curiosity scale did. We hypothesize that 'curiosity' may be a short state that is more difficult to self-assess than the use of emotion words. Ultimately, measuring curiosity directly remains a challenging task — it is either determined by indirect measures, or measured by interventions that can impact the measure by their mere presence. Nevertheless, the results of this study provide evidence that a combination of behavioral and affective measures provides insights about curiosity for exploration.

8 LIMITATIONS

The results of any study should be considered within the limitations of its design. Our primary goal was to examine the effects of level design patterns for spatial exploration, as well as the presence of a goal and its impact on those patterns.

As stated previously, games are complex systems with many interacting elements. With the design of *Shinobi Valley*, one of the goals was to strike a balance between controlling for confounding factors, while still giving players the experience of playing a 'proper' game. However, this meant that the game lacked many characteristics typical of the type of commercial game its design is based on. While this was a deliberate decision, level design patterns are not usually used in isolation. Instead, they are used to guide players towards specific objects, objectives or interaction possibilities. As such, they likely raise expectations in players that their exploratory behavior will somehow be acknowledged or rewarded. This is shown in our qualitative data, which suggests that many players were expecting to find something as the result of their exploration. While exploratory behavior probably was not impacted (due to the short duration of the game), emotional investment likely was because the game did not provide the satisfaction that 'real' games do. For this study, we consider this acceptable, as the intention was to investigate curiosity driven exploration in particular. However, future studies should carefully consider the implementation of rewards (*if* it benefits their purposes), especially when player experience is the main focus of study.

In relation to this, we point out the influence of receiving compensation for participating in a research experiment. It should be noted that our participants who did not receive direct financial compensation, still could receive some compensation by way of a random draw. Although there was a measurable difference between these participants and those who were assured compensation, suggesting this had a more direct influence on player behavior, we cannot fully exclude the possibility that a potential extrinsic reward had some effect as well.

The patterns tested in this study were formulated based on the design of 3D, open world games. As such, the results of this study cannot be generalized to all types of games. However, we argue that the patterns themselves can be implemented in different types of virtual environments (e.g., smaller, 'closed' levels) and other game forms. It is, for example, possible to have hard-to-reach places or out-of-place objects in a 2D game as well. Their implementation will require careful thought, however, and whether or not they are experienced similarly as they were in this study remains to be investigated.

Even a game as simple as *Shinobi Valley* adds layers of complexity to empirical assessment. Our decision to include additional variables, rather than only focusing on the presence or absence of level design patterns, complicated the presented experiment considerably. Testing the game with only one independent variable would have been easier, but also miss important findings related to interaction affects with, for example, a stated goal. As discussed previously, level design patterns do not occur in a vacuum. Therefore, we decided it was important to include at least some aspects very integral to many games, rather than draw conclusions from a very selective experiment. In turn, however, this makes the gathered results more complex to interpret. We argue that this is inherent to disentangling player experience and that further studies should be aware of this and make their own informed decisions.

We also decided to have more participants play with patterns present to increase the sample size and be able to assess their individual impact. Naturally, this skewed the sample sizes of our other condition groups. To run ANOVA tests between condition groups, it is generally recommended for the groups to be of similar size. Instead, some of our condition groups ended with relatively few participants due to this decision to focus our available resources. For example, the smallest condition group (no pattern, goal, nature, no compensation) had 5 participants, while the largest group (pattern, goal, alien, compensation) had 32 (mean = 16, SD = 8.9). Because of this difference, any interaction effects detected in the data lend themselves to further study with more participants and should be considered for future work.

Another consideration in our study design was the decision to make participants wait after finding the Master. Although this decision provided interesting findings relating to how the stated goal and level design patterns interacted, we cannot say that exploration while waiting happened solely due to curiosity and was not, at least in part, motivated by boredom as well. As stated previously, curiosity is difficult to capture. Although our efforts uncovered interesting results, we cannot fully interpret the motivations of players. Likely, they are multi-dimensional. More thorough in-game measures (e.g., observation, questions, think aloud protocol) might assist with unpacking the player experience further, but a balance always needs to be struck between invasiveness of the method and the result.

Finally, we also administered the GUESS after participants had completed the entire game. This meant that players filled out the survey after possibly spending the final minutes of the game annoyed or bored. As a result, the GUESS was likely influenced by these final minutes and not particularly useful in assessing differences between conditions. Although the in-game measures offset this, these do not assess game experience as thoroughly as the GUESS does. While the decision to make players wait provided us with interesting data, the use of any post-game questionnaire should be considered carefully if a similar design is used in future studies. Even if the study design is different, however, we urge caution in relying on post-game measures in capturing temporary states, such as curiosity.

9 CONCLUSION

In this study, we found empirical evidence for level design patterns eliciting curiosity-driven exploration in players with the impact being affected by having an explicit goal statement, and whether compensation is provided. Without patterns, players engaged in less exploration, and formed fewer expectations about being rewarded for doing so. Being given a goal was shown to strongly impact players' likelihood of exploration. Players were most engaged in curiosity-driven exploration when patterns in the environment provided opportunities, and the goal of the game was left sufficiently ambiguous to pay attention to the larger environment. Comments of players also show that exploration in games is understood as a mechanic in itself. It can be concluded that the dramatic principle of Chekhov's gun in literature also applies to elements that invoke exploration: if something promises to be an interesting area for exploration, it should provide acknowledgment to players when they do so. This is already a common practice in game development, and the exclusion of rewards in this study was primarily motivated by reducing confounding experiment variables.

Based on the experiment results, whether or not players explore was not impacted by their general disposition for curiosity. This could mean that the threshold for developing curiosity was not very high in the game experiment, or that general disposition is not a strong predictor for a curious state in a video game. We found evidence that exploration motivated by boredom differed from curiosity-driven exploration. Additionally, curiosity-driven exploration can have different motivations behind it, such as looking for rewards or interaction possibilities, or testing the boundaries of the environment. Design patterns can only increase the likelihood of curiosity, not enforce it. By controlling for environment aesthetic, we consider our findings applicable not just to a specific implementation, but to other game environments as well. Finally, whether or not participants were compensated also affected exploratory behavior and should be considered especially in study designs looking into player behavior and experience.

With our work, we have presented empirical evidence for the overall efficacy of a design practice that is already part of many video games, and is clearly having a demonstrable impact. Based on analysis of in- and post-game measures we have further created a framework for understanding the impact of individual patterns, and for mapping their efficacy in light of relevant factors, such as having a stated goal, and free exploration versus waiting time. Through a study design that incorporates and evaluates various elements common to games, we show the many complexities that result from the interaction between such elements. Although this results in a more complex narrative, we believe that this account can inform future empirical studies of player experience and shows the need to explore variables whose influence may otherwise be ignored. We do not suggest that every study incorporates a variable on whether or not players have a goal — rather, we encourage designers and researchers to consider the effects *all* their decisions have on players, especially when studying player experience and complex constructs.

Future work may expand on the lexicon of design patterns that invoke curiosity for exploration. Given the vast design space of creating game worlds, more data from different implementations of the discussed patterns is needed to either support or scrutinize our findings. The promise of engaging in this work is a better theoretical understanding for how to intentionally design for curiosity-driven exploration in games. In time, we foresee that such work can also support efforts for better procedural creation of game content, or even real-world implementations of explorable architecture, such as the design of playgrounds or amusement parks.

ACKNOWLEDGMENTS

We would like to thank all participants for the time and effort they invested in this study, as well as our reviewers for their extensive feedback and helpful suggestions.

REFERENCES

[1] Linda Altieri, Daniela Cocchi, and Giulia Roli. 2019. Advances in spatial entropy measures. Stochastic Environmental Research and Risk Assessment 33, 4-6 (May 2019), 1223–1240. https://doi.org/10.1007/s00477-019-01686-y
[2] Amazon. 2021. Amazon Mechanical Turk. https://www.mturk.com/
[3] Yoav Bar-Anan, Timothy D. Wilson, and Daniel T. Gilbert. 2009. The Feeling of Uncertainty Intensifies Affective Reactions. Emotion 9, 1 (2009), 123–127. https://doi.org/10.1037/a0014607
[4] Chris Bateman. 2016. The Aesthetic Motives of Play. In Emotion in Games. Springer, 3–20.
[5] Daniel E Berlyne. 1954. A Theory of Human Curiosity. British Journal of Psychology 45, 3 (1954), 180–191.
[6] Daniel E Berlyne. 1960. Conflict, Arousal, and Curiosity. McGraw-Hill Book Company.
[7] Bethesda Game Studios. 2011. The Elder Scrolls V: Skyrim. Game [Windows, Consoles]. (Nov. 2011).
[8] Staffan Björk and Jussi Holopainen. 2006. Games and Design Patterns. The game design reader (2006), 410–437.
[9] Blizzard Entertainment Inc. 2004. World of Warcraft. [Macintosh, Windows]. (Nov. 2004).
[10] Julia Ayumi Bopp, Elisa D. Mekler, and Klaus Opwis. 2016. Negative Emotion, Positive Experience?: Emotionally Moving Moments in Digital Games. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. ACM, San Jose California USA, 2996–3006. https://doi.org/10.1145/2858036.2858227
[11] Greg Costikyan. 2013. Uncertainty in games. Mit Press.
[12] Nigel Cross. 2001. Designerly Ways of Knowing: Design Discipline versus Design Science. Design issues 17, 3 (2001), 49–55.
[13] Hy I Day et al. 1971. Intrinsic motivation: A new direction in education. (1971).
[14] Edward L Deci and Richard M Ryan. 2010. Intrinsic motivation. The corsini encyclopedia of psychology (2010), 1–2.
[15] John Dewey. 1910. How We Think. Heath & Co.
[16] Robert Eisenberger, W David Pierce, and Judy Cameron. 1999. Effects of reward on intrinsic motivation—Negative, neutral, and positive: Comment on Deci, Koestner, and Ryan (1999). (1999).
[17] Christelle Gillioz, Johnny R. J. Fontaine, Cristina Soriano, and Klaus R. Scherer. 2016. Mapping Emotion Terms into Affective Space. Swiss Journal of Psychology 75, 3 (June 2016), 141–148. https://doi.org/10.1024/1421-0185/a000180
[18] Marcello A Gómez-Maureira and Isabelle Kniestedt. 2019. Exploring video games that invoke curiosity. Entertainment Computing 32 (2019), 100320.
[19] Marcello A Gómez Maureira, Isabelle Kniestedt, Max J Van Duijn, Carolien Rieffe, and Aske Plaat. 2019. Shinobi Valley: Studying curiosity for virtual spatial exploration through a video game. In Extended Abstracts of the Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts. 421–428.
[20] Marcello A Gómez-Maureira, Michelle Westerlaken, Dirk P Janssen, Stefano Gualeni, and Licia Calvi. 2014. Improving level design through game user research: A comparison of methodologies. Entertainment Computing 5, 4 (2014), 463–473.
[21] Emily M Grossnickle. 2016. Disentangling Curiosity: Dimensionality, Definitions, and Distinctions From Interest in Educational Contexts. Educational Psychology Review 28, 1 (2016), 23–60.

[22] John Harris. 2007. Gamasutra - Game Design Essentials: 20 Open World Games. https://www.gamasutra.com/view/feature/1902/game_design_essentials_20_open_.php

[23] Toby E. Huff. 2011. Intellectual Curiosity and the Scientific Revolution: A Global Perspective. Cambridge University Press, Cambridge ; New York.

[24] Kenneth Hullett and Jim Whitehead. 2010. Design Patterns in FPS Levels. In Proceedings of the Fifth International Conference on the Foundations of Digital Games - FDG '10. ACM Press, Monterey, California, 78–85. https://doi.org/10.1145/1822348.1822359

[25] Ilhan Inan. 2013. The philosophy of curiosity. Routledge.

[26] JASP Team. 2020. JASP (Version 0.14.1)[Computer software]. https://jasp-stats.org/

[27] Harold Jeffreys. 1961. Theory of probability (3rd ed.). Oxford University Press, Oxford.

[28] Hugh Jeremy. 2015. Using a real-time feedback system to make better games. [Gamasutra GDC Vault]. https://www.gamasutra.com/view/news/244257/Video_Using_a_realtime_feedback_system_to_make_better_games.php Accessed: 2021-04-15.

[29] Todd B Kashdan and John E Roberts. 2004. Trait and state curiosity in the genesis of intimacy: Differentiation from related constructs. Journal of Social and Clinical Psychology 23, 6 (2004), 792.

[30] Todd B. Kashdan, Melissa C. Stiksma, David J. Disabato, Patrick E. McKnight, John Bekier, Joel Kaji, and Rachel Lazarus. 2018. The Five-Dimensional Curiosity Scale: Capturing the Bandwidth of Curiosity and Identifying Four Unique Subgroups of Curious People. Journal of Research in Personality 73 (April 2018), 130–149. https://doi.org/10.1016/j.jrp.2017.11.011

[31] Heidi Keller, Klaus Schneider, and Bruce Henderson. 2012. Curiosity and Exploration. Springer Science & Business Media.

[32] A. Khalifa, F. de Mesentier Silva, and J. Togelius. 2019. Level Design Patterns in 2D Games. In 2019 IEEE Conference on Games (CoG). 1–8. https://doi.org/10.1109/CIG.2019.8847953

[33] Christoph Klimmt. 2003. Dimensions and determinants of the enjoyment of playing digital games: A three-level model. In Level up: Digital games research conference. 246–257.

[34] Schulamith Kreitler, Edward Zigler, and Hans Kreitler. 1975. The nature of curiosity in children. Journal of school psychology 13, 3 (1975), 185–200.

[35] 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.

[36] Craig Lindley, Lennart Nacke, and Charlotte Sennersten. 2008. Dissecting Play–Investigating the Cognitive and Emotional Motivations and Affects of Computer Gameplay. In 13th International Conference on Computer Games (CGames 2008). UNIV WOLVERHAMPTON.

[37] Jordan Litman. 2005. Curiosity and the pleasures of learning: Wanting and liking new information. Cognition & emotion 19, 6 (2005), 793–814.

[38] Jordan A Litman. 2008. Interest and deprivation factors of epistemic curiosity. Personality and individual differences 44, 7 (2008), 1585–1595.

[39] Jordan A Litman. 2010. Relationships between Measures of I-and D-Type Curiosity, Ambiguity Tolerance, and Need for Closure: An Initial Test of the Wanting-Liking Model of Information-Seeking. Personality and Individual Differences 48, 4 (2010), 397–402.

[40] Jordan A Litman, Robert P Collins, and Charles D Spielberger. 2005. The nature and measurement of sensory curiosity. Personality and Individual Differences 39, 6 (2005), 1123–1133.

[41] Jordan A. Litman and Tiffany L. Jimerson. 2004. The Measurement of Curiosity As a Feeling of Deprivation. Journal of Personality Assessment 82, 2 (April 2004), 147–157. https://doi.org/10.1207/s15327752jpa8202_3

[42] Jordan A Litman and Paul J Silvia. 2006. The latent structure of trait curiosity: Evidence for interest and deprivation curiosity dimensions. Journal of Personality Assessment 86, 3 (2006), 318–328.

[43] George Loewenstein. 1994. The Psychology of Curiosity: A Review and Reinterpretation. Psychological Bulletin 116, 1 (1994), 75.

[44] Maarten Marsman and Eric-Jan Wagenmakers. 2017. Bayesian benefits with JASP. European Journal of Developmental Psychology 14, 5 (2017), 545–555.

[45] Edward McAuley, Terry Duncan, and Vance V Tammen. 1989. Psychometric properties of the Intrinsic Motivation Inventory in a competitive sport setting: A confirmatory factor analysis. Research quarterly for exercise and sport 60, 1 (1989), 48–58.

[46] Mojang AB. 2011. Minecraft. Game [Windows, MacOS, Mobile, Consoles]. (Nov. 2011).

[47] Lennart E Nacke, Chris Bateman, and Regan L Mandryk. 2011. BrainHex: preliminary results from a neurobiological gamer typology survey. In International conference on entertainment computing. Springer, 288–293.

[48] Nintendo EPD. 2017. The Legend of Zelda: Breath of the Wild. [Nintendo Switch, Wii U]. (2017).

[49] Marret K. Noordewier and Seger M. Breugelmans. 2013. On the Valence of Surprise. Cognition and Emotion 27, 7 (Nov. 2013), 1326–1334. https://doi.org/10.1080/02699931.2013.777660
[50] Kyle Orland. 2017. Breath of the Wild May Be the Best Zelda Game Ever. Ars Technica (March 2017).
[51] Christopher A. Oswald, Chris Prorock, and Shane M. Murphy. 2014. The Perceived Meaning of the Video Game Experience: An Exploratory Study. Psychology of Popular Media Culture 3, 2 (2014), 110.
[52] Anthony O'Hagan. 2008. The Bayesian approach to statistics. Handbook of probability: Theory and applications (2008), 85–100.
[53] Anthony O'Hagan. 2008. The Bayesian approach to statistics. Handbook of probability: Theory and applications (2008), 85–100.
[54] Mikki H. Phan, Joseph R. Keebler, and Barbara S. Chaparro. 2016. The Development and Validation of the Game User Experience Satisfaction Scale (GUESS). Human Factors 58, 8 (2016), 1217–1247. https://doi.org/10.1177/0018720816669646 arXiv:https://doi.org/10.1177/0018720816669646
[55] Prolific. 2021. Prolific - Online Participant Recruitment for Surveys and Market Research. https://www.prolific.co/
[56] Jeff Reback, Wes McKinney, jbrockmendel, Joris Van den Bossche, Tom Augspurger, Phillip Cloud, gfyoung, Simon Hawkins, Sinhrks, Matthew Roeschke, Adam Klein, Terji Petersen, Jeff Tratner, Chang She, William Ayd, Shahar Naveh, Marc Garcia, Jeremy Schendel, patrick, Andy Hayden, Daniel Saxton, Vytautas Jancauskas, Ali McMaster, Marco Gorelli, Pietro Battiston, Skipper Seabold, Kaiqi Dong, chris-b1, h-vetinari, and Stephan Hoyer. 2021. Pandas-Dev/Pandas: Pandas 1.2.2. Zenodo. https://doi.org/10.5281/zenodo.4524629
[57] Thomas G Reio and Jamie L Callahan. 2004. Affect, curiosity, and socialization-related learning: A path analysis of antecedents to job performance. Journal of Business and Psychology 19, 1 (2004), 3–22.
[58] Thomas G Reio Jr, Joseph M Petrosko, Albert K Wiswell, and Juthamas Thongsukmag. 2006. The measurement and conceptualization of curiosity. The Journal of Genetic Psychology 167, 2 (2006), 117–135.
[59] Richard M Ryan, C Scott Rigby, and Andrew Przybylski. 2006. The motivational pull of video games: A self-determination theory approach. Motivation and emotion 30, 4 (2006), 344–360.
[60] Mike Schaekermann, Giovanni Ribeiro, Guenter Wallner, Simone Kriglstein, Daniel Johnson, Anders Drachen, Rafet Sifa, and Lennart E Nacke. 2017. Curiously Motivated: Profiling Curiosity with Self-Reports and Behaviour Metrics in the Game Destiny. In Proceedings of the Annual Symposium on Computer-Human Interaction in Play. ACM, 143–156.
[61] Jesse Schell. 2008. The Art of Game Design: A book of lenses. CRC press.
[62] Klaus R. Scherer. 2005. What Are Emotions? And How Can They Be Measured? Social Science Information 44, 4 (Dec. 2005), 695–729. https://doi.org/10.1177/0539018405058216
[63] Frederick F Schmitt and Reza Lahroodi. 2008. The Epistemic Value of Curiosity. Educational Theory 58, 2 (2008), 125–148.
[64] Felix D. Schönbrodt and Eric-Jan Wagenmakers. 2018. Bayes Factor Design Analysis: Planning for Compelling Evidence. Psychonomic Bulletin & Review 25, 1 (Feb. 2018), 128–142. https://doi.org/10.3758/s13423-017-1230-y
[65] Nicola S. Schutte and John M. Malouff. 2020. Connections between Curiosity, Flow and Creativity. Personality and Individual Differences 152 (Jan. 2020), 109555. https://doi.org/10.1016/j.paid.2019.109555
[66] Graham G. Scott, Anne Keitel, Marc Becirspahic, Bo Yao, and Sara C. Sereno. 2019. The Glasgow Norms: Ratings of 5,500 Words on Nine Scales. Behavior Research Methods 51, 3 (June 2019), 1258–1270. https://doi.org/10.3758/s13428-018-1099-3
[67] Regina Seiwald. 2019. Games Within Games. In International Conference on Videogame Sciences and Arts. Springer, 18–31.
[68] Gillian Smith, Ryan Anderson, Brian Kopleck, Zach Lindblad, Lauren Scott, Adam Wardell, Jim Whitehead, and Michael Mateas. 2011. Situating Quests: Design Patterns for Quest and Level Design in Role-Playing Games. In Interactive Storytelling (Lecture Notes in Computer Science), Mei Si, David Thue, Elisabeth André, James C. Lester, Joshua Tanenbaum, and Veronica Zammitto (Eds.). Springer, Berlin, Heidelberg, 326–329. https://doi.org/10.1007/978-3-642-25289-1_40
[69] Charles D. Spielberger and Laura M. Starr. 1994. Curiosity and Exploratory Behavior. Motivation: Theory and research (1994), 221–243.
[70] SurveyCircle. 2021. SurveyCircle – The Largest Community for Online Research. https://www.surveycircle.com/en/.
[71] Alexandra To, Ali Safinah, Geoff F Kaufman, and Jessica Hammer. 2016. Integrating Curiosity and Uncertainty in Game Design. In Proceedings of 1st International Joint Conference of DiGRA and FDG. 1–16.
[72] Unknown Worlds Entertainment. 2018. Subnautica. Game [Windows, MacOS, Xbox One, PlayStation4]. (Jan. 2018).
[73] Sophie von Stumm, Benedikt Hell, and Tomas Chamorro-Premuzic. 2011. The Hungry Mind: Intellectual Curiosity Is the Third Pillar of Academic Performance. Perspectives on Psychological Science 6, 6 (Nov. 2011), 574–588. https://doi.org/10.1177/1745691611421204
[74] M. Fernanda Wagstaff, Gabriela L. Flores, Rawia Ahmed, and Sarah Villanueva. 2020. Measures of Curiosity: A Literature Review. Human Resource Development Quarterly (2020). https://doi.org/10.1002/hrdq.21417

[75] George M. Whitesides. 2018. Curiosity and Science. Angewandte Chemie International Edition 57, 16 (2018), 4126–4129. https://doi.org/10.1002/anie.201800684

[76] Georgios N Yannakakis and Ana Paiva. 2014. Emotion in games. Handbook on affective computing 2014 (2014), 459–471.

[77] Yimin Zheng. 2021. spatialentropy (0.1.0)[Python package]. https://pypi.org/project/spatialentropy/

[78] Marvin Zuckerman. 2007. The sensation seeking scale V (SSS-V): Still reliable and valid. Personality and Individual Differences 43, 5 (2007), 1303–1305.

Received February 2021 ; revised June 2021 ; accepted July 2021