Predicting Player Experience Without the Player
An Exploratory Study

Christian Guckelsberger[1,*]**, Christoph Salge**[2,3]**, Jeremy Gow**[1]**, and Paul Cairns**[4]

[1]Computational Creativity Group, Goldsmiths, University of London, London, UK
[2]Adaptive Systems Research Group, University of Hertfordshire, Hatfield, UK
[3]Game Innovation Lab, New York University, New York, US
[4]Department of Computer Science, University of York, York, UK
[*]Corresponding author. Email: c.guckelsberger@gold.ac.uk

ABSTRACT

A key challenge of procedural content generation (PCG) is to evoke a certain player experience (PX), when we have no direct control over the content which gives rise to that experience. We argue that neither the rigorous methods to assess PX in HCI, nor specialised methods in PCG are sufficient, because they rely on a human in the loop. We propose to address this shortcoming by means of computational models of intrinsic motivation and AI game-playing agents. We hypothesise that our approach could be used to automatically predict PX across games and content types without relying on a human player or designer. We conduct an exploratory study in level generation based on empowerment, a specific model of intrinsic motivation. Based on a thematic analysis, we find that empowerment can be used to create levels with qualitatively different PX. We relate the identified experiences to established theories of PX in HCI and game design, and discuss next steps.

Author Keywords

Player Experience; Procedural Content Generation; Models of Intrinsic Motivation; AI Players; Empowerment

INTRODUCTION

Procedural content generation (PCG) holds great advantages for modern video games: By algorithmically creating levels, as in *Spelunky* [57, 58], or other elements such as characters and music, game developers can satisfy the players' demand for richer and more detailed content, while keeping production costs and time manageable [52]. Adding new content like quests or weapons procedurally allows designers to create open-ended games, and increase their replay value [49]. Moreover, PCG can be used to adapt content to individual players [45], and has been proposed as a game mechanic in itself [9].

PCG algorithms require formal guidelines about the desired content characteristics. A procedurally generated level should without doubt be *playable*, i.e. there must be a way for the player to succeed or fail, or to experience the whole content

instance and not just a small part of it. Content should also be *novel* and *typical* (cf. [38]): a generated quest for instance should be different from existing quests, but still fit the game under consideration. However, nobody would care about a level, character or as a consequence even the overall game, if the content in question did not lead to a desired experience. *Player experience* (PX) substantially determines the *value* of a content instance, and consequently its acceptance and replayability (cf. [46]). The key challenge for PCG arising from this is to evoke a certain PX, when we have *no direct control over the content* which gives rise to that experience.

We argue that established methods in human computer interaction (HCI) and PCG do not address the challenge of predicting PX in PCG sufficiently. HCI's methods rely on human players and thus come with strong time and cost constraints. Dedicated PCG approaches either still involve human players for machine-learning predictive models of PX, or require designers to hand-craft such models based on explicit knowledge of the game's semantics. Any present approach relies on people, and is unlikely to scale when new elements are introduced to the game which might invoke new, unanticipated experiences. The challenge of assessing PX in PCG is thus somewhat paradoxical: can we evaluate PX independently of the players themselves, and of designers' knowledge as proxies to PX? Furthermore, can we specify a predictive model of PX which works across games without constant human involvement?

In this paper, we propose such a general, automatic approach to assess PX based on *computational models of intrinsic motivation*, and evaluate it qualitatively. More specifically, we suggest to predict PX from intrinsic motivation quantities calculated on game state samples from AI players. Our approach is independent of players and of designers' knowledge about the meaning of game tokens, and could thus work across different games and content facets. We support our proposal with an exploratory study based on *empowerment* [27], one specific model of intrinsic motivation. We conduct a thematic analysis of player think-alouds while playing procedurally generated levels of an (in)finite runner game. Formally, empowerment quantifies an agent's potential and perceivable influence on their environment in a given state. We find that using empowerment in our generic procedure allows the creation of levels that elicit qualitatively different PX. We relate our results to theories of PX in game design and HCI, and outline the next steps towards a full, quantitative proof-of-concept.

THE CHALLENGE OF ASSESSING PX IN PCG

PX describes the personal, transient and dynamic qualities an individual player experiences from interacting with a game [54]. Amongst these manifold experiences are competence, autonomy, control, immersion, presence, flow, engagement, relatedness, challenge, tension, curiosity, and affect.

HCI and PCG assess PX for different reasons, and HCI's instruments are not applicable to PCG straight away. In HCI, PX is evaluated e.g. to give directions at key points of game development, or to infer more general insights about how people interact with games. Crucially, PX is usually assessed on very *few* conditions, and *after* the player has experienced them. Denisova and Cairns [14] for instance bias participants with two explanations about a game's AI for the same game, and measure the effect of these conditions on their immersion. HCI researchers rely on rich subjective [5, 8, 54] feedback and objective data [30] from players to make detailed and reliable judgements. However, the speed of evaluation is limited by the participants' game-playing and information processing capacity. In addition, certain devices might be expensive to use, and participants want to be paid, thus imposing costs.

The goals and requirements in PCG are typically very different. HCI's instruments could still be applied when PCG is used as a tool during development (cf. [12]), and generated content is carefully curated by designers for the final game. However, the full power of PCG, especially for replayability, can only be unleashed when PCG is used in the shipped game, either offline before-, or online during play. In this scenario, usually *many* content instances must be assessed *before* they are experienced by a player. Each instance can be considered one experimental condition resulting from e.g. different parameter combinations in the generator. Speed is of prime importance, and HCI's traditional subjective and objective approaches do not scale. Acquiring explicit feedback from the player is rarely an option, as it is either too obtrusive, e.g. when using questionnaires, or too imprecise, e.g. when asking for binary preference ratings.

These shortcomings impose a bottleneck on PCG development: to guarantee a particular PX, designers are forced to keep their generators within tight constraints. In *Spelunky*, some *experiential control* is retained by predefining *experiential chunks* [46] during development, i.e. rooms that can be individually experienced by the player. These rooms are then procedurally arranged in a grid during runtime (cf. [58]). In other cases, generators are kept simple by relying on small grammars or noise-based algorithms [49]. These constraints make the *explicit* [12] prediction of player experience obsolete, but they considerably limit the *expressive range* [47] of generators, and thereby the actual benefits of PCG in practice.

Researchers in PCG aim to realise the benefits of less constrained and more complex generative approaches by using predictive models of PX that only require minimal human involvement. Yannakakis and Togelius [56] survey related work in their *experience-driven PCG* framework. A predictive model is essentially a mapping from game content to experiential qualities. We demonstrate the breadth and shortcomings of existing approaches by distinguishing two extreme means to establish this mapping. They belong to the classes of *model-based* and *theory-driven* approaches, respectively.

Recent *model-based* approaches reduce the dependency on players drastically by using player feedback only once for the *training* of a machine-learning model [49], which can then be applied over and over. Guzdial, Sturtevant and Li [19], for instance, create a predictive model of player enjoyment, difficulty and visual aesthetics for *Infinite Mario Bros.* [35] levels, a clone of *Super Mario Bros.* [31]. They use a convolutional neural network and player feedback to automatically extract content features that correlate with the targeted experiences.

Some *theory-driven* approaches are entirely independent of the player, but rely on designers instead: guided by their intuition and/or some qualitative theory, the designers must manually establish a functional mapping between content or gameplay features and experiential qualities [12]. While Yannakakis and Togelius [56] only consider approaches based on *static* game content, we explicitly include *simulation-based* approaches [52] in this category, where *gameplay* features are extracted from the simulation of human play with AI agents. Togelius, De Nardi and Lucas [51] predict a player's enjoyment of racing game tracks based on AI player's learning progress, performance variation and difference in driving speed. Sorenson, Pasquier and Di Paola [48] determine challenge in platformer levels similar to *Super Mario Bros.* [31] based on static level features such as gap size, the player's maximum jump length and the presence of enemies. Recently, Khalifa et al. [24] go one step further and present a more general model of player enjoyment for any game level encoded in the *Video-Game Description Language* [44]. Their model is based on both static level features like the presence of harmful sprites and gameplay features including the score difference of AI players.

Current approaches to predict PX in PCG thus still involve human players or designers at some point. We argue that this dependency renders them inflexible and imprecise: they are likely to fail as soon as new game elements are introduced or present elements are changed. This is commonplace during development, e.g. to balance a game. Adding the *Berserker* power-up to a battle scene in *Doom* [22] for instance is likely to change the player's experience drastically, but this correlation would not be reflected in a previously trained model. With present *model-based* approaches, we would have to acquire new feedback from players each time the game is changed. The *theory-driven* approach in contrast is constrained by the designers' anticipation skills. PX strongly depends on the *interaction* of game elements, and identifying all possible interactions and every possible situation can be very complex. We expect designers to not anticipate a substantial part of these interactions, making their hand-crafted models imprecise. The same inflexibility applies when manually crafting gameplay metrics to gather human player feedback *implicitly* [12].

PX can be assessed based on algorithmic prediction mechanisms, or by gathering explicit and/or implicit designer and player feedback. In this section, we have shown exhaustively (cf. [12]) that there is no means to assess PX in PCG in a sufficiently precise and unobtrusive way without constraining the power of the content generator. In summary, the identified

challenge of assessing PX in PCG is twofold: we have to assess PX independently of human players, while not relying on designers' knowledge about a game's semantics.

PREDICTING PX WITH INTRINSIC MOTIVATION MODELS

We propose to compensate for human feedback in the prediction of PX with computational models of intrinsic motivation. Being *intrinsically motivated* means to engage in "an activity for its inherent satisfactions rather than for some separable consequence" [39]. We could easily be mislead to consider games as intrinsically motivating because players act on game-internal goals, in contrast to external rewards such as a prize in an e-sports competition. However, this does not go far enough, as acting on such game-internal goals is not necessarily inherently satisfying and can still be considered a separate consequence from the player's perspective. We can most clearly observe intrinsic motivation in games without apparent goals, or when goal pressure is particularly low. In *Skyrim* [21] for instance, there is no time pressure to complete goals in terms of quests in the main storyline. Despite this, players do not become passive but explore the world and the skill-crafting system. The goals which determine this behaviour are then shaped within the players. The reason why most games are considered intrinsically motivating even in the presence of strong goal pressure is that these goals are designed to align with the players' intrinsic motivations: it is more common e.g. to find quests about exploring the world and gaining power than quest that would require one to stay in a boring place and do nothing. This concept of intrinsic motivation is what Pavlas et al. refer to as *autotelic experience* [33], and constitutes an important factor in many other theories of PX [40, 13, 1].

Given the concept's omnipresence in PX theories, we consider the use of *computational models* of intrinsic motivation a natural fit to predict PX. We adopt a strict, formal definition of intrinsic motivation put forward by Oudeyer and Kaplan [32]. Here, a motivation as basis for an agent's actions only depend on the "collation or comparison of information from different stimuli [. . .] independently of their semantics" [32]. Goals must therefore not only be generated from within the agent, but must also not refer to anything outside. We thus only need to know about a player's interaction with the game, but not about the meaning of game tokens. How game tokens affect our motivational values depends on the specific computational model under consideration. The independence of sensor semantics allows us to apply a model across different games and content types. In contrast to existing work, we use an AI agent to reproduce human play, and do not depend on actual players.

Shaker [45] has recently also proposed to embrace models of intrinsic motivation in PCG, but her focus is fundamentally different. She suggests reinforcement learning as an approach to content generation, and intrinsic motivations as reward signals to support the generation of novel and diverse content. While her generative system is driven by intrinsic rewards, she proposes to estimate PX based on the traditional methods outlined earlier, and relies on player feedback.

We describe our approach here informally and as it was employed in our later study. It comprises three consecutive steps.

1. Recording state samples

Our primary concern is to determine players' typical experience of a game. As basis for our prediction, we consequently have to record a state sample that closely resembles the *typical states* that players would pass through in our game. While only few will discover all the secrets in *Doom* [22], every player will engage in the major battles. At the same time, we have to identify *inaccessible states*: while the game parameters could realise many potential states, a player only perceives a small subset. We cannot determine these typical and accessible states from the static game content, but have to use playouts. PCG's time constraints enforce a *simulation-based approach* [52], i.e. we sample from the play of AI controllers.

2. Calculating intrinsic motivation quantities

We do not use models of intrinsic motivation to steer behaviour, but leverage the quantities underlying decision-making as predictors of human PX. To calculate these quantities efficiently, we condense the information in our sample to what is required by the model used. For our study, we only need to retain the visited states, but not their order. We then calculate our motivational quantity on every state in the sample.

3. Aggregating quantities into experience prediction

We finally aggregate the previously calculated, raw motivational quantities into a scalar *experience prediction*, which could eventually inform the experimenter or be used e.g. in the objective function of a content generator. We normalise the quantities across different content instances to allow for comparison. We aggregate all values of a playout in the mean.

A more detailed but generic description of our approach must take the specific requirements of other models of intrinsic motivation into account, and is yet to come. We believe that different models can be used to predict the effect of different types of content on PX. A model of curiosity, for instance, could be employed on *decorative content* [46] such as sound and visuals. For our exploratory study, we use a specific model of intrinsic motivation, *empowerment*, to predict the effect of *functional content* [46] in the form of level structure on PX.

EMPOWERMENT

Empowerment [27] is an information-theoretic quantity, defined over the relationship between the actuators and sensors of an artificial agent. Informally, empowerment measures the influence of an agent's actions on its environment (controllability), and the extent to which the agent can perceive this influence afterwards (observability). In short, it quantifies *an agent's potential, perceivable influence on the world*.

We have chosen empowerment for several reasons: It has already been briefly investigated in HCI to quantify the impact of uncertainty in manual control [53]. Interfaces with reduced empowerment were correlated with feelings of frustration, providing us with a preliminary connection between the formal model and user experience. Crucially, *empowerment maximisation* has been hypothesised as proxy for behaviour in animals and humans [27, 37, 17]. Some argue that "understanding games is approaching a phase where it is close to understanding the psychology of individual life experiences in general" [50], and we believe that empowerment could relate
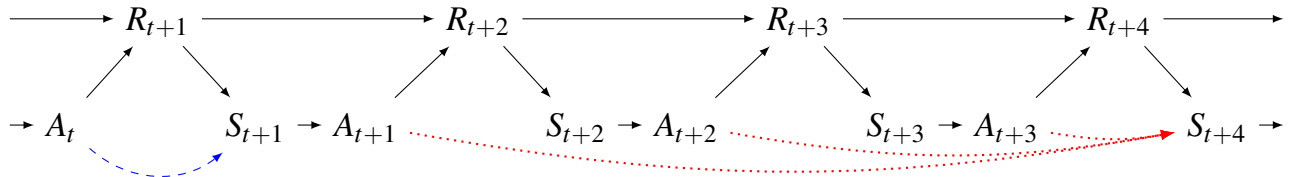
Figure 1: Causal Bayesian network of the perception-action loop of a memoryless agent unrolled in time, with the agent's sensors S, actuators A and the rest of the world R. The blue, dashed line represents the anticipation of follow-up states from $t \to t+1$. The red, dotted lines represent 3-step empowerment at time $t+1$.

to such experiences. Empowerment maximisation has already been used to drive the behaviour of general game-playing agents [3] and non-player characters [18]. The agents successfully identified latent game features, leading to meaningful behaviour in the absence of explicit goals [3]. This leads us to believe that empowerment might also influence how people play and consequently experience games.

At the centre of the empowerment definition is the interpretation of an agent's embodiment as an information-theoretic communication channel. For any arbitrary separation between an agent and a world R we can define the sensors S and actuators A as random variables whose states allow for the in- and outflow of information to the agent, respectively. This interaction with the world is usually described as a perception-action loop, which can be analysed by means of a causal Bayesian network as in Figure 1. Here, arrows imply causation between random variables: the agent's actions A only depend on its sensor input S, which in turn is determined by the rest of the system R. The latter is affected by the preceding system state and the agent's actions. Games can be understood as similar feedback systems [41], where R corresponds to the partially observable game state that the player influences and perceives with their available actions A and their senses S, respectively.

The causal probability distribution $p(S_{t+1}|S_t, A_t)$ represents the (potentially noisy) communication channel between a sensor state and actions at time t, and the future sensor states at $t+1$. In a game, this distribution is reflected in the player's *forward model*. Empowerment $\mathfrak{E}_{s_t}$ in a given state $S_t = s_t$ is defined as the *channel capacity*, i.e. the maximum potential flow of information that could possibly be induced into the future sensor state by a suitable choice of actions. More generally, we consider a sequence of actions $A_t^n = (A_t, \ldots, A_{t+n-1})$ corresponding to a lookahead of n. With *n-step empowerment* we can then quantify the amount of information an agent can potentially inject into its sensor n steps in the future:

$$\mathfrak{E}_{s_t}^n = \max_{p(a_t^n)} I(S_{t+n}; A_t^n = (A_t, \ldots, A_{t+n-1})|S_t = s_t) \quad (1)$$

The term being maximised represents the mutual information between the actuator and future sensor states, given the current sensor state s_t. Cover and Thomas [10] treat the information-theoretic notions in depth. For simplicity, we have defined empowerment for discrete interactions in time and space, but continuous implementations exist (cf. [43] for a survey).
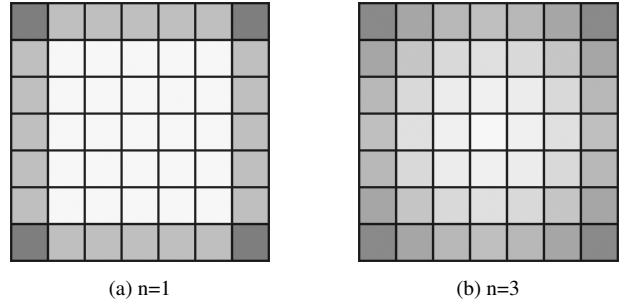

(a) n=1   (b) n=3

Figure 2: n-step empowerment of an agent at different positions in a gridworld, surrounded by walls. The agent can move north, east, south, west and idle, but it cannot go beyond the boundaries. The agent's sensor captures its position. Darker values indicate lower empowerment.

The red, dotted line (⸱⸱⸱⸱) in Figure 1 represents an agent's 3-step empowerment at $t+1$. Figure 2 illustrates empowerment and the difference between lookaheads in a simple gridworld surrounded by walls. The agent can sense its absolute position. It can idle or move north, east, south, and west, but it cannot cross the worlds' boundaries. There is no uncertainty about the consequences of actions. In this deterministic scenario, empowerment is proportional to the number of distinct states an agent can reach within the next n steps. For $n = 1$, it is very distinguished and low at the edges and in the corners, reflecting the impact the walls have on the agent's mobility. The larger lookahead ($n = 3$) produces a smooth reachability gradient. Empowerment is highest in the middle where the agent can access the most states in the specified time frame.

Empowerment has been hypothesised as driving principle in decision-making [27, 37, 17], i.e. animals and humans might choose their actions depending on the expected empowerment of the future situations that these actions might lead to. In our exploratory study, we want to quantify empowerment on game states, and consequently look at the player's *state-expected empowerment*. For the calculation of state-expected empowerment, we consider the action-expected empowerment for all available actions in the current state s_t, weighted by the probability that the player is going to choose these actions:

$$E[\mathfrak{E}^n]_{s_t} = \sum_{a_t} p(a_t|s_t) \sum_{s_{t+1}} p(s_{t+1}|s_t, a_t) \, \mathfrak{E}_{s_{t+1}}^n \quad (2)$$

Figure 3: The *RoboRunner* testbed: a deterministic, one-button (in)finite runner game.

This definition accounts for the possibility of noise in the agent's local dynamics, i.e. the player's forward model: a player might not be sure about the consequences of their actions, or the action outcomes might objectively be uncertain. Figure 1 shows the two stages in the calculation of 3-step state-expected empowerment at time t: the agent first anticipates which states its actions might yield at $t+1$ (- -), and then calculates 3-step empowerment for each state (·····).

Empowerment comes with several important properties. It is *local*, i.e. the agent's knowledge of the local dynamics $p(S_{t+1}|S_t, A_t)$ is sufficient to calculate the quantity. Its information-theoretic formulation makes empowerment *domain-independent* and *universal*, i.e., it can be applied to every possible agent-world interaction. Finally, empowerment is *task-independent*, i.e., it is not evaluated in regard to a specific external goal or reward, but determined by the agent's embodiment in the world. Empowerment can thus be calculated for any interaction of a player with a game, independent of the meaning of game states. Furthermore, we do not need to know about the game's complete mechanics or parameters, but only about how the player's actions affect what they can perceive. Existing studies demonstrate that empowerment is not restricted to single-player games [16, 18], and can deal with uncertainty and dynamic game worlds [42].

EXPLORATORY STUDY
Our overall goal is to predict PX with computational models of intrinsic motivation. This study represents the first step towards a proof-of-concept based on empowerment as one particular model used with our generic approach. While empowerment formally translates to an agent's potential and perceivable influence over the environment, we cannot assume the same experience for a human player. In order to conduct a quantitative study based on established, reliable HCI instruments for a full proof-of-concept, we first have to explore the aspects of PX to which empowerment potentially pertains.

Method
We approach this study as an instance of an *experiential vignette* [20] where the goal is to manipulate explicitly the response of players by providing them different games to play, but the data is analysed qualitatively in order to explore the concepts in play around the manipulation. For this reason, only modest numbers of participants are required. Different conditions here are given by different level instances of an (in)finite runner game. Our hypothesis is that levels with low mean state-expected empowerment evoke qualitatively different experiences than levels with high values. We conduct a *thematic analysis* [6] on player think-alouds to find out which experiences the different conditions give rise to. We decided against a more quantifiable approach such as content analysis, as the goal was not to see how often people engaged in a set of previously known experiences, but to explore the range of yet unknown experiences that people have in response to our manipulation. Given ambiguity and noise in the think-aloud data, we only coarsely state the frequency of specific experiences.

We have evaluated two conditions, given by levels with low- and high mean state-expected empowerment. We refer to these conditions by *low* and *high*. We have chosen a mixed experimental design, in which the two conditions are evaluated within subjects, while two instantiations of the same condition are considered between subjects. The levels were procedurally generated, and, due to stochasticity in the generator, can vary locally despite having the same mean state-expected empowerment. The goal of our mixed design was to balance effects of these within-condition differences on PX. Each participant first had to complete one of two tutorial levels with medium state-expected empowerment to familiarise with the game. We thus evaluated 8 level combinations (2 instances × 2 conditions × 2 tutorials), presented in balanced order.

Testbed: RoboRunner
Participants were asked to play different levels of *RoboRunner* (Figure 3), a simple game from the infinite runner genre specifically developed for this study. The goal is to drive a yellow robot from the left to the right of a space station. In order to reach the end of the level, the robot has to jump across chasms. The game has one implicit and one explicit action: the robot will drive to the right automatically with constant speed, while the player can invoke jumps only when the robot is on the floor. While *RoboRunner* has all characteristics of an infinite runner, our levels are in fact finite to ensure that all participants experience the same content.

We have chosen this particular testbed for several reasons. Infinite runners are, despite their simplicity, very popular. Our participants have likely seen or played similar games before, and we have assumed that they can thus familiarise with our

(a) Condition: low mean state-expected empowerment ($E[\mathfrak{C}]^1_{S_t} = 0.65$)

(b) Tutorial: medium mean state-expected empowerment ($E[\mathfrak{C}]^1_{S_t} = 0.75$)

(c) Condition: high mean state-expected empowerment ($E[\mathfrak{C}]^1_{S_t} = 0.85$)
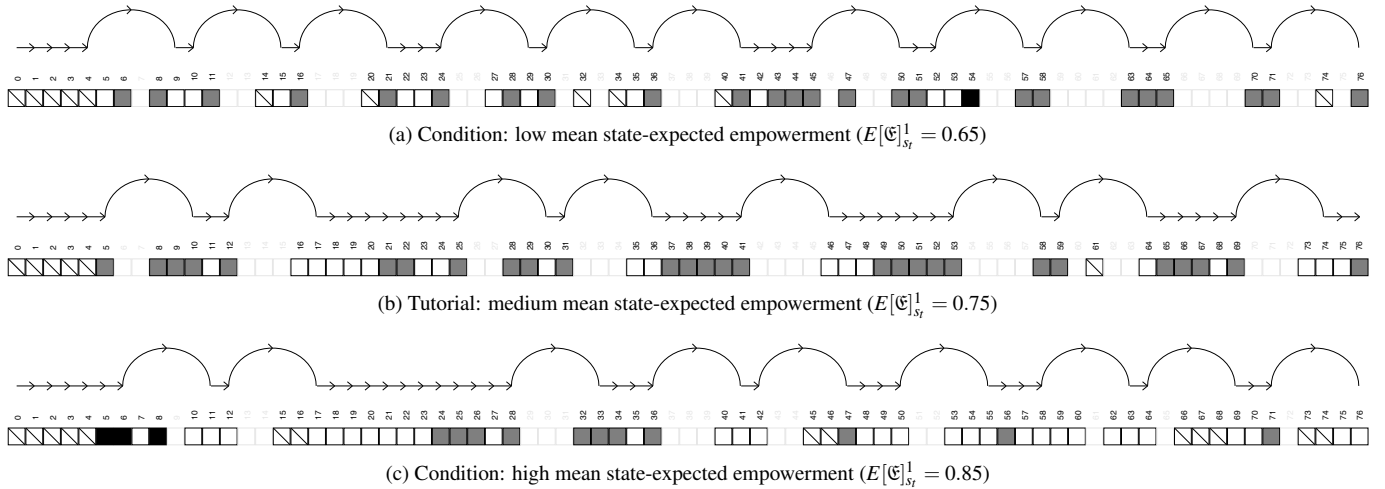
Figure 4: Left third of three levels used in the study, with different mean 1-step state-expected empowerment. Striked-through modules represent irrelevant states. The AI play trajectory is shown above. Each level starts and ends with a 5-module safety zone.

testbed more easily. Due to its linear nature, *RoboRunner* allows us to determine the relevant states the player is going to pass through with little ambiguity: while the player can jump at different positions, there are no alternative routes to reach the end of the level. The simple controls and linearity also eased the task to create an AI controller which resembles human play. The game appears continuous in time and space to the player, but it is composed of modular blocks and can consequently be treated as discrete in space to simplify our analysis. *RoboRunner* is deterministic, i.e. only the player causes changes to the game state. Each state is thus fully characterised by the player's position, and we can visualise the whole state space at once. Furthermore, the game's low complexity allows us to develop a level generator without biases prior to the evaluation with an objective function.

Participants

The study was completed by eight participants (5 male, 3 female). Six were aged between 25-34, and two belonged to age groups 18-24 and 35-44, respectively. They were all native speakers, and most were recruited from our local MSc and PhD games programmes. We can thus presume a good video game playing experience, supported by a reported average of 16.25 hours of video game playing per month ($sd = 5.29$). Their experience is critical, as making progress in our levels is essential for rich and representative responses in the think-aloud. The participants were incentivised with chocolates.

Materials

We have generated and selected our levels in a two-stage process. We first procedurally generated level candidates over the whole range of possible instances with a genetic algorithm, to avoid any design bias from the authors. We employed a simple AI player to determine if a level is generally playable, and to find all accessible states that a human player could possibly experience. To increase the human-likeness, the controller realises a *reaction time constraint*, i.e. our AI player could not jump from the same module in which it landed. We then

calculated the player's state-expected empowerment (Equation 2) for each state of the current genotype, except for the inaccessible states. The first- and last five modules of the level were considered unchangeable to form a "safety zone" for the player, and were also excluded from the calculation. Together, these states were labelled *irrelevant*. We then aggregated all quantities in the mean to form our experience prediction. The fitness of each level was determined from the distance of this prediction to a target mean.

In a second step, we selected the stimuli for our evaluation from the generated candidate levels. We observed that levels with a mean state-expected empowerment below 0.65 had very short platforms and large gaps, and were thus almost unplayable. Levels with a mean state-expected empowerment above 0.95 in contrast had almost no gaps, and did not represent typical examples of an infinite runner. We consequently picked three values from the *playable and typical range*: 0.65 for our *low* condition, 0.85 for *high*, and 0.75 for our tutorial level. We generated two instances of each condition and the tutorial level to balance for local structural differences that may affect PX. We ended up with eight unique level combinations. Figure 4 shows the left third of an instance of each level type with a trajectory of our AI player. Modules with brighter hues indicate higher state-expected empowerment for an agent in that position. Irrelevant states are crossed-through.

For an example consider the small platform at position 47 in Figure 4a. State-expected empowerment here is lower, as only jumping would allow the agent to get on solid ground and perform more actions in the future. The opportunity to hit the small platform is reflected in the higher value at position 42, as this is where the agent would have to jump to hit the platform.

Procedure

The experiment was run individually under lab conditions. Each participant was assigned one level combination, starting with the tutorial and followed by the conditions in balanced order. They were introduced to *RoboRunner*, and asked to play the tutorial level to get familiar with the game and controls.

They were then prompted to play the two remaining levels until completion, or for at most five times, while thinking aloud. After completing both conditions, we asked them:

- Did you spot any differences between the levels?
- Did they feel different while playing?
- Which level did you enjoy more, and why?

This interview was structured to consist of only these three questions. The entire session lasted about 25 minutes per participant, and was recorded on audio.

Results

The thematic analysis was jointly conducted by the authors, following the procedure by Braun and Clark [6]. We first exhaustively transcribed the think-alouds, and familiarised ourselves with the data. In an iterative process, we then assigned codes and clustered them into themes that occurred across participants. A code dictionary was used between the authors to ensure consistency. The targeted interview data was analysed similarly to the think-aloud. However, because interview data is more reflective by nature, it also gave the opportunity to interpret some of the more immediate responses generated by the think-aloud data. This informed the development of the thematic analysis. Our results show how structural differences, based on empowerment, correlate with PX. We identified six major themes: *challenge*, *involvement*, *attention and engagement*, *learning* and *emotions*.

Physical & Cognitive Challenge

The players consistently considered *low* as more *challenging*:

> Ouh god it's tricky! I see, yes! God is that ...! (1)

Figure 4 shows that *low* has smaller platforms and more gaps than *high*. Players were sensitive to how these structural differences affected them, allowing us to distinguish two types of challenge: *physical challenge* addresses the player's physical limitations to interact with the game, affecting the speed and accuracy of actions [11]. *Cognitive challenge* in contrast is based on the player's cognitive capacities, and the speed and accuracy of problem solving [11]. In *RoboRunner*, the avatar's speed and the level layout determine how often the player has to jump and judge distances to succeed, influencing both physical and cognitive challenge and making them hard to separate. An exception is given by short platforms as around position 47 in Figure 4a: here, the player has to jump on- and off quickly, while the immanent gap makes further reasoning unnecessary. Participants described their struggle in terms of reaction time:

> This is tricky, because you feel like you should jump in the middle, and I suspect when I jump on the middle one, I won't jump off in time. (2)

Players experienced *cognitive challenge* in the levels' demand for *spatial reasoning* and *planning*:

> My spatial reasoning, to determine whether or not I can actually make that [jump]. So, it feels like the biggest question it is asking me is whether my judgement of the distance, *pause*, whether I should jump sooner or later. It makes me question my ability to do that. (3)

If you chose the wrong thing, you would be left in a situation where you had to jump and land in a hole. So you wouldn't jump at the earliest opportunity. (4)

> The small platforms are those that are challenging, because you have to bounce off them really quickly, and measure your distance appropriately so you don't fall down the gaps, which I've now done 5 times! (5)

Both types of challenge were increased by decision pressure:

> The second one (*low*), it felt like [the gaps] were a lot smaller, and less frequent, so there wasn't so much of a challenge of 'I have to make this decision now'. (6)

Involvement

Participants described the experience of being *actively involved* or being *passive*, depending on the frequency with which they had to take action to succeed in the game. Almost all participants considered themselves as more passive in *high*:

> The [*low* condition] seemed to have less of those really short platforms, and they had longer stretches of where you just didn't, didn't jump. So I kind of waited. (7)

They however preferred *low* for being more involving:

> Cause this one (*high*) has like too many ... where you're just not jumping. It's got quite a loong platform. So it's kind of, just less interesting, than hard games. (8)

A lack of threats did not discourage all participants from acting, with some jumping wildly on a long stretch without gaps:

> Oh, I'm trying to do a double for fun. And another! (9)

Attention and Engagement

Almost all participants talked about *attention* in terms of *concentration* and *focus*. More than half of them had to *concentrate* harder while playing *low*, expressed explicitly and through interrupted speech:

> Yeah, I'm having to concentrate a lot harder (pause) than before. And I'm, and I'm, ..., och, I'm keeping doing that. I'm keeping falling down this one. (10)

While statements on concentration were rather unspecific, participants reported on their *focus* with respect to the game elements. They e.g. focused on jumping, while blocking out the game's visuals or details of the level structure:

> I don't know if the gaps were set on random or not. I was so focussing on jumping, I wasn't actually looking at the sizes of them. I think maybe, *pause*, the spaces changed in the game as well. (11)

In more focussed situations, players also expressed a stronger *engagement*. The majority of participants were *determined* to succeed and willing to *re-engage*:

> I'm gonna make these lives count. Now it's personal! (12)

Some participants described their engagement as "being on a roll", "zoning in" and as "flow":

> [*Low* is] a tougher game that you get into flow quite easily, because you're quite focused on just jumping over the obstacles. (13)

Learning

Almost all participants referred to the concept of *learning*. They talked about remembering mistakes and the level structure for their next attempt:

> And the gaps are, the gaps are, *pause*, whoops! Oh no, that was an error! (robot dropped into gap) I'm gonna remember, I'm gonna remember the end now. (14)

The majority of participants touched on their *performance* and *progress* without being asked explicitly, and by relating back to previous attempts:

> I felt like it (*low*) was quite hard, but I feel like, even though it is frustrating to constantly die playing a level, I feel like I would've gotten there eventually. (15)

Emotions

Our impression of engagement is supported by the players' varied expression of emotions. Almost everyone expressed *pleasure* when succeeding in the game and *disappointment* when failing, but some were also *angry*, *tense* and even *anxious*. They were particularly upset when failing at the same position repeatedly:

> So, cause I know what's coming now, I played it enough times, *pause*, I kind of, *pause*, bloody hell! (16)

This previous quote also expresses some *tension*. The feeling of tension was usually followed by relief:

> Sh! Uh oh oh oh, oh! Oh no! Ah! *Deep breath, exhaling.* Puh! Release now, I'm really upset! (17)

Some participants even expressed some *anxiety* from anticipating and assessing upcoming situations in the game:

> It's this little one that gives me the heebie-jeebies. (18)

As is common in thematic analysis, the themes are intended to stand alone as broad qualitative descriptions of the data. The method is not designed to look more causally at links between the themes but future work will exam the links explicitly.

DISCUSSION

Altering game levels based on empowerment does not only yield different structural outcomes; the preceding analysis supports our hypothesis that it also evokes qualitatively different experiences in human players. We have reported our results independently of established PX theories. We now make up for this and address potential connections between empowerment and PX theories both in game design and HCI. We selected the theories based on the formal interpretation of empowerment, and related them to the results of our exploratory study.

Empowerment & PX in Game Design

Empowerment resonates with several theories on game design. Salen and Zimmerman describe *play* in the most general sense as *"free movement within a more rigid structure"* ([41], p. 304). This rigid structure emerges from the interplay of game rules and content, and is explored and exploited in play. Such exploitation however requires the player to become aware of the structure's boundaries. Formally, empowerment quantifies an artificial agent's perceivable influence on the environment in terms of controllability and observability, and thus allows to trace the boundaries of an agent's rigid world structure and to identify latent game features [3]. To engage in play, human players might employ a similar metric to trace the structure imposed on them: When talking about challenge in our study, players referred to particularly constrained areas (Quote 4, 5).

The notion of play as free movement within a rigid structure is strongly related to a player's *control* in a game. *Super Mario Bros.*'s [31] creator Shigeru Miyamoto puts *Tegotae*, a satisfaction coming from being in control, at the core of their games design [34]. He explains "if the player jumps from a high place, the amount of time they stay in the air needs to be just right, or else they'll feel disconnected from the experience". The player's empowerment would be high on the ground, and drop to zero while being in the air. We thus hypothesise that empowerment correlates with *Tegotae* as PX. The player who jumped around wildly on a long stretch might have probed their *Tegotae* in a situation where they are otherwise just passively pushed forward by the game mechanics (Quote 9).

Salen and Zimmerman also deem *outcome uncertainty* essential for meaningful play: "If the outcome of the game is completely predetermined – then any choices a player makes are meaningless, because they do not impact the way that the game plays out" ([41], p. 174). Caillois [7] directly relates outcome uncertainty to enjoyment. Empowerment measures the consequences of a player's action sequences, and could thus, given a sufficiently large lookahead, quantify the player-distinguishable game outcomes. In practice however, this lookahead is bound by the available computational resources. Empowerment is then restricted to the near future and captures the richness of intermediate rather than terminal game states. Outcome uncertainty however is also influenced by noise in the player's actions: it can be crucial whether a given action leads to one distinct outcome, or to a set of different possible outcomes with equal probability. This noise might be inherent to the game content and rules, but it might also only exist in the agent's local model of the game dynamics. Our testbed is deterministic, but players were uncertain about their reaction times and spatial reasoning (Quote 2, 3). As an agent-centric measure, empowerment captures both types of noise.

Empowerment & PX in HCI

These design theories are supported by empirical studies in HCI. Abuhamdeh, Csikszentmihalyi and Jalal [1] for instance show that greater *outcome uncertainty* leads to an increase in enjoyment. This effect was mediated by *suspense*, which also shows in our participant's *tension* in *low* (Quote 17).

Perceivable control as in *Tegotae* is also considered an important element in many theoretical models of *engagement*, e.g. *cognitive absorption* [2], *flow* [13], and *immersion* [23]. Some of our participants said to "be on a roll", "zoned in" and in "flow" (Quote 13). Nevertheless, we focus on the more immanent relationship between empowerment and players' *perceived control*. Strictly speaking, empowerment does not directly correspond to control; it relates more to *effectance*, which can be considered a part or precondition of control. Klimmt, Hartmann and Frey define effectance as the experience of "receiving immediate, direct feedback on one's action and of influencing the game world" [26]. They empirically link

effectance to game enjoyment, but their experiment is limited to situations where available actions are expected to, but do not affect the current state. In our study, the player who exercise their effectance by jumping wildly clearly enjoyed themselves (Quote 9). Empowerment additionally differs from control in that it also quantifies how the game constrains the *availability* of actions, but this was not the case in *RoboRunner*.

The players' perceived *challenge* represents the strongest theme in our study. Challenge is an element of many PX theories, and there is general agreement that being optimally challenged leads to stronger enjoyment [25]. In our study, most participants preferred *low*, the levels that were also generally considered more challenging. Optimal challenge is also an element in theories of immersion [11] and flow [13], and it might not be a coincidence that our participant said to be "in flow" for *low* (Quote 13). Malone [29] has related challenge to a player's uncertainty, and our participants mentioned their uncertainty in spatial reasoning when talking about challenge (Quote 3). Poels et al. [36] show that challenge can also trigger negative emotions such as tension, disappointment or anger, and our study supports their claim (Quote 16 –18). Our participants also appeared to talk more about their *attention* and *learning* as a result of cognitive challenge (Quote 10, 14).

We finally relate empowerment to *Self-Determination Theory* (SDT), a popular theory of intrinsic motivation which has also been applied to PX [40]. SDT is formed of several sub-theories; *Cognitive Evaluation Theory* highlights the critical role of a person's *competence* and *autonomy* in intrinsic motivation. Ryan, Rigby and Przybylski [40] relate these arguably loosely defined concepts to enjoyment in games. In-game autonomy is "the degree of choice one has over the sequence of actions, or the tasks and goals undertaken" [40]. Competence in contrast is defined as a player's "need for challenge and feelings of effectance" [40]. We believe that empowerment relates to autonomy in that it quantifies the availability of actions in different game states, and relates to competence in terms of effectance as discussed earlier.

MOVING FORWARD

Our exploratory study highlights the next steps for a strong proof-of-concept on employing computational models of intrinsic motivation to predict PX. Empowerment as one particular model evokes qualitatively different PX and thus qualifies as a candidate for future studies. However, we cannot tell yet whether the identified relationships, e.g. between empowerment and challenge, will persist beyond our very minimal testbed, and if the direction of the observed effects remains. The most important step thus is to investigate if our approach *generalises*. In our next study, we will validate the identified relationships across different game genres and reward structures based on quantitative player feedback.

We furthermore have to *discriminate the PX* that empowerment directly relates to rather than mediating effects. Many experiences have been described in terms of other, more fine-grained ones, such as effectance as part of control, or different types of challenge [15]. We have to use players' quantitative feedback to identify which experiences empowerment correlates with strongest. We must also consider the effect of player

skill on experience. Ultimately, we want to compile a list of intrinsic motivation models, mapped to the types of PX they are best suited to predict in different game genres.

The accuracy of our prediction is strongly influenced by how closely the behaviour of our AI agents resembles human play. A truly general, human-independent predictor of PX must be complemented with an accurate, *general model of player behaviour*. We can consider the challenge of automatically predicting PX as an iterative refinement process: on the one hand, the prediction of PX requires us to make general game-playing agents more human-like. On the other hand, our models of intrinsic motivation might provide insights on how players experience and consequently act in games, which could contribute to developing more human-like AI players. General game-playing agents are not truly general yet [4], and research on making AI agents more human-like still has some way to go. Luckily, we can separate these two sides in the lab: for our next study, we will record the play traces of human players, use them as basis for our prediction, and relate the outcome to their quantitative feedback on PX.

The experience of individual participants in our study varied strongly within a level. In order to clearly separate experimental conditions in cases where participants will unlikely complete them, we have to ensure more homogeneity of experience within conditions. We can realise this by not collating all motivational values into an average, but preserving the time of experienced states and fitting the intrinsic motivation quantities to a function, e.g. a constant. We plan to use this *time-series approach* for content generation in future studies.

We have focussed on predicting a player's *typical experience* in a video game, but suggest to adopt this approach for two more scenarios. On the one hand, we can learn a model from the behaviour of one individual or use player persona models [28] to predict the experience of one person or a small group of players for the generation of *custom-tailored content*. On the other hand, we can move from individual pieces of content to analysing the entire content generator. We suggest to take Smith and Whitehead's concept of *expressive range* analysis [47] further and calculate the "*experiential range*" of a given generator, i.e. all experiences it can potentially give rise to, given possible parameter configurations.

CONCLUSION

PCG yields many advantages for modern video games, and we believe that assessing the player's experience of procedurally generated content represents an important and interesting challenge for HCI research. We have argued that neither the rigorous methods to assess PX in HCI, nor the specialised methods developed by PCG researchers are sufficient, because they rely on a human in the loop. We proposed to solve this paradoxical challenge with computational models of intrinsic motivation. We hypothesise that these models could be used to automatically predicting PX across different games and content types without relying on a human player or designer.

As a first step towards a proof-of-concept, we conducted a qualitative, exploratory study in level generation based on empowerment. A thematic analysis allowed us to identify

correlates between empowerment and *challenge*, *involvement*, *attention and engagement*, *learning* and *emotions*. After relating our findings to established PX theories, we think that these experiences are more mediating effects of empowerment, while empowerment might be more directly related to *effectance*, *(outcome) uncertainty* and *perceived control*. The most important next step is to investigate the generality of our claims, and the predictive power of empowerment in respect to the identified experiences. We hope that our findings encourage researchers to examine the relationship between other models of intrinsic motivation and PX. While empowerment might be most suitable for predicting PX arising from *functional content* [46], other models are likely to account for the impact of *decorative content* [46], e.g. sounds and visuals, too.

We have motivated our approach with the tough challenges that PCG imposes on present methods for assessing PX. However, our proposal is not limited to PCG. We do not think that our approach is going to match the precision of traditional HCI instruments; nevertheless, its speed and potential flexibility could be used in *creativity support tools* to quickly assess game designers' hand-crafted content without interrupting their workflow, or to propose alterations to their content in a *co-creative* manner [55]. This could guide designers to convey the experiences they envisaged, and to create better games.

ACKNOWLEDGEMENTS

REFERENCES

1. S. Abuhamdeh, M. Csikszentmihalyi, and B. Jalal. 2015. Enjoying the Possibility of Defeat: Outcome Uncertainty, Suspense, and Intrinsic Motivation. *Motiv. Emot.* 39, 1 (2015), 1–10.

2. R. Agarwal and E. Karahanna. 2000. Time Flies When You're Having Fun: Cognitive Absorption and Beliefs About Information Technology Usage. *MIS quarterly* (2000), 665–694.

3. T. Anthony, D.l Polani, and C. L. Nehaniv. 2014. General Self-Motivation and Strategy Identification: Case Studies Based on Sokoban and Pac-Man. *IEEE T-CIAIG* 6, 1 (2014), 1–17.

4. P. Bontrager, A. Khalifa, A. Mendes, and J. Togelius. 2016. Matching Games and Algorithms for General Video Game Playing. In *Proc. Conf. AIIDE*.

5. E. A. Boyle, T. M. Connolly, T. Hainey, and J. M. Boyle. 2012. Engagement in Digital Entertainment Games: A Systematic Review. *Computers in Human Behavior* 28, 3 (2012), 771–780.

6. V. Braun and V. Clarke. 2006. Using Thematic Analysis in Psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101.

7. R. Caillois. 1961. *Man, Play and Games*. University of Illinois Press.

8. P. Cairns, A. Cox, and A. I. Nordin. 2014. Immersion in Digital Games: Review of Gaming Experience Research. *Handbook of Digital Games* (2014), 337–361.

9. M. Cook, M. Eladhari, A. Nealen, M. Treanor, E. Boxerman, A. Jaffe, P. Sottosanti, and S. Swink. 2016. PCG-Based Game Design Patterns. *arXiv preprint arXiv:1610.03138* (2016).

10. T. M. Cover and J. A. Thomas. 1991. *Elements of Information Theory*. Wiley-Interscience.

11. A. Cox, P. Cairns, P. Shah, and M. Carroll. 2012. Not Doing But Thinking: The Role of Challenge in the Gaming Experience. In *Proc. ACM Conf. CHI*. 79–88.

12. R. Craveirinha, N. Barreto, and L. Roque. 2016. Towards a Taxonomy for the Clarification of PCG Actors' Roles. In *Proc. ACM Conf. CHI Play*. 244–253.

13. M. Csikszentmihalyi, S. Abuhamdeh, and J. Nakamura. 2014. Flow. In *Flow and the Foundations of Positive Psychology*. Springer, 227–238.

14. A. Denisova and P. Cairns. 2015. The Placebo Effect in Digital Games: Phantom Perception of Adaptive AI. In *Proc. ACM Conf. CHI Play*. 23–33.

15. A. Denisova, C. Guckelsberger, and D. Zendle. 2017. Challenge in Digital Games: Towards Developing a Measurement Tool. In *Proc. ACM Conf. CHI*.

16. C. Guckelsberger and D. Polani. 2014. Effects of Anticipation in Individually Motivated Behaviour on Survival and Control in a Multi-Agent Scenario with Resource Constraints. *Entropy* 16, 6 (2014), 3357–3378.

17. C. Guckelsberger and C. Salge. 2016. Does Empowerment Maximisation Allow for Enactive Artificial Agents?. In *Proc. Conf. ALIFE*. 704–711.

18. C. Guckelsberger, C. Salge, and S. Colton. 2016. Intrinsically Motivated General Companion NPCs. In *Proc. IEEE Conf. CIG*. 150–157.

19. M. Guzdial, N. Sturtevant, and B. Li. 2016. Deep Static and Dynamic Level Analysis: A Study on Infinite Mario. In *Experimental AI in Games Workshop*, Vol. 3.

20. M. Hudson and P. Cairns. 2014. Interrogating Social Presence in Games with Experiential Vignettes. *Entertainment Computing* 5, 2 (2014), 101–114.

21. id Software. 2011. *The Elder Scrolls V: Skyrim*. Game [Windows, PlayStation 3/4, Xbox 360/One]. (11 Nov 2011). Bethesda Game Studios.

22. id Software. 2016. *Doom*. Game [PC/PS4/XBoxOne]. (13 May 2016). id Software, Richardson, USA.

23. C. Jennett, A. L. Cox, P. Cairns, S. Dhoparee, A. Epps, T. Tijs, and A. Walton. 2008. Measuring and Defining the Experience of Immersion in Games. *Human Computer Studies* 66, 9 (2008), 641–661.

24. A. Khalifa, D. Perez-Liebana, S. M. Lucas, and J. Togelius. 2016. General Video Game Level Generation. In *Proc. Conf. Genetic and Evolutionary Computation*.

25. M. Klarkowski, D. Johnson, P. Wyeth, M. McEwan, C. Phillips, and S. Smith. 2016. Operationalising and Evaluating Sub-Optimal and Optimal Play Experiences

through Challenge-Skill Manipulation. In *Proc. ACM Conf. CHI*. 5583–5594.

26. C. Klimmt, T. Hartmann, and A. Frey. 2007. Effectance and Control as Determinants of Video Game Enjoyment. *Cyberpsychology & Behavior* 10, 6 (2007), 845–848.

27. A. S. Klyubin, D. Polani, and C. L. Nehaniv. 2008. Keep Your Options Open: An Information-Based Driving Principle for Sensorimotor Systems. *PloS one* 3, 12 (2008), 1–14.

28. A. Liapis, C. Holmgård, G. N. Yannakakis, and J. Togelius. 2015. Procedural Personas as Critics for Dungeon Generation. In *Proc. European Conf. Applications of Evolutionary Computation*. 331–343.

29. T. W. Malone. 1981. Toward a Theory of Intrinsically Motivating Instruction. *Cognitive science* 5, 4 (1981), 333–369.

30. L. E. Nacke. 2013. An Introduction to Physiological Player Metrics for Evaluating Games. In *Game Analytics*. Springer, 585–619.

31. Nintendo R&D4. 1985. *Super Mario Bros.* Game [NES]. (13 September 1985). Nintendo, Kyoto, Japan.

32. P.-Y. Oudeyer and F. Kaplan. 2008. How Can We Define Intrinsic Motivation?. In *Proc. Conf. Epigenetic Robotics*. 93–101.

33. D. Pavlas, F. Jentsch, E. Salas, S. M. Fiore, and V. Sims. 2012. The Play Experience Scale: Development and Validation of a Measure of Play. *Human Factors* 54, 2 (2012), 214–225.

34. A. Perry. 2016. 'Tegotae' is essential to Nintendo's philosophy on video games – here's what it means. `http://www.techinsider.io/tegotae-is-essential-to-nintendos-philosophy-2016-7`. (2016). [Online; accessed 21-July-2016].

35. M. Persson. 2010. *Infinite Mario Bros.* Game [Cross-Platform/Java]. (2010).

36. K. Poels, Y. De Kort, and W. Ijsselsteijn. 2007. It Is Always a Lot of Fun!: Exploring Dimensions of Digital Game Experience Using Focus Group Methodology. In *Proc. ACM Conf. Future Play*. 83–89.

37. D. Polani. 2009. Information: Currency of Life? *HSFP Journal* 3, 5 (2009), 307–316. Issue 3.

38. G. Ritchie. 2007. Some Empirical Criteria for Attributing Creativity to a Computer Program. *Minds and Machines* 17, 1 (2007), 67–99.

39. R. M. Ryan and E. L. Deci. 2000. Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. *Contemp. Educ. Psych.* 25, 1 (2000), 54–67.

40. R. M. Ryan, C. S. Rigby, and A. Przybylski. 2006. The Motivational Pull of Video Games: A Self-Determination Theory Approach. *Motiv. Emot.* 30, 4 (2006), 347–363.

41. K. Salen and E. Zimmerman. 2004. *Rules of Play: Game Design Fundamentals*. MIT Press.

42. C. Salge, C. Glackin, and D. Polani. 2014a. Changing the Environment Based on Empowerment as Intrinsic Motivation. *Entropy* 16, 5 (2014), 2789–2819.

43. C. Salge, C. Glackin, and D. Polani. 2014b. Empowerment – an Introduction. *Guided Self-Organization: Inception* (2014), 67–114.

44. T. Schaul. 2013. A Video Game Description Language for Model-Based or Interactive Learning. In *Proc. IEEE Conf. CIG*. 1–8.

45. N. Shaker. 2016. Intrinsically Motivated Reinforcement Learning: A Promising Framework for Procedural Content Generation. In *Proc. IEEE Conf. CIG*. 1–8.

46. G. Smith. 2014. Understanding Procedural Content Generation: a Design-Centric Analysis of the Role of PCG in Games. In *Proc. ACM Conf. CHI*. 917–926.

47. G. Smith and J. Whitehead. 2010. Analyzing the Expressive Range of a Level Generator. In *Proc. 2010 Workshop on PCG in Games*. 1–7.

48. N. Sorenson, P. Pasquier, and S. DiPaola. 2011. A Generic Approach to Challenge Modeling for the Procedural Creation of Video Game Levels. *IEEE T-CIAIG* 3, 3 (2011), 229–244.

49. A. Summerville, S. Snodgrass, M. Guzdial, C. Holmgård, A. K. Hoover, A. Isaksen, A. Nealen, and J. Togelius. 2017. Procedural Content Generation via Machine Learning. *arXiv preprint arXiv:1702.00539* (2017).

50. J. Takatalo, J. Häkkinen, J. Kaistinen, and G. Nyman. 2010. Presence, Involvement, and Flow in Digital Games. In *Evaluating User Experience in Games. Concepts and Methods*, R. Bernhaupt (Ed.). Springer, 23–46.

51. J. Togelius, R. De Nardi, and S. M. Lucas. 2007. Towards Automatic Personalised Content Creation for Racing Games. In *Proc. IEEE Symp. CIG*. 252–259.

52. J. Togelius and N. Shaker. 2015. The Search-Based Approach. In *Procedural Content Generation in Games*, N. Shaker, J. Togelius, and M. J. Nelson (Eds.). Springer, 17–30.

53. D. Trendafilov and R. Murray-Smith. 2013. Information-Theoretic Characterization of Uncertainty in Manual Control. In *Proc. 2013 IEEE Conf. Systems, Man, and Cybernetics*. 4913–4918.

54. J. Wiemeyer, L. Nacke, C. Moser, and F. Mueller. 2016. Player Experience. In *Serious Games*, R. Dörner, S. Göbel, W. Effelsberg, and J. Wiemeyer (Eds.). Springer, 243–271.

55. G. N. Yannakakis and C. Alexopoulos. 2014. Mixed-Initiative Co-Creativity. In *Proc. Conf. FDG*.

56. G. N. Yannakakis and J. Togelius. 2011. Experience-Driven Procedural Content Generation. In *IEEE Trans. Affective Computing*. 147–161.

57. D. Yu. 2008. *Spelunky*. Game [PC]. (21 December 2008). Mossmouth.

58. D. Yu. 2016. *Spelunky*. Boss Fight Books.