Assessing the Effects of Interacting with MAP-Elites

Alberto Alvarez,[1] Jose Font,[2] Steve Dahlskog,[3] Julian Togelius[4]
[123]Game Lab, Computer Science and Media Technology Department, Malmö University, Sweden
[4]Game Innovation Lab, New York University, USA
{[1]alberto.alvarez,[2]jose.font,[3]steve.dahlskog}@mau.se
[4]julian@togelius.com

Abstract

MAP-Elites has been successfully applied to the generation of game content and robot behaviors. However, its behavior and performance when interacted with in co-creative systems is underexplored. This paper analyzes the implications of synthetic interaction for the stability and adaptability of MAP-Elites in such scenarios. We use pre-recorded human-made level design sessions with the Interactive Constrained MAP-Elites (IC MAP-Elites). To analyze the effect of each edition step in the search space over time using different feature dimensions, we introduce Temporal Expressive Range Analysis (TERA). With TERAs, MAP-Elites is assessed in terms of its adaptability and stability to generate diverse and high-performing individuals. Our results show that interactivity, in the form of design edits and MAP-Elites adapting towards them, directs the search process to previously unexplored areas of the fitness landscape and points towards how this could improve and enrich the co-creative process with quality-diverse individuals.

Introduction

Mixed-initiative co-creativity (MI-CC) (Yannakakis, Liapis, and Alexopoulos 2014), is a human-AI collaborative approach where both human and computer have a proactive role in the creation of content (Liapis, Yannakakis, and Togelius 2013b). Recent research shows the importance of using quality-diversity (QD) algorithms (Pugh, Soros, and Stanley 2016; Gravina et al. 2019) to better drive the evolutionary process in complex search spaces by generating stepping stones that barely resemble the optimal solution (Gaier, Asteroth, and Mouret 2019). A popular QD implementation in recent research is MAP-Elites (Mouret and Clune 2015), which has been applied to procedurally generate levels for bullet hell games (Khalifa et al. 2018), as well as dungeon rooms for mixed-initiative generation of adventure games (Alvarez et al. 2019), and levels for puzzle games (Charity, Khalifa, and Togelius 2020).

The rising interest of the evolutionary computation and computational intelligence in games research community in PCG, MI-CC and MAP-Elites, calls for improving the ways for evaluating these novel approaches. Some of the main problems in mixed-initiative tools are the inadequate consideration of the costs and benefits for every automated action, as well as failing to spot the opportunities for users to guide the invocation of the automated services (Horvitz 1999).

The area of eXplainable AI for Designers (XAID) (Zhu et al. 2018) strives for achieving system explainability necessarily built on *understandings of both algorithmic properties of the underlying AI techniques and the needs of human designers*. Similarly, Compton reflects on the grokloop (Compton 2019): the creative feedback MI-CC loop where a user builds a hypothesis, modifies a system, evaluates the result, and then updates the system. Shorter grokloops improve the overall performance of the mutual inspiration, and attempting to shorten this loop implies a clear understanding and interpretation of the relationship between each user action and the changes that it triggers in the system.

MAP-Elites have shown excellent results at generating QD individuals in games (Fontaine et al. 2020; Alvarez et al. 2020), and offline adaptation based on its generated repertoire (Cully et al. 2015; Gonzalez-Duque et al. 2020). However, MAP-Elites generation capabilities have been mostly evaluated in non-interactive scenarios and based on its final result even when used in interactive situations (Charity, Khalifa, and Togelius 2020; Alvarez et al. 2019). This results in a lack of research to assess the effects and consequences of interacting with MAP-Elites, and the adaptability and stability properties of MAP-Elites in dynamic scenarios.

Therefore, the contributions of this paper are two-fold. On the one hand, we present Temporal Expressive Range Analysis (TERA) as a novel way to analyze interactive PCG. TERAs allow us to inspect and analyze the changes in the expressive range over a defined period, which in our case, are design editions. On the other hand, using TERAs, we explored and analyzed how population dynamics react and adapt to constant changes in the IC MAP-Elites for level generation of 2D adventure games. IC MAP-Elites is evaluated using simulated pre-recorded design sessions with different design goals that display the algorithm's stability and adaptability properties and benefits. Our results show that IC MAP-Elites stably encounters high-performing solutions while adapting to changes in the design, and by doing this, regions of the search space which previously seemed inaccessible are opened for exploration.

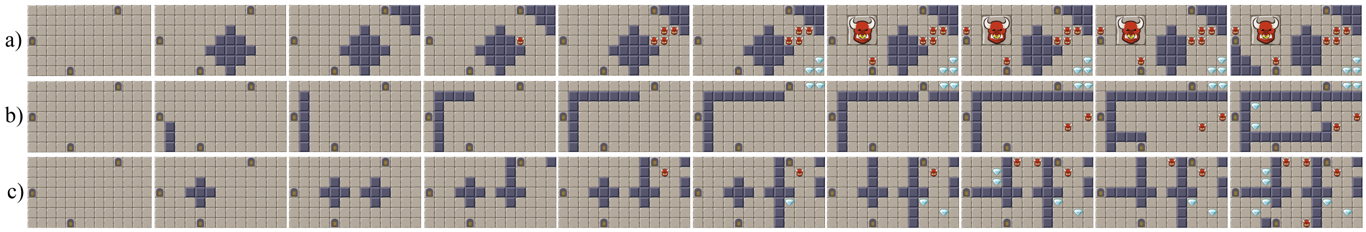

Figure 1: Sequences of rooms used in the three scenarios, targeting (a) low leniency, (b) high linearity, (c) and high meso-pattern concentration respectively. The leftmost and rightmost rooms correspond to the start and end rooms in each sequence, while intermediate steps are shown in between, limited to eight due to space restrictions.

Background

The Evolutionary Dungeon Designer (EDD) is a MI-CC tool to co-create 2D dungeons in the style of the seminal game *The Legend of Zelda* (Nintendo 1986). Designers manually edit the dungeon structure as well as the interior of every room in it. EDD constantly offers tailored room suggestions on the fly that designers may decide to incorporate to their designs at any moment.

In EDD, the system analyzes the level-design patterns (i.e., micro- and meso-patterns) that exist in each room, calculating and utilizing their quality to assess rooms. Micro-patterns are the building blocks in a design, which in EDD are categorized as *spatial micro-patterns*: chamber, corridor, intersections, connector; and *inventorial micro-patterns*: enemy, treasure, and door. On the other hand, Meso-patterns are defined as the relation between micro-patterns or other meso-patterns, and by the composition between inventorial micro-patterns and spatial micro-patterns. Meso-patterns are used to identify structures in the room that join together a set of micro-patterns and can be: *ambush*, *guard chamber*, *treasure chamber*, and *guarded treasure*. All patterns are shown in figure 2, and further information and discussion can be found in (Baldwin et al. 2017; Alvarez et al. 2018).

Methods for evaluating PCG

Recent research focuses on methods for evaluating procedural algorithms. The work in (Gravina, Liapis, and Yannakakis 2019) evaluates fitness, offspring and selection for five MAP-elite methods, whereas (Cook, Gow, and Colton 2016) shows how users can improve the generator with the aid of automatic parameter tuning and, consequently, evaluates the effect that it has on the generator. The work is continued in (Cook et al. 2019) where two analytical techniques, smoothness and co-dependence, were introduced in order to help analyze the impact of a parameter change and its effect on a generative system. Liapis et al. (Liapis, Yannakakis, and Togelius 2014) did a similar evaluation as the one in this paper, using artificial agents simulating designer's choices of suggestions to evaluate and display properties of their designer's style model.

Previously suggested evaluation methods include a top-down approach (Smith and Whitehead 2010; Horn et al. 2014), called the *expressive range analysis* (ERA) which refers to the idea of exploring and visualizing the content space. Summerville (Summerville 2018) proposes techniques for visually assessing and analyzing procedural systems, with other means of visual assessment including analysis of generative space and individual procedural artefacts.

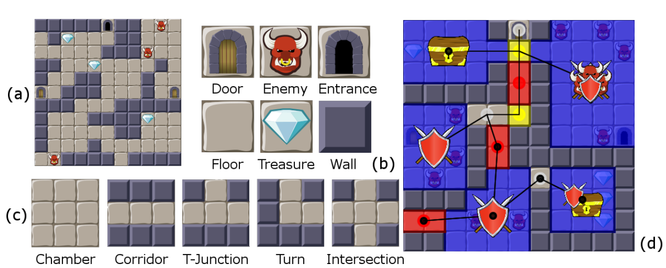

Figure 2: Main components in EDD. (a) Basic room, (b) different tiles, (c) micro-patterns and (d) meso-patterns

Variants of MAP-Elites

MAP-Elites, a quality-diversity (QD) algorithm, seeks to *illuminate* a *behavior space* by trying to find the best solutions across a feature-dimension grid (Mouret and Clune 2015). Some versions skip the grid in favour of voronoi tesselation to decide which elite individuals to keep in the map (Vassiliades, Chatzilygeroudis, and Mouret 2016). Other works combine the effective adaptive search of Covariance Matrix Adaptation Evolution Strategies with a map of elites, yielding large improvements for real-valued representations in terms of both objective value and number of elites discovered (Fontaine et al. 2020). ME-MAP-Elites (Cully 2020) creates a set of emitters to focus on different optimization processes that are active at different generations, generating higher performing and diverse individuals.

Constrained MAP-Elites (Khalifa et al. 2018) combines divergent search with a two-population approach to constraint satisfaction, taken from the FI-2Pop algorithm (Kimbrough et al. 2008). Constrained MAP-Elites has been used as the basis for subsequent experiments, e.g., to find sets of levels implementing diverse game mechanics (Charity et al. 2020). This algorithm was later combined with interactive evolution to yield the aforementioned Interactive Constrained MAP-Elites (Alvarez et al. 2020). Moreover, MAP-Elites has been shown to be robust at adapting to changing conditions after running the algorithm thanks to

Feature	Definitions
Similarity (Sim)	Aesthetic (tile-by-tile) similarity between a generated level and the designer's design.
Inner Similarity (IS)	Different tiles' sparsity and density similarity between a generated level and the designer's design.
Symmetry	Room's aesthetic symmetry.
Leniency (Len)	Challenge based on enemies and treasures.
Linearity (Lin)	Paths that exist connecting entry points in a level.
#Meso-Patterns (Meso)	Amount of meso-patterns that exist within a level. This is a discrete dimension rather than continuous.
#Spatial-Patterns (Spa)	Amount of spatial-patterns that exist within a level.

Table 1: Level design based dimensions used in EDD with IC MAP-Elites.

its generated behavioral repertoire. This was proposed and tested in the intelligent trial-and-error algorithm (Cully et al. 2015; Gonzalez-Duque et al. 2020). Related work extended MAP-Elites with *Adaptive Sampling* and *Drifting-Elites* to be more robust in noisy environments and domains where the fitness and behavior evaluation might be stochastic such as games (Justesen, Risi, and Mouret 2019).

Approach

IC MAP-Elites is a variation of MAP-Elites and Constrained MAP-Elites that incorporates adaptive mechanisms and implements continuous and interactive evolution. IC MAP-Elites uses an adaptive fitness function that continuously adapts from a user and enables users to flexibly change dimensions and cells' granularity at runtime (Alvarez et al. 2020). Our evaluation described in the following section is applied to EDD, which implements IC MAP-Elites, allowing us to evaluate the effects and dynamics of interacting with MAP-Elites.

EDD's IC MAP-Elites implementation uses a single-objective weighted fitness function with a FI2Pop genetic algorithm (Kimbrough et al. 2008). Individuals are deemed infeasible when they contain unreachable areas from any of the room's entry points and are evaluated based on how many unreachable tiles they have. Feasible individuals are evaluated as the following equally divided weighted sum:

$$f_{fitness}(r) = \frac{1}{2} f_{inventorial}(r) + \frac{1}{2} f_{spatial}(r) \quad (1)$$

This evaluation is adaptive, meaning that the tile's ratios, patterns, and balance between open areas and corridors are related to the target and collected by MAP-Elites after every modification to the target. $f_{inventorial}$ calculates the quality of all inventorial micro-patterns in relation to the current edited room, and $f_{spatial}$ calculates both the quality of spatial micro-patterns and the distribution and composition of tiles in the room described by the meso-patterns.

Seven level-design related feature dimensions are implemented in EDD. The designer can pick dimension pairs at a time and change the dimensions' granularity. When the designer changes dimensions, IC MAP-Elites seamlessly reshape the cells and move around the current elites, allowing the designer to switch between features to explore the search space. The seven features are briefly described in table 1, although an extensive discussion can be found in (Alvarez et al. 2020).

Experiment Setup

Dimensions	Low Leniency Scenario		
	◇	†	◯
Len-Meso	⋆56.6%	⋆19.46±5.55	**0.96±0.013**
Len-Spa	71.4%	⋆19.34±7.979	0.94±0.016
Lin-Meso	78.7%	**43.5±3.76**	0.92±0.013
Lin-Spa	⋆57.8%	28.28±4.109	0.92±0.011
Meso-Spa	**78.9%**	27.15±7.315	0.93±0.012
Sym-Len	68.1%	21.54±5.88	**0.97±0.014**
Sym-Spa	**83.4%**	24.87±10.334	0.95±0.014
Sim-Lin	58.4%	30.78±2.602	⋆0.9±0.012
Sim-Meso	62.3%	27.98±3.385	⋆0.9±0.014
Sym-IS	70.8%	21.63±3.877	**0.96±0.011**
Average.	67.09±3.35	25.75±2.48	0.93±0.01
Dimensions	**High Linearity Scenario**		
	◇	†	◯
Len-Spa	72.6%	⋆16.31±3.035	0.91±0.006
Lin-Meso	67.2%	**40.79±3.221**	0.85±0.015
Lin-Spa	62.7%	33.55±0.963	⋆0.84±0.013
Sym-Len	72.6%	20.5±2.646	**0.94±0.008**
Sym-Lin	**84.3%**	**37.94±2.103**	0.87±0.013
Sym-Spa	**85.5%**	18.94±4.366	0.92±0.011
Sim-Len	⋆58.7%	20.93±2.174	0.87±0.011
Sim-Lin	62.3%	27.78±1.424	⋆0.84±0.014
Sim-Meso	68%	25.6±1.946	⋆0.84±0.011
Sym-IS	82.8%	21.24±3.005	**0.94±0.011**
Average.	71.8±3.12	24.52±2.81	0.89±0.01
Dimensions	**High Meso-Pattern Scenario**		
	◇	†	◯
Len-Lin	⋆56.3%	32.49±1.536	0.92±0.007
Lin-Spa	⋆55.7%	30.47±1.959	0.9±0.007
Sym-Len	65.7%	⋆22.73±5.025	**0.96±0.009**
Sym-Lin	83.4%	**44.4±1.019**	0.93±0.007
Sym-Meso	83.6%	38.45±5.136	**0.96±0.01**
IS-Meso	**90.2%**	34.82±3.749	0.95±0.009
Sim-IS	59.9%	⋆21.16±2.237	0.91±0.009
Sim-Lin	66.3%	36.29±1.585	⋆0.89±0.008
Average.	72.34±4.12	29.88±2.84	0.93±0.01

Table 2: Results from the three metrics in all scenarios across the relevant dimension pairs. ◇ relates to coverage, † relates to average coverage per step, and ◯ average population fitness throughout all generations. Higher scores per column are highlighted in bold. ⋆ marks the lower values. Confidence intervals are shown for each average value (†, and ◯), and in the last row we show the average of all the 21 dimensions per metric and scenario.

We have conducted a series of experiments on EDD to analyze the *adaptability* and *stability* of IC MAP-Elites, as well as the effects of the interaction for MAP-Elites and the user. *Stability* relates to the steady generation of high-performing individuals, while gradually growing the search and stably covering the generative space at each edit step. *Adaptability* relates to the ability of the search to adapt to changing conditions, adjusting the search to the new goals, while still generating high-performing individuals. Both features, relate to the notions of evolvability (Doncieux et al. 2020), the ability of the search to generate creative individuals in problems with changing conditions.

We recorded three different design sessions, called scenarios, where we manually designed, step by step, dungeon rooms with specific target design goals. They are shown in Figure 1: a) a boss room - identified as a **low leniency** design goal; b) a linear room with specific paths and targets - identified as a **high linearity** design goal, and finally, c) a room where each chamber within is usable - identified as a **high meso-pattern** design goal. We chose these design goals as they represent key metrics with clear distinct representative goals and design styles that one might create in a dungeon.

The experiments consist on running these pre-recorded scenarios separately on EDD, step by step with a lapse of 100 evolutionary generations between steps. Each scenario implies 21 evolutionary runs, one per each pair of feature-dimension, where the dimensions are (table 1): leniency (len), linearity (lin), spatial patterns (spa), meso-patterns (meso), symmetry (sym), similarity (sim), and inner similarity (IS). Each edition (i.e., design step) is included as-is in the population and used as a target in the fitness function by IC MAP-Elites, updating corridor, chamber, and inventorial ratios affecting the quality of micro and meso patterns. Further, every 100 generations we gather population data related to novel generated individuals to later analyze how the search and the fitness landscape vary after each design step. Step after step, we measure how the explored generative space grows, as well as how distribution and concentration of elites together with the manually edited room traverse the generative space.

In all the experiments, the initial population was set to 1000 mutated individuals. All cells were set to a maximum capacity of 25 individuals each. In every generation, we selected 5 cells random, and 5 parents per cell through tournament selection. The random selection followed a uniform distribution. Offspring were produced through a two-point crossover and a 30% mutation chance. Using this setup, between 150 to 2001 individuals were produced every 100 generations, with an average of 373 unique individuals generated every 100 generations throughout all runs.

Metrics

All our experiments are evaluated and analyzed following the same procedure and metrics, focusing in the novel generated individuals. In particular, we calculated the *coverage* (◇), the *average coverage per step* (†), and *average fitness* (◯). *Coverage* relates to the percentage of space covered by the search in total and is calculated as the cumulative amount of covered hexagons at the final step divided by the maximum amount in our experiments. *Average coverage* relates to the average coverage per step, i.e., how much of the space is covered at each design step in average, calculated as the cumulative coverage per step divided by the amount of design editions. Finally, *average fitness* simply calculates the average individual fitness in the search throughout all generations.

Results and Analysis

Table 2 shows an extract of all test results. The results were filtered to only show the pair of dimensions with the lowest or highest values in different metrics. This means that some pair of dimensions are not shown since all their metric values were in between lower and higher scores. Each subtable represents one of the three design scenarios, each of them displaying the metrics described above. The higher and lower scores per column are highlighted in bold and with $\star$, respectively. Confidence intervals are shown per value when using averages. In general, table 2 shows IC MAP-Elites' stability to cover the space while encountering high-performing individuals; thus, it is able to adapt to the new editios which change the fitness landscape. Coverage (◇), supported by average coverage per step (†) and average fitness (◯), shows that in average the algorithm keeps exploring and generating novel and high-performing individuals rather than sporadically generating them.

The symmetry and spatial pattern (Sym-Spa) dimension pair explore and cover on average more of the space across all scenarios than others (◇). In general, when using either *sym* or *spa*, MAP-Elites is pressured to generate content that maximizes the utility of walls since both use walls as a core building block. Similarly, *Meso* and *IS* are other two dimensions that perform well with others, especially *Meso*. However, *Meso* requires the combination of shapes with walls and correct placement of individuals; thus, in principle involving more complex operations to achieve high dimensional values.

Overall the average population fitness (◯) is very high with 0.915 in average (of a maximum 1.0) with a narrow confidence interval ± 0.01. Symmetry and leniency (Sym-Len) scored the highest in all scenarios while Similarity and Linearity (Sim-Lin) scored the lowest. This shows a stable behavior in the IC MAP-Elites as that the average quality of individuals is high even when large portions of the generative space are explored (high diversity), and regardless of the dimension pair chosen.

IC MAP-Elites works in constant interaction with the edited design. Each step reshapes the fitness landscape to a certain extent, which could hinder the evolutionary process. However, per step, an avg. of 27.34% of the space is covered by generating novel individuals (i.e., not encountered previously in the population). This, coupled with the relatively narrow confidence intervals, means that the search is constantly exploring the space, diversifying and encountering new spaces. However, as it will be exemplified through the cases, the implicit explorative and exploitative mechanisms of MAP-Elites might not be enough to explore new areas, which can be introduced interactively to MAP-Elites.

The following cases examine TERAs following the different scenarios presented in figure 1. Different cases will use either an aggregated TERA step by step or a non-aggregated

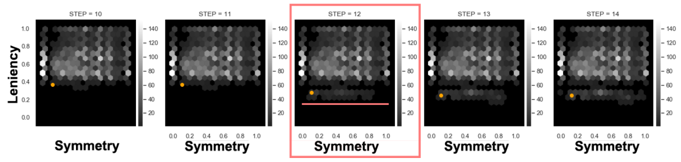

Figure 3: Aggregated TERA using Sym and Len following the low leniency scenario (fig. 1.a). Highlights how the design introduces new generative area to the algorithm. In red (step 12) it is highlighted when the design enters a new area of the space and MAP-Elites is then able to generate individuals in that area, explained in detail in Case 1.

version showing each step's specific scores in the evaluated dimensions. Each figure's caption and respective case will indicate what type of TERA is used. Non-Aggregated TERAs show the delta maps in the search, meaning where the search has focused and the space covered for a specific step. On the other hand, aggregated TERAs show the density over all steps of the generated individuals and the covered space up to the specific step. In each TERA, we also show as an orange dot where the current design is in relation to the used and tested dimensions. Case 1 and 2 examine aggregated TERAs, while case 3 examines non-agreggated TERAs.

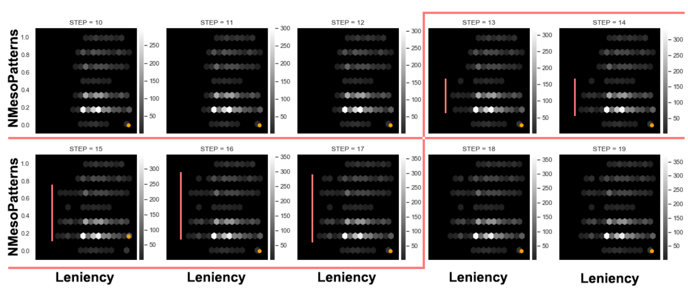

Figure 4: Aggregated TERA using Len and Meso following the high linearity scenario (fig. 1.b). Highlights subtle discovery of new generative areas. In red (step 12), this subtle discovery is highlighted, explained in detail in Case 2.

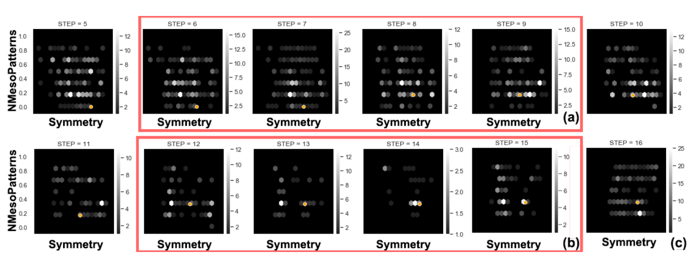

Figure 5: Non-aggregated TERA using Sym and Meso following the high meso-pattern level scenario (fig. 1.c). Highlights overall properties of interacting with MAP-Elites. Lighter areas identified with a, b, c, and d, represent the main areas of focus explained in detail in Case 3.

Case 1 - Design opens new areas of the generative space

In this case, we analyze the interaction with IC MAP-Elites through examining the generative space when using Sym and Len as dimensions (figure 3) following the low leniency scenario depicted in figure 1.a. The scenario aimed to gradually increase the room's challenge while dividing it into two clear and connected areas.

Table 2 shows that this pair of dimensions do not have the best scores except for the $\overline{fitness}$ (◯). This indicates that these dimensions are able to stably find high-performing individuals (above the avg.) while adapting to the new designs exploring an average of 21.54 and a total of 68.1%. Moreover, in fig 3, for several steps, half of the generative space is completely unexplored, which could indicate that in those areas, dimensions would be mutually exclusive.

However, at step 12 (highlighted in red in fig 3), when reaching a low leniency score, the design enters an unexplored area of the generative space, which subsequently enables IC MAP-Elites to search and generate high-performing individuals in the new region. Furthermore, there is a significant rise in the number of unique individuals generated with a high concentration on the new area, spreading over the already explored space.

Case 2 - Subtle changes in the design reflected in the generation of MAP-Elites

Case 1 showed that by entering an unexplored area of the generative space, the designer could show possibilities for the algorithm and influence the search, yet more subtle guidance is possible. Figure 4 presents such a case. We focus on the high linearity scenario (fig. 1.b), where the goal was to create a single narrow corridor between the top and left door and add some objective at the bottom entrance. Through this, we not only aimed at high linearity but also tried to promote other main characteristics such as the open chamber/corridor balance, or the combination of meso-patterns.

Similar to the previous case, in figure 4, it is shown that for many steps, the search does not explore a big part of the generative space. In this case, the room does not move either in the generative space as the changes are not affecting the dimensions. However, between steps 13 to 17, a new area of the generative space is filled (highlighted in red in fig 4), which indicates that even if changes in the room do not have a direct influence by moving in the generative space, they still can foster exploration in new areas. These main steps are visible in fig. 1.b, subfigure 5, and 6 from the left. Specifically, lower leniency areas are generated once the room is strongly divided into a representative corridor and a big open area with a treasure meso-pattern. These stepping-stones gave the needed "building blocks" to MAP-Elites to cross and mutate until the new generative space was explored. Moreover, similarly to the previous case, the algorithms generates significantly more novel individuals during this time (around 531 novel individuals), and the search covers 20% of the space, exploiting the new region, akin to Novelty search behavior (Liapis, Yannakakis, and Togelius 2015).

Case 3 – Exploring multiple properties

In this case, we analyze the TERA of unique individuals generated each step using Sym and Meso as dimension pairs (fig. 5). We heed to the high meso-pattern level scenario (fig. 1.c), where we subdivided the room into small open chambers with clear objectives. Overall, in fig. 5 two aspects stand out; firstly, the generative space is explored more at the early steps, as there are fewer constraints from the edited design. Secondly, the generated individuals seem to follow the path taken by the design in the generative space. Supported by the other cases, this indicates that the design can filter the search and point areas of interest for the IC MAP-Elites.

Moreover, opposite to case 1 and as a consequence of filtering the generative space, when the design leaves low scoring areas, the algorithm rapidly disregards creating individuals in those spaces. For instance, the bottom area after step 7 (fig. 5.a). Further, as the room is changing but without any type of influence in the searched dimensions e.g., steps 12 to 15 (fig. 5.b), MAP-Elites has difficulties exploring the generative space. A similar challenge was found and discussed by Alvarez et al. (Alvarez et al. 2020), where their experiments showed that the search got to a plateau after 1000 generations due to the MAP-Elites lacking the incentive to explore. Even if their experiments focused on static environments, this case gives further evidence that minimal changes to the design and lack of influence in the generative space, conditions the exploration of space and the generation of novel individuals.

Lastly, at step 16 (fig. 5.c), we encounter a similar situation as with fig. 4; where a design edition enables the needed "building blocks" for MAP-Elites. In this case, it was triggered by the forming of a dead-end chamber pattern, which enables even more meso-patterns to be used and discovered. Rather than finding a new area of the generative space, this time, the search gets rebooted, and therefore, explores all areas generating novel individuals.

Discussion

Our evaluation shows that IC MAP-Elites has a high degree of adaptability to dynamic environments, adapting the generated content to the design process and design goals while stably generating high-performing and diverse solutions. For MI-CC systems and interactive approaches as in EDD, this is especially relevant and important. The fitness function adapts to the current design; thus, adaptability and stability go hand in hand. Furthermore, the deployment of an MI-CC approach in a scenario such as the ones presented would benefit both Map-Elites and the human designer. On the one hand, it enables MAP-Elites to explore more of the generative space while producing quality solutions. On the other hand, users would have more control over the suggestions as they influence and guide the search and generation similar to (Anderson et al. 2000), but more seamlessly.

We also observe that when using Linearity as a dimension, IC MAP-Elites performed quite stably in all our scenarios regardless of the design traversing around the generative space or not, which indicates that Linearity is more robust and stable and more agnostic and independent from the design. These characteristics are beneficial in certain cases, but based on the results presented in table 2, this stability comes at the expense of adaptability and higher fitness scores.

Furthermore, Alvarez et al. (Alvarez et al. 2020) presented an analysis of IC MAP-Elites in a static scenario, where on average the covered space of MAP-Elites after 5000 generation was 52.4% using pair of dimensions and 51.7% using all seven dimensions. Our results show a clear advantage for MAP-Elites when used continuously and interactively with an avg. coverage of 70.9%. However, when the design remains still in the space defined by the dimensions, exploration is hindered; thus, what dimensions are used and how the design maps to them is crucial.

Finally, we used and introduced TERAs to analyze the dynamic behaviors in generative systems and algorithms, and observe the effects of changes over time in the expressive range based on the edition steps. We used two variations, non-aggregated TERA, which shows the delta maps of the search, and aggregated TERA, showing the search density and aggregated results. TERAs are generic and could be used with other generative system to evaluate their dynamics by simply defining a pair of features and a step period such as design editions, amount of generations, or whenever a suggestion is applied. TERAs can also be used to spot key and non-trivial steps or changes that have an effect in the search, which can help to understand more in-depth the sensibility of the algorithm and the system.

Conclusions and Future Work

This paper analyzes and evaluates the benefits of dynamically interacting with quality-diversity algorithms, specifically, the IC MAP-Elites. We have examined the adaptability and stability of MAP-Elites in relation to 21 dimension pairs highlighting different characteristics and properties through different simulated design scenarios. We examined key metrics when exploring the generative space, as depicted in table 2, and conducted three different case studies that highlighted different dynamics with the algorithm.

While our results show several MAP-Elites' properties and promising ways to improve the MI-CC workflow, further evaluation is needed with human users to assess these properties in-the-wild and evaluate more in-depth the interactive dynamic between humans and algorithms. To further highlight the importance of interaction, it would be interesting to analyze and compare with both MAP-Elites disabling adaptive mechanisms (i.e., rendering the algorithm static and agnostic to changes) and with other non QD algorithms. Likewise, another interesting project for future work, would be to evaluate and compare IC MAP-Elites using TERAs with other co-creative systems using different algorithms such as reinforcement learning (Delarosa et al. 2020; Guzdial, Liao, and Riedl 2018) or constraint solving algorithms (Karth and Smith 2019).

Finally, a promising step is to analyze MAP-Elites together with surrogate models that capture the preference, style, and process of designers (Liapis, Yannakakis, and Togelius 2013a; Alvarez and Font 2020; Alvarez, Font, and Togelius 2020), and how these influence the properties discussed in this paper. For instance, Designer Personas (Alvarez, Font, and Togelius 2020) could be used to explore how the user's design moves through the space, identifying possible paths, and analyzing if key changes, i.e., moving between style clusters, connect to key moments in the MAP-Elites generation.

References

Alvarez, A.; Dahlskog, S.; Font, J.; Holmberg, J.; Nolasco, C.; and Österman, A. 2018. Fostering Creativity in the Mixed-initiative Evolutionary Dungeon Designer. In *Proceedings of the 13th International Conference on the Foundations of Digital Games*, FDG '18.

Alvarez, A.; Dahlskog, S.; Font, J.; and Togelius, J. 2019. Empowering Quality Diversity in Dungeon Design with Interactive Constrained MAP-Elites. In *2019 IEEE Conference on Games (CoG)*.

Alvarez, A.; Dahlskog, S.; Font, J.; and Togelius, J. 2020. Interactive Constrained MAP-Elites: Analysis and Evaluation of the Expressiveness of the Feature Dimensions. *IEEE Transactions on Games* 1–1. doi:10.1109/TG.2020.3046133.

Alvarez, A.; and Font, J. 2020. Learning the Designer's Preferences to Drive Evolution. In *Applications of Evolutionary Computation*, volume 12104 of *Lecture Notes in Computer Science*, 431–445. Cham: Springer International Publishing.

Alvarez, A.; Font, J.; and Togelius, J. 2020. Designer Modeling through Design Style Clustering. *Arxiv* Preprint available: https://arxiv.org/abs/2004.01697.

Anderson, D.; Anderson, E.; Lesh, N.; Marks, J.; Mirtich, B.; Ratajczak, D.; and Ryall, K. 2000. Human-guided simple search. In *AAAI/IAAI*, 209–216.

Baldwin, A.; Dahlskog, S.; Font, J. M.; and Holmberg, J. 2017. Towards Pattern-based Mixed-initiative Dungeon Generation. In *Proceedings of the 12th International Conference on the Foundations of Digital Games*, FDG '17. New York, NY, USA: ACM.

Charity, M.; Green, M. C.; Khalifa, A.; and Togelius, J. 2020. Mech-Elites: Illuminating the Mechanic Space of GVG-AI. In *International Conference on the Foundations of Digital Games*, 1–10.

Charity, M.; Khalifa, A.; and Togelius, J. 2020. Baba is Y'all: Collaborative Mixed-Initiative Level Design. In *2020 IEEE Conference on Games (CoG)*, 542–549. doi:10.1109/CoG47356.2020.9231807.

Compton, K. 2019. Casual Creators: Defining a Genre of Autotelic Creativity Support Systems. PhD Dissertation, University of California Santa Cruz.

Cook, M.; Colton, S.; Gow, J.; and Smith, G. 2019. General Analytical Techniques For Parameter-Based Procedural Content Generators. In *2019 IEEE Conference on Games (CoG)*. ISSN 2325-4270. doi:10.1109/CIG.2019.8848024.

Cook, M.; Gow, J.; and Colton, S. 2016. Towards the automatic optimisation of procedural content generators. In *2016 IEEE Conference on Computational Intelligence and Games (CIG)*. ISSN 2325-4289. doi:10.1109/CIG.2016.7860424.

Cully, A. 2020. Multi-Emitter MAP-Elites: Improving quality, diversity and convergence speed with heterogeneous sets of emitters.

Cully, A.; Clune, J.; Tarapore, D.; and Mouret, J.-B. 2015. Robots that can adapt like animals. *Nature* 521(7553): 503–507. ISSN 1476-4687. doi:10.1038/nature14422.

Delarosa, O.; Dong, H.; Ruan, M.; Khalifa, A.; and Togelius, J. 2020. Mixed-Initiative Level Design with RL Brush.

Doncieux, S.; Paolo, G.; Laflaquière, A.; and Coninx, A. 2020. Novelty Search makes Evolvability Inevitable. In *Proc. of the Genetic and Evolutionary Computation Conf.*, GECCO '20.

Fontaine, M. C.; Togelius, J.; Nikolaidis, S.; and Hoover, A. K. 2020. Covariance Matrix Adaptation for the Rapid Illumination of Behavior Space. In *Proceedings of the 2020 Genetic and Evolutionary Computation Conference*, GECCO '20, 94–102. ISBN 9781450371285. doi:10.1145/3377930.3390232.

Gaier, A.; Asteroth, A.; and Mouret, J.-B. 2019. Are quality diversity algorithms better at generating stepping stones than objective-based search? In *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, 115–116. ACM.

Gonzalez-Duque, M.; Berg Palm, R.; Ha, D.; and Risi, S. 2020. Finding Game Levels with the Right Difficulty in a Few Trials through Intelligent Trial-and-Error. In *Proceedings of the IEEE Conference on Games*.

Gravina, D.; Khalifa, A.; Liapis, A.; Togelius, J.; and Yannakakis, G. N. 2019. Procedural Content Generation through Quality Diversity. In *2019 IEEE Conference on Games*.

Gravina, D.; Liapis, A.; and Yannakakis, G. N. 2019. Blending Notions of Diversity for MAP-elites. In *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, GECCO '19, 117–118. New York, NY, USA: ACM. doi:10.1145/3319619.3321987.

Guzdial, M.; Liao, N.; and Riedl, M. 2018. Co-Creative Level Design via Machine Learning. In *Proceedings of the 5th EXAG Workshop at AIIDE*.

Horn, B.; Dahlskog, S.; Shaker, N.; Smith, G.; and Togelius, J. 2014. A Comparative Evaluation of Procedural Level Generators in the Mario AI Framework. In *Proceedings of the 9th International Conference on the Foundations of Digital Games*, FDG '14.

Horvitz, E. 1999. Principles of Mixed-Initiative User Interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '99, 159–166. New York, NY, USA: Association for Computing Machinery. ISBN 0201485591. doi:10.1145/302979.303030.

Justesen, N.; Risi, S.; and Mouret, J.-B. 2019. MAP-Elites for Noisy Domains by Adaptive Sampling. In *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, GECCO '19, 121–122. New York, NY, USA: Association for Computing Machinery.

Karth, I.; and Smith, A. M. 2019. Addressing the Fundamental Tension of PCGML with Discriminative Learning. In *Proceedings of the 14th International Conference on the Foundations of Digital Games*, FDG '19. New York, NY, USA: Association for Computing Machinery. ISBN 9781450372176. doi:10.1145/3337722.3341845.

Khalifa, A.; Lee, S.; Nealen, A.; and Togelius, J. 2018. Talakat: Bullet hell generation through constrained map-elites. In *Proceedings of The Genetic and Evolutionary Computation Conference*. ACM.

Kimbrough, S. O.; Koehler, G. J.; Lu, M.; and Wood, D. H. 2008. On a Feasible–Infeasible Two-Population (FI-2Pop) genetic algorithm for constrained optimization: Distance tracing and no free lunch. *European Journal of Operational Research* 190(2): 310–327.

Liapis, A.; Yannakakis, G.; and Togelius, J. 2013a. Designer modeling for personalized game content creation tools. In *Artificial Intelligence and Game Aesthetics - Papers from the 2013 AIIDE Workshop, Technical Report*, volume WS-13-19, 11–16. AI Access Foundation. ISBN 9781577356349.

Liapis, A.; Yannakakis, G. N.; and Togelius, J. 2013b. Generating Map Sketches for Strategy Games. In *Proceedings of Applications of Evolutionary Computation*, volume 7835, LNCS. Springer.

Liapis, A.; Yannakakis, G. N.; and Togelius, J. 2014. Designer modeling for Sentient Sketchbook. In *2014 IEEE Conference on Computational Intelligence and Games*, 1–8. doi:10.1109/CIG.2014.6932873.

Liapis, A.; Yannakakis, G. N.; and Togelius, J. 2015. Constrained Novelty Search: A Study on Game Content Generation. *Evolutionary Computation* 23(1): 101–129.

Mouret, J.-B.; and Clune, J. 2015. Illuminating search spaces by mapping elites. *preprint arXiv:1504.04909* .

Nintendo. 1986. The Legend of Zelda.

Pugh, J. K.; Soros, L. B.; and Stanley, K. O. 2016. Quality diversity: A new frontier for evolutionary computation. *Frontiers in Robotics and AI* 3: 40.

Smith, G.; and Whitehead, J. 2010. Analyzing the Expressive Range of a Level Generator. In *Proceedings of the 2010 Workshop on Procedural Content Generation in Games*. New York, NY, USA: ACM. ISBN 978-1-4503-0023-0. doi:10.1145/1814256.1814260.

Summerville, A. 2018. Expanding Expressive Range: Evaluation Methodologies for Procedural Content Generation. In *Proceedings of the 14th AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*.

Vassiliades, V.; Chatzilygeroudis, K. I.; and Mouret, J. 2016. Scaling Up MAP-Elites Using Centroidal Voronoi Tessellations. *CoRR* abs/1610.05729.

Yannakakis, G. N.; Liapis, A.; and Alexopoulos, C. 2014. Mixed-Initiative Co-Creativity. In *Proceedings of the 9th Conference on the Foundations of Digital Games*.

Zhu, J.; Liapis, A.; Risi, S.; Bidarra, R.; and Youngblood, G. M. 2018. Explainable AI for Designers: A Human-Centered Perspective on Mixed-Initiative Co-Creation. In *2018 IEEE Conference on Computational Intelligence and Games (CIG)*, 1–8. ISSN 2325-4270. doi:10.1109/CIG.2018.8490433.