Mixed-initiative generation of virtual worlds
A comparative study on the cognitive load of WFC and HSWFC

Alaka, Shaad; Bidarra, Rafael

DOI
10.1145/3649921.3659852

Publication date
2024

Document Version
Final published version

Published in
FDG '24

Citation (APA)
Alaka, S., & Bidarra, R. (2024). Mixed-initiative generation of virtual worlds: A comparative study on the cognitive load of WFC and HSWFC. In G. Smith, J. Whitehead, B. Samuel, K. Spiel, & R. van Rozen (Eds.), *FDG '24: Proceedings of the 19th International Conference on the Foundations of Digital Games* Article 75 Association for Computing Machinery (ACM). https://doi.org/10.1145/3649921.3659852

Mixed-initiative generation of virtual worlds - a comparative study on the cognitive load of WFC and HSWFC

Shaad Alaka
s.alaka@tudelft.nl
Delft University of Technology
Delft, The Netherlands

Rafael Bidarra
R.Bidarra@tudelft.nl
Delft University of Technology
Delft, The Netherlands

ABSTRACT

Procedural Content Generation (PCG) has long been proposed as an answer to the increased workload imposed on designers of virtual worlds, although often at the cost of controllability and expression of designer intent. Mixed-initiative approaches promise to overcome this, and valid proposals have been made of mixed-initiative editors, e.g. driven by the Wave Function Collapse (WFC) algorithm [12]. However, stock WFC operates on a simple tileset without any hierarchy or semantic clustering, preventing designers from fluently expressing level of detail, leaving the constraint solving burden partly on them instead of on the algorithm. Hierarchical Semantic Wave Function Collapse (HSWFC) claims to solve this problem by hierarchically structuring its tileset of semantic tiles, featuring meta-tiles that can undergo intermediate collapses [2]. Employing an HSWFC interactive editor offering such features may significantly reduce cognitive load and, thus, make such an editor more appealing for widespread use. We describe a user study to test this hypothesis, by comparing and discussing the cognitive load assessed on designers using either a stock WFC-driven editor or an HSWFC-driven editor. Our findings confirm that there is a significant reduction in cognitive load when the HSWFC editor is employed in comparison to the stock WFC editor.

CCS CONCEPTS

• **Computing methodologies → Computer graphics**; • **Human-centered computing → Collaborative interaction**.

KEYWORDS

procedural content generation, wave function collapse, world editing, mixed-initiative, cognitive load

ACM Reference Format:
Shaad Alaka and Rafael Bidarra. 2024. Mixed-initiative generation of virtual worlds - a comparative study on the cognitive load of WFC and HSWFC. In *Proceedings of the 19th International Conference on the Foundations of Digital Games (FDG 2024), May 21–24, 2024, Worcester, MA, USA.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3649921.3659852

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

ACM ISBN 979-8-4007-0955-5/24/05
https://doi.org/10.1145/3649921.3659852

1 INTRODUCTION

Using PCG to decrease the effort involved with creating virtual worlds for games, movies and the like has been commonplace over the past years. With that however, the limitations imposed on controllability and designer intent also have become apparent. Mixed-initiative (MI)-PCG aims to bridge the gap between arduous minute work and the loss of direct control as a result of using a PCG algorithm. For this, it offers a reduction in effort from PCG, and the precise control from manually sculpting a world. PCG in this sense acts as an *amplifier*, boosting the output of the designer, but ultimately it is the designer who is still in full control, at least within the context of the expressive power of the algorithm.

A particular class of PCG algorithms have proven to be very promising for MI application, due to their intuitive entry points for manual input, built-in strong adherence to learned or user-defined constraints, and real-time performance. These are algorithms that descend from the WFC family, which are essentially constraint solvers that operate on a graph. Often, the representation of choice for this graph is that of a grid, which can be interpreted as a tile canvas for both the designer and the algorithm to paint on. Tiles with constraints pre-defined among them are placed on the grid, and their restrictions get propagated, resulting in an output that always satisfies all constraints among tiles (or no output at all, in the case of a conflict).

This idea of using WFC in a MI setting was adopted and tested in several previous works (miWFC, several other editors built by Maxim Gumin, Adam Newgas and others) [5, 12, 17], although widespread adoption by the industry remains limited. While on paper the MI adaptation of WFC sounds promising, the experience shows that the burden of solving constraints is then largely passed onto the designers, who have to watch over preserving detailed tile semantics, leading to an increase in cognitive load [1, 12].

As a reaction to this, the observation was made that designers possibly do not want to paint in exact tiles, but rather in unconsolidated semantic concepts, much akin to making a rough sketch or ground plan [22, 27]. This coincides with the perspective proposed by HSWFC [2], which claims to be able to reduce cognitive load by largely taking away from the designer the burden of constraint solving and forced exactness, thus improving a weakness of MI-PCG with WFC-type algorithms.

In this work we compare the cognitive load of using WFC vs. HSWFC for mixed-initiative generation of virtual worlds. For this, we designed a user study that deploys a 2D editor offering the respective features of either algorithm. Participants were given a world building task, after which they answered a questionnaire assessing cognitive load and general usability.

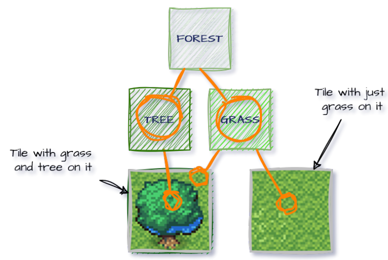


Figure 1: Meta-tiles T_{FOREST}, T_{TREE} and T_{GRASS} represent the semantic traits present in other tiles; the remaining tiles, known as terminal tiles, are the manifestation of these traits as concrete instances.

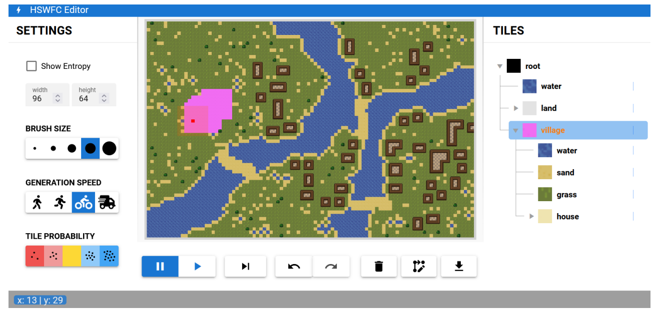


Figure 2: Interface of the HSWFC-driven web editor. At the right-hand side, it offers the hierarchical tileset with meta-tiles, while the stock WFC version of the editor presents a simple flat tileset.

2 RELATED WORK

In 2010, Paul Merrel published the Model Synthesis algorithm [15], which can be considered the main predecessor of the WFC algorithm, of which the initial implementation was unleashed by Maxim Gumin in 2016 [5]. Since then, WFC has had a considerable impact on researchers, technical artists and game developers, getting adopted, adapted and used in commercially published projects (Caves of Qud, Townscaper, Matrix Awakens), and in many research projects. The repository has become a hub of anything related to WFC, linking to research, derived works, alternative implementations, etc. [3, 8, 9, 11, 14, 19, 21].

Recently, the HSWFC algorithm was published, aiming at improving the MI usability of Wave Function Collapse (stock WFC)-based world editors, and providing a structure to work with semantics [1, 2]. Its main contribution was the proposal of *meta-tiles*, a new type of tile that can semantically represent other tiles. This allows one to move from the simple tileset of WFC into the hierarchical and more meaningful tile structure of HSWFC. This new tile structure is known as the meta-tree, and always has a single root meta-tile known as T_{ROOT}, from which all other tiles eventually descend. A canvas cell that collapses into a meta-tile can collapse again, until it eventually collapses into one of its descendent terminal tiles (comparable to the tiles used in stock WFC). This construct allows, for example, a "village" tile to collapse into a "house" tile (while also having had the choice to collapse into "grass", "road", "tree", etc) and a "house" tile to then collapse into a concrete "floor" or "wall" terminal tile. An example with a "forest" meta-tile is given in Figure 1, representing part of a meta-tree.

Adding mixed-initiative interactivity to WFC to make the generation process more spatially controllable has been proposed with miWFC, an interactive editor that allows you to place/overwrite tiles with a brush, create snapshots, regenerate marked parts, and spatially modify the tile weights [12].

Other types of mixed-initiative interactivity have been proposed for WFC that do not involve spatial control. Karth and Smith [10], for example, propose to allow the designer to intuitively adapt constraints by providing positive and negative examples of tile combinations; to fulfil them, a back-and-forth process progresses towards the generated result.

There are also applications of mixed-initiative WFC interactivity within a game. In Townscaper, for example, Oskar Stålberg shows the concept of letting users interactively collapse specific parts of the grid, while letting WFC choose the appropriate tile for the context [24]. Wildtile is a recent Unity 3D world editing asset powered by WFC that attempts to turn Townscaper's methodology into a powerful world authoring tool[1]. Adam Newgas also explored the concept of being able to edit WFC grids interactively, and built a proof of concept editor, while also briefly touching upon how this can be implemented using 'driven WFC', as used in Townscaper [16, 18]. All these approaches demonstrate the flexibility of WFC when it comes to involving humans in the generative process.

3 METHODOLOGY

Finding the difference in cognitive load between stock WFC and HSWFC requires interactive editor applications implementing the algorithms, and a user study to measure the cognitive load of their use. In this section, we present both the editor and the setup used in this study.

3.1 The editor

A general purposes 2D tile editor from the original HSWFC project is available [1]. It offers both HSWFC[2] and stock WFC[3] under the same interface, with HSWFC unlocking a range of unique editing facilities.

We therefore opted to use this editor, provided in the form of a web application; see Figure 2. This makes it very accessible for users, as it runs on any modern browser and requires no application installation. This editor comes with a set of basic editing facilities that are generally expected in an interactive editor:

- Painting, erasing and replacing tiles on the canvas
- The ability to undo and redo operations
- Saving and loading a canvas state (snapshotting), including clearing the canvas
- Visualizing the brush, and changing its size

Furthermore, some interactive control is provided on the PCG algorithm, so that the user can steer the generation process:

- Starting and stopping the PCG

[1]https://assetstore.unity.com/packages/tools/level-design/wildtile-215355
[2]HSWFC editor: https://archer6621.github.io/hswfc-editor-b/
[3]stock WFC editor: https://archer6621.github.io/hswfc-editor-a/

- Generating a fixed amount of tiles at once
- Altering the generation speed
- Adjusting tile generation weights

The mixed-initiative characteristic here is noteworthy because both the user and the algorithm work in the same (output canvas) space. Therefore, features such as undo/redo also apply to the PCG algorithm, and the user can freely manipulate anything that gets generated as if they had painted it themselves. This also means that the user and the PCG algorithm can almost simultaneously work together. In this setting, while the algorithm is running and simply collapses any cell that is eligible, the user can steer the generation process by e.g. erasing or replacing output cells that are deemed undesirable. This way of working allows users to focus on the design areas which they are explicitly concerned with, while letting the algorithm fill in the surroundings.

As mentioned above, the HSWFC version of the editor unlocks some unique editing facilities; see Figure 3. These include the ability to paint with meta-tiles, essentially empowering the user with semantic mark-up, which can be used to induce the algorithm towards the desired concrete result. Furthermore, probabilities can now be tweaked hierarchically per meta-tile instead of globally, allowing one to precisely control tile distributions. Another editing facility in the HSWFC editor, which cannot be supported under stock WFC, is *collapse-path resetting*, which is part of a more general class of new facilities that involve querying and editing using collapse paths. This particular feature allows the user to target all tiles that descend from a given meta-tile, and reset them all back to that meta-tile; this issues the algorithm to re-collapse those tiles, which may result in a different terminal tile configuration. This is very convenient for generating variants of explicit semantic constructs, e.g. villages.

3.2 User study

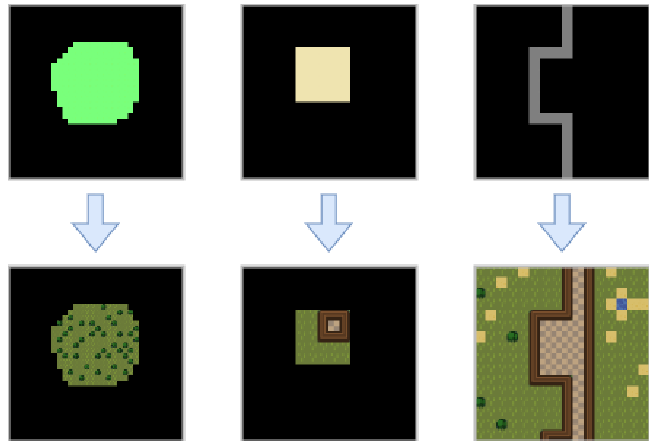
Participants were gathered through the following methods:

- Posting in various game technology communities, among which the FDG 2023 discord channel, the TU Delft CGV group, and the Dutch Games Industry slack server.
- Contacting game development companies and asking them to forward the user study to their environment designers.
- Attending a variety of game developer oriented gatherings and approaching people there, such as the Dutch Game Garden networking lunches.
- Approaching friends and acquaintances either directly or via email.

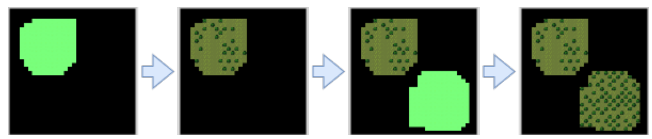
To compare the cognitive load of using stock WFC vs. HSWFC for mixed-initiative generation of virtual worlds, we conducted a blind A/B test, with the hypothesis that there will be a significant difference in the cognitive load measured for group A using stock WFC, and for group B using HSWFC.

In the experiment, participants of both groups need to be given the same task, for which we devised the following reasonable conditions:

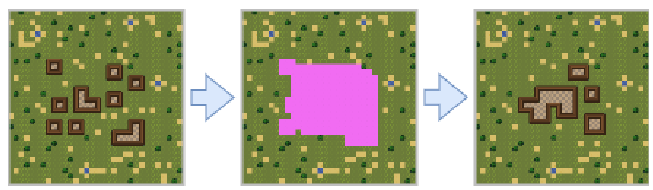
- The task should be complex enough, so that, if desired, a participant in group B can use the additional editing facilities.
- The task should not be too lengthy, in order to minimize participant dropout.



(a) Meta-tile painting: forest (left), house (center), wall (right)



(b) Tweaking tree weights within a forest



(c) Village regeneration, by collapse path resetting

Figure 3: Several unique HSWFC editing facilities, with (a) various examples of meta-tile painting, (b) hierarchical probability tweaking of trees in a forest, and (c) collapse-path resetting, where a village gets regenerated with different tile probabilities.

- Some parts of the task should be left open to interpretation, in order to allow the user to explore the editor on their own creative terms.

We, therefore, asked participants the following task: *create a little world with two villages adjacent to a body of water and some forest and a sandy path that connects the villages, with some additional constraints and properties* (full details can be found in Appendix B). They were given the role of "level designer" for some game under development, receiving a task from the "producer" to create a game world, which should instigate some sense of purpose. After creating the world, they were free to make some variants of it according to their own creative touch, in order to explore the editor a bit more and get a more solid opinion on it.

For this experiment, the tileset presented in Appendix A was used, as it was appropriately detailed for the task described above.

After being done with the tasks, the participants filled in a modified NASA TLX survey on a 7-point scale, which is used to measure the overall cognitive load [6]. The survey was modified, suppressing

ID	Question
Q1	I felt that the editor was able to capture my intent
Q1	I felt like I had the freedom to tweak things easily
Q3	I understood why certain tiles could not be placed in certain locations
HSWFC Q1	I found using the hierarchy for selecting a tile to paint with to be intuitive
HSWFC Q2	Using a single situation-dependent brush for painting and erasing felt intuitive
HSWFC Q3	Painting with meta tiles gave me the results I expected
HSWFC Q4	Adjusting the meta-tile probabilities had the results I expected
HSWFC Q5	I found using the regeneration tool useful for creating variations of my design

Table 1: Additional questions asked to the groups about the editor. Both groups were asked the first three questions, while the remaining questions were only for group B.

the 'physical effort' rubric, as it was deemed irrelevant for the task at hand.

Beyond measuring cognitive load, some additional questions were asked to all participants, focusing on the user experience of the editor itself; in addition, for group B, using the HSWFC editor, we also asked how intuitive the new editor facilities were. The virtual worlds designed by the participants were collected as well, mostly to find qualitative differences between environments created in the two groups. For details regarding this survey, refer to Appendix B.

4 RESULTS

The user test had a total of 18 participants, 9 per group. While the web app provided mobile support, all participants used a desktop. Before the user study, participants were asked about their relation and experience with PCG; see Figure 4. Many of the participants were programmers, though the ratio of programmers and artists/designers was approximately equal between the two groups. The testers were predominantly male, with ages varying between 20 and 50 years old.

The resulting aggregation of the scores for the NASA TLX rubrics can be inspected in Figure 5. A Monte Carlo hypothesis test was performed (significance level 5%, using skew as test statistic) to confirm that none of the resulting distributions are significantly different from a normal distribution. Along with that, the statistical significance of the results with respect to the comparison between the groups was evaluated using Student's t-test [23] and can be found in Table 2, obtained using the independent t-test implemented by `scipy`[4], which tests for the null hypothesis; whether the 2 independent collections of samples (in our case, group A and group B) have identical expected values. Rejection of this hypothesis means that the two groups are potentially different. All the metrics with p-values lower than 0.05 can be assumed to be significantly different between the two groups: these are 'Performance', 'Effort' and 'Frustration'.

Overall, the NASA TLX results show that group B felt significantly more successful in their performance, had to exert significantly less effort, and experienced significantly less frustration while using the HSWFC-driven editor. Temporal demand and mental demand show hints of being improved as well, but the pool of participants in the user study was not large enough to be conclusive about the significance of these differences.

Rubric	t(16) statistic	p value	Significant
Mental Demand	2.08514	0.05343	Doubt
Temporal Demand	2.01246	0.06132	Doubt
Performance	-4.00000	0.00103	Yes
Effort	2.60412	0.01918	Yes
Frustration	3.45218	0.00328	Yes

Table 2: The results of performing the independent t-test for the NASA TLX survey.

[4]https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.ttest_ind.html

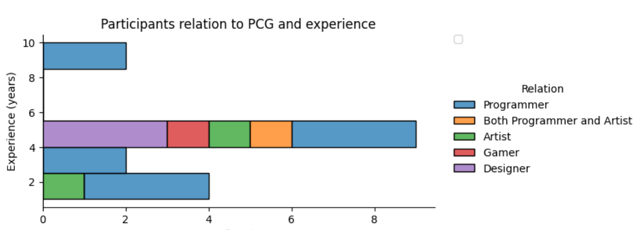


Figure 4: The distribution of the participants' relation and experience with PCG.

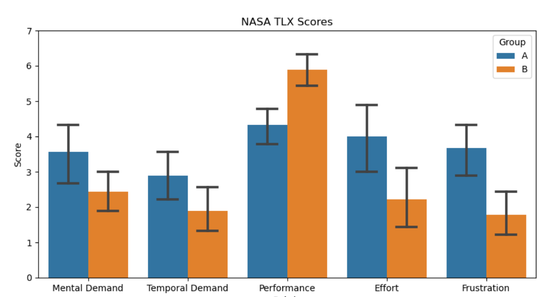


Figure 5: The NASA TLX results aggregated into a bar plot.

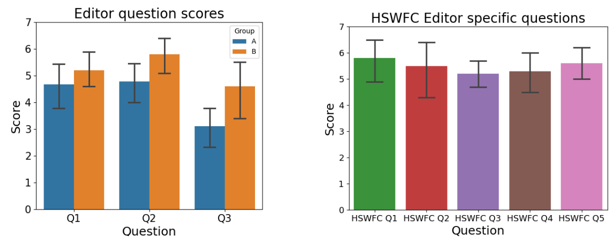


Figure 6: The resulting scores for the editor questions aggregated into two bar plots. Left: common editor questions asked to both groups; Right: questions specifically asked to group B, tailored at the HSWFC feature set.

Question	t(16) statistic	p value	Significant
Q1	-1.20605	0.24533	Unlikely
Q2	-1.76505	0.09663	Doubt
Q3	-1.91156	0.07401	Doubt

Table 3: The results of performing the independent t-test for the common editor questions.

Aside from NASA TLX, there were also some questions that were more specifically aimed at the usability of the two editors (see Table 1). The results for these can be found in Figure 6, along with statistical significance of the differences in Table 3. It appears that none of these questions had significant differences in their answers among the two groups. Interestingly, even though the HSWFC-driven editor did not specifically focus on making it easier to understand why certain tiles can be placed, Q3 still shows a difference that is bordering on significant. In addition, Q2 scored exceptionally high for the HSWFC editor, with a low variance, which means that participants in group B almost unanimously had the feeling that they were able to tweak their creation easily.

As for the HSWFC-specific questions, it seems that `HSWFC Q2` had the most divided opinions, pointing out that not everyone finds using meta-tiles for erasing intuitive. Participants were least divided over `HSWFC Q3` and `HSWFC Q5`. `HSWFC Q1` had the highest average score, meaning that participants found the hierarchy to be quite intuitive to use. Participants were least enthusiastic about the probability tweaking, potentially because of the somewhat inconvenient User Interface (UI) for it.

Analyzing the images of the requested task that users created in Figure 7, a distinct difference appears to be that users from group *A* were much less likely to draw a river than users from group *B*, even though a river was not explicitly mentioned in the task (see Appendix B).

Another visible difference is that the users from group *A* preferred to draw relatively few houses, that were much larger and less detailed on average, while group *B*, in general, presents more houses per village. Overall, group *B* was able to get closer to the objectives that were given in the task than group *A*. Finally, going over some of the open feedback that users have given, these were the most common themes and most interesting points raised:

- A/B: some feedback for when something goes wrong would be nice, currently it seems as if just nothing happens because of auto-undoing.
- A/B: users found the snapshotting feature useful.
- A: trees were problematic to place due to them not being allowed to be adjacent (tileset constraint).
- A: making a building with an irregular shape was quite hard.
- B: the collapse path resetting tool targets all tiles, nicer would be if one could use the brush to mark specific regions instead.
- B: the root tile acting as eraser was not intuitive.
- B: probability tweaking was a bit arduous.

Interestingly, group *A* was more likely to leave open feedback than group *B*, though most of the group *A* feedback was about how hard it was to place trees.

5 DISCUSSION

The results seem to confirm that *an editor driven by HSWFC, with the additional editing facilities that come forth from it, generates significantly less cognitive load than an editor driven by stock WFC without these facilities*.

It seems that a lot of the excess cognitive load generated by the stock WFC editor that group *A* experienced comes from frustration and effort. A caveat here is that some of the excess effort and frustration may be exacerbated in the stock WFC editor due to the fact that at the time of the user study, the editor did not try to collapse as much of a brush stroke as possible yet, and instead rejected the entire stroke if it caused a contradiction. This affected both editors, but the lack of meta-tiles makes this effect more apparent in group *A* due to a higher probability for contradictions while painting, manifesting itself mostly in complaints about the tree tiles.

There were also differences in the mean values of the temporal and mental demand metrics of the NASA TLX survey between the two groups, with group A having higher mean values than group B, though the significance of these differences is doubtful, but only barely upon inspecting the significance values in Table 2. A study on the correlation between the NASA TLX dimensions shows that temporal and mental demand are strongly positively correlated to effort and frustration, so one can assume that the new editing facilities have a reducing effect on temporal and mental demand as well [20]. Noteworthy to mention is that there was no time limit on the tasks to induce temporal demand, though we did give, in the pitch, an estimated amount of time it would take (around 15 minutes), hence temporal demand may still be self-imposed from users not willing to spend too much time on the study.

An important consideration with these results from the NASA TLX survey and the extra questions is the fact that mean value of several Likert scales responses may not be fully representative, because Likert scales are considered to be ordinal, and the levels of such a scale may not be equidistant if the underlying distribution is skewed [7, 26]. We could however not find significant evidence against the normality of the score distributions in the result data with a Monte Carlo hypothesis test, as mentioned in the results. While there may be issues with interpreting mean values of Likert scale samples in isolation, we nevertheless believe that the results are insightful in showing that there may be a significant difference between the two groups by interpreting them as interval scales.

(A)

(B)

Figure 7: User creations from the requested task. Top batch corresponds to group A; the bottom one, to group B.

The fact that there was such a large difference regarding performance (see Figure 5) was also reflected in the fact that users were able to get closer to the objective in group *B* than in group *A*. It therefore seems that editor *B* made users feel more successful at accomplishing their tasks. One interesting result was the lack of rivers in group *A*. Intuitively, a river made sense for the request we had in the task. Drawing a river is not more difficult between the two editor versions, so perhaps participants were distracted by other usability issues, or were more conservative with their brush strokes in group *A* because of the effort involved with fixing things again after overwriting, a phenomenon that has been observed before in some other studies that investigate the impact of cognitive load on productivity and creativity [4, 13, 25].

The answers to the questions that were not part of the NASA TLX survey seem to show that users are positive about the new editing facilities enabled by a HSWFC-driven editor, especially regarding hierarchical tilesets. This may be because of the reducing effect they have on cognitive load, and more specifically the better sense of accomplishment they give due to being able to better implement the tasks, as confirmed by the NASA TLX performance metric.

An important observation from the open feedback in both groups was that a substantial amount of the users seem to prefer a destructive approach over a conforming approach. This can also be gathered from Q3, as the users who gave a low score here for HSWFC were the users who (according to the open feedback) chose to use the terminal tile T_{TREE} and were not able to effectively place it. In other words, it is likely that users care more about exerting their will, rather than having the constraints be fulfilled, though this would have to be confirmed through further user studies.

6 CONCLUSION

MI-PCG shows to be a promising direction with respect to building virtual worlds, but is hampered by a lack of suitable algorithms for employing the paradigm. While WFC type algorithms are promising here, they also incur additional cognitive load which makes them unappealing at first glance. In this work we have shown that HSWFC, a particular variant of the stock WFC algorithm, is able to significantly reduce this cognitive load in the setting of a 2D tile editor when compared to stock WFC in a nearly identical setup. This result brings MI-PCG driven by WFC-type algorithms closer to being suitable for mainstream use.

Since the currently established feature set of the HSWFC editor was merely a first iteration, there is much potential for even further reductions in cognitive load through more novel tooling that is enabled by HSWFC. One interesting addition to the feature set would be the ability to use complex heterogeneous brushes that paint pre-made templates of meta-tiles, e.g. a house with its walls in a certain shape and a door in a specific spot, or a template for a park that has the same spatial division but generates differently due to its meta-tiles. Such templates can generate under certain meta-tiles as well. One recent work is already exploring this facet within the context of stock WFC [3].

While promising overall, HSWFC still presents some other open challenges that require attention. For one, an additional input term is required for HSWFC over stock WFC, namely the meta-tree, which may make it harder to establish a resource/asset base. Finding ways to make input specification easier for HSWFC, either through intuitive user-driven input editors or through machine learning efforts on output and/or geographic layered data, could be the tipping point for its potential widespread adoption in virtual world design.

In addition, because hierarchical tilesets are more sophisticated in the sense that there are many possible hierarchies for the same set of concrete tiles, it can also be useful to investigate through a user study what users deem to be an optimal guideline for laying out such hierarchies: what hierarchical representation suits designers the best?

REFERENCES

[1] Shaad Alaka. 2023. *Hierarchical Semantic Wave Function Collapse*. Master's thesis. Delft University of Technology, Delft, NL. URL: http://resolver.tudelft.nl/uuid:e79264cc-1d0a-4151-9158-342a96d46764.

[2] Shaad Alaka and Rafael Bidarra. 2023. Hierarchical Semantic Wave Function Collapse. In *Proceedings of the 18th International Conference on the Foundations of Digital Games*. ACM, New York, NY, USA, Article 68, 10 pages. https://doi.org/10.1145/3582437.3587209

[3] Michael Beukman, Branden Ingram, Ireton Liu, and Benjamin Rosman. 2023. Hierarchical WaveFunction Collapse. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, Vol. 19. AAAI, Salt Lake City, Utah, USA, 23–33.

[4] Edith Galy, Magali Cariou, and Claudine Mélan. 2012. What is the relationship between mental workload factors and cognitive load types? *International journal of psychophysiology* 83, 3 (2012), 269–275.

[5] Maxim Gumin. 2016. *Wave Function Collapse Algorithm*. personal repository. https://github.com/mxgmn/WaveFunctionCollapse

[6] Sandra G Hart. 1986. NASA task load index (TLX). https://humansystems.arc.nasa.gov/groups/TLX/

[7] Susan Jamieson. 2004. Likert scales: How to (ab) use them? *Medical education* 38, 12 (2004), 1217–1218.

[8] Isaac Karth and Adam Smith. 2021. WaveFunctionCollapse: Content Generation via Constraint Solving and Machine Learning. *IEEE Transactions on Games* PP (05 2021), 1–1. https://doi.org/10.1109/TG.2021.3076368

[9] Isaac Karth and Adam M. Smith. 2017. WaveFunctionCollapse is Constraint Solving in the Wild. In *Proceedings of the 12th International Conference on the Foundations of Digital Games* (Hyannis, Massachusetts). ACM, New York, NY, USA, Article 68, 10 pages. https://doi.org/10.1145/3102071.3110566

[10] Isaac Karth and Adam M. Smith. 2019. Addressing the Fundamental Tension of PCGML with Discriminative Learning. In *Proceedings of the 14th International Conference on the Foundations of Digital Games* (San Luis Obispo, California, USA) *(FDG '19)*. ACM, New York, NY, USA, Article 89, 9 pages. https://doi.org/10.1145/3337722.3341845

[11] Hwanhee Kim, Seongtaek Lee, Hyundong Lee, Teasung Hahn, and Shinjin Kang. 2019. Automatic Generation of Game Content using a Graph-based Wave Function Collapse Algorithm. *Proceeding of IEEE Conference on Games* 1, 1 (08 2019), 1–4. https://doi.org/10.1109/CIG.2019.8848019

[12] Thijmen SL Langendam and Rafael Bidarra. 2022. miWFC - Designer empowerment through mixed-initiative Wave Function Collapse. In *Proceedings of the 17th International Conference on the Foundations of Digital Games* (Athens, Greece) *(FDG '22)*. ACM, New York, NY, USA, Article 66, 8 pages. https://doi.org/10.1145/3555858.3563266

[13] Christine S Lee and David J Therriault. 2013. The cognitive underpinnings of creative thought: A latent variable analysis exploring the roles of intelligence and working memory in three creative thinking processes. *Intelligence* 41, 5 (2013), 306–320.

[14] Alain Lioret, Nicolas Ruche, Etienne Gibiat, and Cédric Chopin. 2022. GAN Applied to Wave Function Collapse for Procedural Map Generation. In *ACM SIGGRAPH 2022 Posters* (Vancouver, BC, Canada) *(SIGGRAPH '22)*. Association for Computing Machinery, New York, NY, USA, Article 59, 2 pages. https://doi.org/10.1145/3532719.3543198

[15] Paul Merrell and Dinesh Manocha. 2010. Model synthesis: A general procedural modeling algorithm. *IEEE Transactions on Visualization and Computer Graphics* 17, 6 (2010), 715–728.

[16] Adam Newgas. 2021. Driven WaveFunctionCollapse. Article can be viewed at: https://www.boristhebrave.com/2021/06/06/driven-wavefunctioncollapse/.

[17] Adam Newgas. 2021. Tessera: A Practical System for Extended WaveFunctionCollapse. In *Proceedings of the FDG Workshop on Procedural Content Generation*. ACM, Montreal, Canada, 1–7. https://doi.org/10.1145/3472538.3472605

[18] Adam Newgas. 2022. Editable WFC. https://www.boristhebrave.com/2022/04/25/editable-wfc/

[19] Yuhe Nie, Shaoming Zheng, Zhan Zhuang, and Xuan Song. 2023. Extend Wave Function Collapse Algorithm to Large-Scale Content Generation. In *2023 IEEE Conference on Games (CoG)*. ACM, Lisbon, Portugal, 1–8. https://doi.org/10.1109/CoG57401.2023.10333214

[20] Christopher Nikulin, Gabriela Lopez, Eduardo Piñonez, Luis Gonzalez, and Pia Zapata. 2019. NASA-TLX for predictability and measurability of instructional design models: case study in design methods. *Educational Technology Research and Development* 67 (2019), 467–493.

[21] Tobias Nordvig Møller, Jonas Billeskov, and George Palamas. 2020. Expanding Wave Function Collapse with Growing Grids for Procedural Map Generation. In *Proceedings of the 15th International Conference on the Foundations of Digital Games* (Bugibba, Malta) *(FDG '20)*. ACM, New York, NY, USA, Article 28, 4 pages. https://doi.org/10.1145/3402942.3402987

[22] Ruben M. Smelik, Tim Tutenel, Klaas Jan de Kraker, and Rafael Bidarra. 2010. Interactive creation of virtual worlds using procedural sketching. In *Proceedings of Eurographics 2010 - Short papers*. Eurographics, Norrköping, Sweden, 1–4. http://graphics.tudelft.nl/Publications-new/2010/STDB10e

[23] Student. 1908. The probable error of a mean. *Biometrika* 6, 1 (1908), 1–25.

[24] Oskar Stålberg. 2021. Townscaper. Raw Fury, Steam, Epic Games Store, GOG, Nintendo Switch, xBox, App Store, Google Play. https://oskarstalberg.com/Townscaper/

[25] John Sweller. 1988. Cognitive load during problem solving: Effects on learning. *Cognitive science* 12, 2 (1988), 257–285.

[26] Huiping Wu and Shing-On Leung. 2017. Can Likert scales be treated as interval scales?—A simulation study. *Journal of social service research* 43, 4 (2017), 527–532.

[27] Ceyuan Yang, Yujun Shen, and Bolei Zhou. 2021. Semantic hierarchy emerges in deep generative representations for scene synthesis. *International Journal of Computer Vision* 129 (2021), 1451–1466.

A TILESET

An input for HSWFC consists of the triple (T, A, E), where T is the set of tiles, A is the set of adjacency constraints over T, and E is the set of edges in the meta-tree over T as triples $(T_A, T_B, w_{T_A-T_B})$, where $w_{T_A-T_B}$ represents the weight in collapsing from T_A to T_B. In figure 8, T and E are shown in the form of a meta-tree. In Section A.1, the adjacency constraints specified in the input are shown.

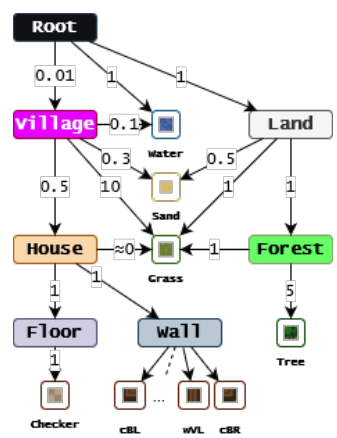


Figure 8: The meta-tree input used for the user study. Meta-tiles are shown as rounded rectangles. The rounded squares represent terminal tiles. The numbers on the edges correspond to the weights of the semantic representations. The wall terminals have been abbreviated and are only partially shown, all of their edges carry weight 1.

A.1 Adjacency constraints and (terminal) tile list

Figure 9 shows the adjacency constraints in the form of an adjacency matrix.

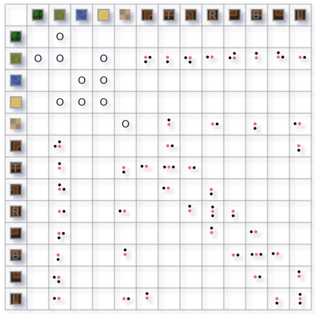


Figure 9: The 'O' represents omni-directional constraints. In the asymmetric constraints, the red dot always represents the vertical axis tile, while the other dots represent the possible relative positioning of the horizontal axis tile.

B USER STUDY

This is a description of the user study that was performed. Two groups were made, A and B; group A worked with the stock WFC editor, and group B worked with the HSWFC editor. For group B, some additional questions were asked pertaining to HSWFC functionality.

B.1 Tasks

Users were given two tasks; the first one involved following some loose specification of requirements and wishes for a game environment/world in a game studio setting, creating a design for it. The second task was more about playing around with the result of that, creating variations of the design. The main intention of the second task was to give users ample opportunity to use all the tools that were given to them, so that their opinion would be somewhat more solidified for the editor/tooling-specific questions that followed. Below you can find both task descriptions:

T1 - Creating a game world: you are the environment designer of Dragon Bane, a top-down role-playing game where the player character may traverse the world to find and complete quests in search of eternal fame and glory. You have been asked by your producer to create environment concepts for one of the outdoor areas of the game, with the deadline being today.

In this game, the player cannot pass through water, can get quests and gear in towns, and the quests often take place in the forests and the wilds. Areas in the game, such as the one you will design, are connected to each other at their borders. The gameplay designer and the lore wizard did have some more specific requirements for this particular area though:

- there should be two villages that are clearly separated
- one of the villages should be surrounded by forest as far as possible
- one of the villages should have a considerably larger house in the center of it with an interesting shape
- there should be an additional but very dense forest somewhere in the area
- the villages should be connected via a sandy path
- both villages should be adjacent to some body of water

Besides these minimum requirements (which can be interpreted fairly loosely), they trust that you will be able to fill up the rest of the environment with interesting features, as long as the points above are not violated. You can use the description given earlier to do this in a way that makes sense for the game.

Make sure to make a snapshot of your final result, and keep the window open, as there will be a second task.

T2 - Making variations: with conceptual design, it is quite usual to make several variations and iterations of some basic idea, as this allows you to home in on a final direction that both you and your clients are happy with. In this case, you have been asked to make at least one or more variants of the landscape that you have created before.

In fact, you have been given the freedom to violate the rules that were given before for these variants, with the idea of using your previous work as a starting point and experimenting from that point onward with the functionality that the editor provides.

B.2 Questions

All non-open questions used a 7-point scale. The survey for the non-HSWFC version of the editor omits the HSWFC-specific questions.

B.2.1 Questions about experience and relation to PCG. Before the main part of the survey, some questions were asked in order to determine the participants' amount of experience and their relation to PCG:

- What is your relation to procedural content generation?
- How many years of experience do you have at least in this field?

B.2.2 NASA TLX. These questions were shown after **T1**.

Mental Demand The task was mentally demanding
Temporal Demand The pace of the task felt hurried or rushed
Performance I felt successful in accomplishing the task
Effort I felt that I had to work hard to achieve my desired level of performance in the task
Frustration I felt insecure, discouraged, stressed and/or annoyed with the editor while performing the task

B.2.3 General WFC questions. See Table 1, specifically the questions prefixed with 'WFC'. These were shown after **T2**.

B.2.4 HSWFC specific questions. See Table 1, specifically the questions prefixed with 'HSWFC'. These were shown after **T2**, given that the HSWFC-editor was used.

B.2.5 Open questions.

- (After making some variations) Can you briefly describe your train of thought for coming up with the variant(s)?
- (At the end) Any particular feedback about the hierarchical tile set, the single situation-dependent brush, the regeneration tool, or elaboration on any of the scores given above?