Human Computation (2022) 9:1:58-65

ISSN: 2330-8001, DOI: 10.15346/hc.v9i1.135

# A Case Study of Quality-Diversity Search in Human Computation

SETH COOPER, NORTHEASTERN UNIVERSITY


## ABSTRACT

Human computation, applying human problem solving to computational problems, has shown promise in numerous applications. In some applications of human computation, it may be useful to find not just a single best solution, but a variety of good solutions with different properties that can be used for further analysis. Recent work in quality-diversity search, such as MAP-Elites, has developed techniques that aim to find a variety of solutions. Thus, in this work, we explore the potential of combining quality-diversity and human computation approaches. We ran a crowdsourced study of the Traveling Salesperson Problem in which some participants were provided with a visualization of their MAP-Elites archive and some were not. We did not find a difference in the quality of the best solutions found by participants between the two groups. However, we did find that participants provided with the archive visualization searched more of the MAP-Elites behavior space than those without the visualization. This demonstrates potential for quality-diversity approaches to lead to finding a wider variety of solutions in human computation search.


## 1. INTRODUCTION

Human computation (Quinn and Bederson, 2011), the application of human problem solving to computational problems, has shown promise in numerous applications, including those which can be considered searching a large space for solutions—e.g. biomolecule design (Koepnick et al., 2019; Lee et al., 2014), route planning (Anderson et al., 2000; Williams et al., 2016), and software verification (Bounov et al., 2018; Walter et al., 2019). However, in some applications, it may be useful to find not just a single "best" solution, but a variety of good solutions with different properties that can be used for further analysis—e.g. multiple promising biomolecules or routes.

Recent work in automated search has developed quality-diversity (QD) algorithms (Pugh et al., 2016), such as MAP-Elites (Mouret and Clune, 2015). These search techniques aim to find not just a single best solution, but a variety of good solutions. In basic MAP-Elites search, solutions from a high-dimensional search space (e.g. robot arm designs) are projected into a low-dimensional space by computing *behaviors* of the solutions (e.g. given a robot arm, behaviors such as its length and weight can be computed). Solutions can also be compared using a fitness function (e.g. arm cost) (Mouret and Clune, 2015). The search will aim to find good solutions according to the fitness

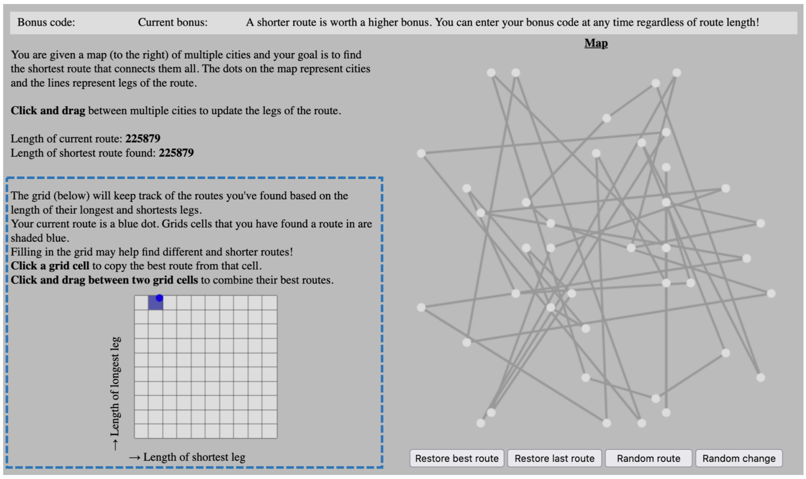



***Figure 1.*** ***Screenshot of the TSP task. In the Hidden condition, the lower-left portion showing the MAP-Elites archive and related information (outlined by a blue dotted line, which itself was not shown in the task) was hidden. The 10 × 10 grid of cells in the lower left shows the current state of the archive in the search. The location of a solution in the archive is determined by the two behaviors of the solution: the length of the its shortest leg (x-axis) and its longest leg (y-axis). The current solution is a blue dot; cells containing an elite are shaded blue. In this screenshot, there is an elite in only the cell containing the current solution.***

function (e.g. low-cost arms) that vary across this lower-dimensional behavior space (e.g. arms with a variety of lengths and weights). To do this, the low-dimensional behavior space is discretized into an *archive* of *cells*. As the search progresses, the best (i.e. most fit) solution found in each cell is stored as its *elite*. These elites can be used as the basis of a genetic search, using operators such as mutation and crossover. This results in an archive filled with good solutions that vary across the behavior dimensions. Thus, a *behavior* is a dimension of the low-dimensional representation of a solution; the *archive* is a discretized representation of the behavior space; a *cell* is one discretized region in the archive; and an *elite* is the best solution found thus far in a cell.

In this work, we explore the potential of combining QD and human computation approaches. We hypothesized that when given access to a MAP-Elites archive of their solutions, participants would (i) search more of the behavior space and (ii) find better solutions. We ran a crowdsourced study, via Amazon Mechanical Turk, of the Traveling Salesperson Problem (TSP), in which some participants were provided with an interactive visualization of their MAP-Elites archive and some were not. We did not find a difference in the quality of the best solution found between the two groups, but did find that participants provided with the archive visualization searched more of the MAP-Elites behavior space than those without the visualization. This points to QD techniques as potentially impacting the variety of solutions found in human computation search, and as an area for further study.

## 2. RELATED WORK

Previous work in human computation and crowdsourcing has explored how people and crowds can solve problems. Some work has looked specifically at, for example, how people search collaboratively (Mason et al., 2008; Bernstein et al., 2018) or in response to financial incentives (Mason and Watts, 2009) in crowdsourced contexts. As the work presented here looks at how people solve an artificially constructed task, it relates to a larger body of work that considers how people make decisions (Simon, 1956; Cohen et al., 1972), search for solutions (Billinger et al., 2014; Vuculescu et al., 2020), and trade off exploration and exploitation (March, 1991; Billinger et al., 2021). TSP has long been used as an example task that humans are able to solve efficiently (Macgregor and Ormerod, 1996). Due to this, TSP and related route-planning problems have been used as test tasks when studying human computation, problem solving and collaboration (Bernstein et al., 2018; Anderson et al., 2000; Williams et al., 2016). Additionally, several *citizen science* projects have applied the problem solving of crowds towards real scientific problems (Kawrykow et al., 2012; Koepnick et al., 2019; Lee et al., 2014; Bounov et al., 2018; Walter et al., 2019; Heck et al., 2018; Jensen et al., 2021).

QD algorithms (Pugh et al., 2016) include MAP-Elites (Mouret and Clune, 2015), SHINE (Smith et al., 2016), and novelty search with local competition (Lehman and Stanley, 2011). These are a relatively recent development in evolutionary computation, which seek to search a solution space for a variety of solutions that have good performance, rather than optimize for a single best solution. MAP-Elites and variants have found applications in domains including robotics (Nordmoen et al., 2018), workforce scheduling and routing problems (Urquhart and Hart, 2018), maze solving (Colas et al., 2020) and video game level generation (Withington, 2020; Sarkar and Cooper, 2021). In this work we are directly inspired by MAP-Elites, which, in its standard form, uses an automated evolutionary search including genetic operators such as mutation and crossover. While we retain the the archive used by MAP-Elites, we replace the automated search with a human-guided search.

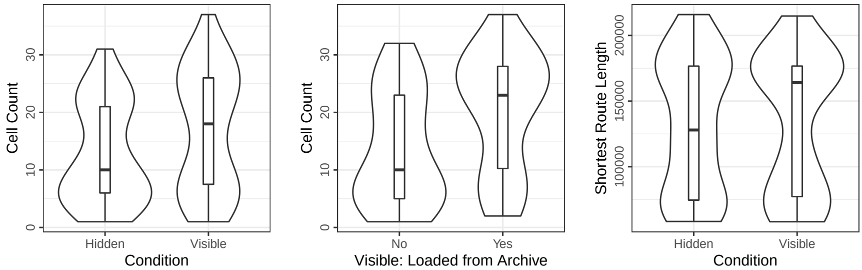
Other work has allowed humans to interact with or guide evolutionary algorithms. In human-based genetic algorithms (Kosorukoff, 2001), humans replace the genetic operators themselves; in interactive evolutionary computation (Takagi, 2001), human evaluation provides the fitness function. Recently, video game level generation approaches using MAP-Elites have introduced interactivity, allowing level designers to co-create with the evolutionary search (Alvarez et al., 2020). In this work we explore how the MAP-Elites archive impacts problem solving.

## 3. TASK DESCRIPTION

Here we describe the task used in this study. For this work we used a task based on the TSP, which has been used as an example task in studies of human computation as described in the Related Work. In the TSP, a map of cities is given, and the goal is to find the shortest route that visits all the cities and returns to the starting city. All participants were given the same (randomly-generated) layout of 40 cities, but started with a random initial route. A screenshot of the task page is given in Figure 1.

Participants are provided with short instructions in the top-left of the page, told that their goal is to find the shortest route, and provided with the length of their current route and the shortest route they found so far. They are provided with a display of the map on the right, and can click and drag on the map to draw parts of a new route. There are buttons for restoring the previous or best route found, as well as making a random change to the route or getting a completely random route.

As new routes are found by a participant, the routes are stored in their archive. Each participant has their own independent archive populated just by their search. We used a two-dimensional archive, where the low-dimensional behaviors are the length of the shortest and longest legs of the route (a leg




(a) (b) (c)

***Figure 2.*** ***Violin and quartile plots comparing distributions of: (a) number of cells in which a solution was found for each condition; (b) number of cells in which a solution was found in the Visible condition for participants who did not and did use the archive; and (c) shortest route length found for each condition.***

being a direct connection between two cities). These two lengths are rescaled based on the possible legs available in the map, and discretized into a $10 \times 10$ archive, resulting in 100 cells. Fitness is the length of the whole route. We do not argue that these are necessarily the best behavioral dimensions to use, but believe they suffice for this work; determining how different behaviors impact the search is a potential area for future work.

Information related to the MAP-Elites archive is shown in the bottom-left of the page. There is a short description of the grid display, and an encouragement to use the grid to find different and shorter routes. Participants can click on the grid to load routes from the archive, either by restoring an elite or performing crossover of two elites. In this work we use a simple version of ordered crossover (Davis, 1985). There are instructions for using the grid. The page is set up so that the section containing the archive grid and related instructions can be hidden.

A bar along the top provides information related to Mechanical Turk (described below).

## 4. STUDY

We recruited participants via a Human Intelligence Task (HIT) on Mechanical Turk. The base payment for the HIT was \$1.50. The HIT informed participants they would receive a higher bonus for finding shorter routes; the current bonus and a bonus code (which could be entered back on the Mechanical Turk site) were displayed at the top of the task. Notably, there was no additional direct monetary incentive for searching the behavior space.

After accepting the HIT and agreeing to a consent page, participants were randomly assigned into either the *Visible* condition, in which they could see the archive and related instructions and interact with the archive as described above, or the *Hidden* condition, in which their MAP-Elites archive and related instructions were hidden. Note that for participants in the Hidden condition the archive was still used to track their search for analysis, even though it was not visible.

The HIT recruited 127 participants; after filtering out 1 participant who had an invalid log and 10 who did not log any routes after the initial route, we included 116 participants in the analysis (61 in Hidden, 55 in Visible).

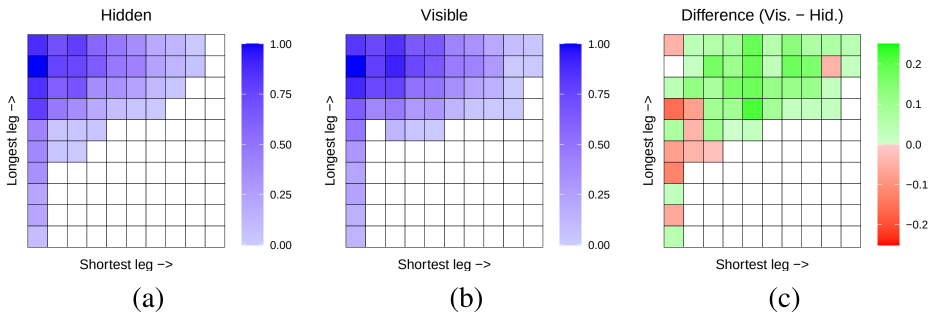



***Figure 3.*** ***(a-b) Density of routes found in each cell by each participant for the (a) Hidden and (b) Visible conditions. Values are scaled within each condition so that the cell found by the most participants is 1. Darker blue indicates relatively more routes found in that cell; white cells had no solutions found. (c) Difference between densities (Visible − Hidden). More green or more red indicate relatively more solutions found in that Visible or Hidden conditions, respectively; white cells had the same density.***

To test our hypotheses, we ran a Wilcoxon rank-sum test comparing participants in the two conditions for both (i) the count of archive cells in which a route was found and (ii) the length of the shortest route found. Cell count was found to be significant ($W = 1300.5, p = .037$), with an effect size of 0.194, considered a small effect size. Shortest route length was not significant ($W = 1666, p = .952$). For the Hidden condition, the median and maximum cell counts were 10 and 31; for the Visible condition they were 18 and 37. Figure 2(a) shows their distributions of cell counts. For the Hidden condition, the median and minimum shortest route lengths were 127,986 and 58,304; for the Visible condition they were 164,039 and 58,119. Figure 2(c) shows their distributions of shortest route lengths.

Although the goal of this work was not to determine if participants would be able to find the absolute best solution, we ran an automated method for comparison. The minimum route length found using the `python_tsp` package (https://pypi.org/project/python-tsp/) by 100 runs of the `solve_tsp_simulated_annealing` function was 56,715. Both conditions found routes within 3% of this.

To analyze use of the archive further, we looked at participants in the Visible condition who loaded routes from the archive. Only about half, or 47%, of participants in the Visible condition logged loading a solution from the archive (either through restoring an elite or performing crossover). Of all logged solutions from the Visible condition, 3.6% were restored elites and 1.1% were crossovers. Looking at the cell counts in the Visible condition, those who did not load from the archive had median and maximum cell counts of 10 and 32; those who did load had median and maximum cell counts of 23 and 37. Figure 2(b) shows their distributions of cell counts. Thus it appears that when participants were provided with the archive, interaction with it was somewhat limited, and those who did not interact with it had similar cell counts to those who did not have the archive. However, those who did use the archive seem to have been able to use it to search more of the behavior space.

We also looked at the distribution of logged routes found in the archive for each condition. Figure 3 shows the distribution of routes found by each player for both conditions, as well as the difference of the two distributions (Visible − Hidden). Participants in the Hidden condition found routes in a total of 44 cells; those in the Visible condition found 47. Visually, it appears that participants in the Hidden condition focused more on routes with shorter shortest legs, while those in the Visible condition searched relatively more for solutions in the top-right, with both longer shortest and longest legs.

Although not significant, the best routes found by participants in the Visible condition were, on average, longer than those in the Hidden condition. Potentially participants in the Visible condition spent more effort on finding a variety of routes, rather than focusing on finding the shortest.

We found participants spent a median of 6.0 minutes on the task (after removing blocks of idle time 5 minutes or more). This results in a median pay rate of $15.00 per hour before bonuses. The median and maximum bonuses awarded were $0.29 and $0.59. Although we did not compare times statistically, we found that participants in the Hidden and Visible conditions spent, respectively, a median of 7.1 minutes and 5.5 minutes. Looking further, we found that of participants in the Visible condition, those who did not and did load from the archive spent, respectively, a median of 3.0 and 7.4 minutes. It is not necessarily clear why those who did not load in the Visible condition might have spent less time overall, but it is possible the more complex interface discouraged them from spending more time on the task.

## 5. CONCLUSION

In this work, we examined giving participants in a human computation task access to a MAP-Elites archive of their solutions. Our first hypothesis, that participants provided with the archive would search more of the behavior space, was supported. There was a significant difference in the count of cells in which a route was found, with more found when the archive was visible, on average. Our second hypothesis, that those with the archive would find shorter routes, was not supported. We also found, visually, that those with the archive who load routes from it search the space differently and find more cells than those who don't. Even without a specific incentive to do so, the archive encouraged some participants to search the space of routes more broadly.

There are several limitations to this study that lead to directions for future work. We only examined one instance of one problem (TSP); in the future, we would like to generalize across other problems. It may be that the simplicity of the task limited potential differences between the two groups. We also only looked at one set of behaviors (shortest and longest leg length); this could have impacted the search and it is possible the outcome would have been different if different behaviors were chosen. The results are likely also impacted by the choice of presentation and explanation of the archive, and future work may explore different choices for this aspect of the interface.

Additionally, all search was done manually by participants, whereas MAP-Elites is usually carried out by automated search; we would like to explore combining human and automated search. We only explored one payment incentive; we would like to evaluate the effect of incentivizing more exploration of the archive directly. We also only gathered and analyzed performance data, and in the future could ask participants more qualitative questions about their experience. Finally, in this work each participant had their own archive independent of others; it would be interesting to explore multi-participant collaboration via the archive with a combined archive based on all solutions found collectively.

## 6. ACKNOWLEDGEMENTS

The author would like to thank Andreas Petrides for implementation assistance and Colan Biemer for feedback on paper drafts.

## 7. REFERENCES

Alvarez, A, Fernandez, J. M. M. F, Dahlskog, S, and Togelius, J. (2020). Interactive constrained MAP-Elites: analysis and evaluation of the expressiveness of the feature dimensions. *IEEE Transactions on Games* (2020).

Anderson, D, Anderson, E, Lesh, N, Marks, J, Mirtich, B, Ratajczak, D, and Ryall, K. (2000). Human-guided simple search. In *Proceedings of the 17th National Conference on Artificial Intelligence and 12th Conference on Innovative Applications of Artificial Intelligence*. 209–216.

Bernstein, E, Shore, J, and Lazer, D. (2018). How intermittent breaks in interaction improve collective intelligence. *Proceedings of the National Academy of Sciences* 115, 35 (Aug. 2018), 8734–8739.

Billinger, S, Srikanth, K, Stieglitz, N, and Schumacher, T. R. (2021). Exploration and exploitation in complex search tasks: how feedback influences whether and where human agents search. *Strategic Management Journal* 42, 2 (2021), 361–385.

Billinger, S, Stieglitz, N, and Schumacher, T. R. (2014). Search on rugged landscapes: an experimental study. *Organization Science* 25, 1 (Feb. 2014), 93–108.

Bounov, D, DeRossi, A, Menarini, M, Griswold, W. G, and Lerner, S. (2018). Inferring loop invariants through gamification. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 231:1–231:13.

Cohen, M. D, March, J. G, and Olsen, J. P. (1972). A garbage can model of organizational choice. *Administrative Science Quarterly* 17, 1 (1972), 1–25.

Colas, C, Madhavan, V, Huizinga, J, and Clune, J. (2020). Scaling MAP-Elites to deep neuroevolution. In *Proceedings of the 2020 Genetic and Evolutionary Computation Conference*. 67–75.

Davis, L. (1985). Applying Adaptive Algorithms to Epistatic Domains. In *Proceedings of the 9th International Joint Conference on Artificial Intelligence*. 162–164.

Heck, R, Vuculescu, O, Sørensen, J. J, Zoller, J, Andreasen, M. G, Bason, M. G, Ejlertsen, P, Elíasson, O, Haikka, P, Laustsen, J. S, Nielsen, L. L, Mao, A, Müller, R, Napolitano, M, Pedersen, M. K, Thorsen, A. R, Bergenholtz, C, Calarco, T, Montangero, S, and Sherson, J. F. (2018). Remote optimization of an ultracold atoms experiment by experts and citizen scientists. *Proceedings of the National Academy of Sciences* 115, 48 (Nov. 2018), E11231–E11237.

Jensen, J. H. M, Gajdacz, M, Ahmed, S. Z, Czarkowski, J. H, Weidner, C, Rafner, J, Sørensen, J. J, Mølmer, K, and Sherson, J. F. (2021). Crowdsourcing human common sense for quantum control. *Physical Review Research* 3, 1 (Jan. 2021), 013057.

Kawrykow, A, Roumanis, G, Kam, A, Kwak, D, Leung, C, Wu, C, Zarour, E, Sarmenta, L, Blanchette, M, Waldispühl, J, and Phylo Players, . (2012). Phylo: a citizen science approach for improving multiple sequence alignment. *PLOS ONE* 7, 3 (March 2012), e31362.

Koepnick, B, Flatten, J, Husain, T, Ford, A, Silva, D.-A, Bick, M. J, Bauer, A, Liu, G, Ishida, Y, Boykov, A, Estep, R. D, Kleinfelter, S, Nørgård-Solano, T, Wei, L, Players, F, Montelione, G. T, DiMaio, F, Popović, Z, Khatib, F, Cooper, S, and Baker, D. (2019). De novo protein design by citizen scientists. *Nature* 570, 7761 (June 2019), 390–394.

Kosorukoff, A. (2001). Human based genetic algorithm. In *2001 IEEE International Conference on Systems, Man and Cybernetics. e-Systems and e-Man for Cybernetics in Cyberspace (Cat.No.01CH37236)*, Vol. 5. 3464–3469 vol.5.

Lee, J, Kladwang, W, Lee, M, Cantu, D, Azizyan, M, Kim, H, Limpaecher, A, Yoon, S, Treuille, A, Das, R, and EteRNA Participants, . (2014). RNA design rules from a massive open laboratory. *Proceedings of the National Academy of Sciences* 111, 6 (Feb. 2014), 2122–2127.

Lehman, J and Stanley, K. O. (2011). Evolving a diversity of virtual creatures through novelty search and local competition. In *Proceedings of the 13th annual conference on Genetic and evolutionary computation*. Association for Computing Machinery, 211–218.

Macgregor, J. N and Ormerod, T. (1996). Human performance on the traveling salesman problem. *Perception & Psychophysics* 58, 4 (June 1996), 527–539.

March, J. G. (1991). Exploration and exploitation in organizational learning. *Organization Science* 2, 1 (Feb. 1991), 71–87.

Mason, W and Watts, D. J. (2009). Financial incentives and the "performance of crowds". In *Proceedings of the ACM SIGKDD Workshop on Human Computation*. 77–85.

Mason, W. A, Jones, A, and Goldstone, R. L. (2008). Propagation of innovations in networked groups. *Journal of Experimental Psychology. General* 137, 3 (Aug. 2008), 422–433.

Mouret, J.-B and Clune, J. (2015). Illuminating search spaces by mapping elites. *arXiv:1504.04909 [cs, q-bio]* (April 2015).

Nordmoen, J, Ellefsen, K. O, and Glette, K. (2018). Combining MAP-Elites and incremental evolution to generate gaits for a mammalian quadruped robot. In *Applications of Evolutionary Computation (Lecture Notes in Computer Science)*, Kevin Sim and Paul Kaufmann (Eds.). 719–733.

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 (2016), 40.

Quinn, A. J and Bederson, B. B. (2011). Human computation: a survey and taxonomy of a growing field. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1403–1412.

Sarkar, A and Cooper, S. (2021). Generating and blending game levels via quality-diversity in the latent space of a variational autoencoder. In *The 16th International Conference on the Foundations of Digital Games (FDG) 2021*. Association for Computing Machinery, 1–11.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review* 63, 2 (March 1956), 129–138.

Smith, D, Tokarchuk, L, and Wiggins, G. (2016). Rapid phenotypic landscape exploration through hierarchical spatial partitioning. In *Parallel Problem Solving from Nature – PPSN XIV (Lecture Notes in Computer Science)*, Julia Handl, Emma Hart, Peter R. Lewis, Manuel López-Ibáñez, Gabriela Ochoa, and Ben Paechter (Eds.). Springer International Publishing, 911–920.

Takagi, H. (2001). Interactive evolutionary computation: fusion of the capabilities of EC optimization and human evaluation. *Proc. IEEE* 89, 9 (Sept. 2001), 1275–1296.

Urquhart, N and Hart, E. (2018). Optimisation and illumination of a real-world workforce scheduling and routing application (WSRP) via map-elites. In *Parallel Problem Solving from Nature – PPSN XV (Lecture Notes in Computer Science)*, Anne Auger, Carlos M. Fonseca, Nuno Lourenço, Penousal Machado, Luís Paquete, and Darrell Whitley (Eds.). 488–499.

Vuculescu, O, Pedersen, M. K, Sherson, J. F, and Bergenholtz, C. (2020). Human search in a fitness landscape: how to assess the difficulty of a search problem. *Complexity* 2020 (July 2020), e7802169.

Walter, A. T, Boskin, B, Cooper, S, and Manolios, P. (2019). Gamification of loop-invariant discovery from code. *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing* 7, 1 (Oct. 2019), 188–196.

Williams, M, Gharbi, H, Ulusan, A, Ergun, O, Xiaofeng, Z, Zhang, S, and Harteveld, C. (2016). Toward human in the loop optimization through game-based experiments. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts*. 351–358.

Withington, O. (2020). Illuminating Super Mario Bros: quality-diversity within platformer level generation. In *Proceedings of the 2020 Genetic and Evolutionary Computation Conference Companion*. Association for Computing Machinery, 223–224.