COMP 6666 Fall 2020 Assignment 1d

Dennis Brown, dgb0028@auburn.edu

October 11, 2020

1 GREEN 1

Please see README.md in the repository for details on these deliverables.

2 GREEN 2

Please see README.md in the repository for details on the implementation of these deliverables.

The Multi-Objective Evolutionary Algorithm (MOEA) searches the solution space of the Light-Up puzzle. The puzzle genotype is encoded as a naive binary string representation of the puzzle board, where each bit in the binary bit string represents a white cell on the board; 0 = empty and 1 = bulb. The 2D puzzle board is encoded into the bit string row-by-row.

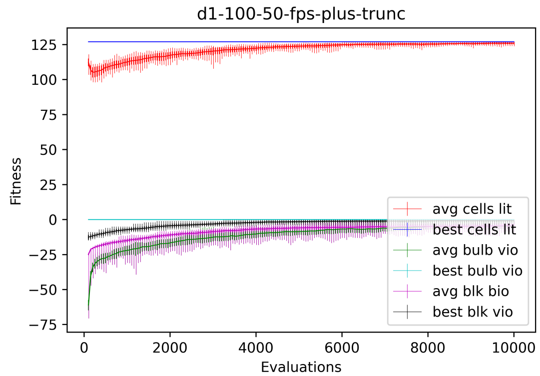
The MOEA ran against all problem instances using the following baseline parameters. Variations on these baseline parameters were used for the remaining experiments, and those variations are called out.

- Steady-state $\mu + \lambda$ EA where $\mu = 100$ and $\lambda = 50$
- 30 runs per experiment; Termination: 10000 evals per run
- Initialization: Validity forced
- Fitness: Multi-Objective
  - Maximize Cells Lit
  - Minimize Lit Bulb violations
  - Minimize Black cell violations
- Parent selection: Fitness proportional selection
- Recombination: $p_c = 1.0$; mutation chance for child $p_{mc} = 0.70$; mutation change per bit $p_m = 0.005$; fixed mutation rate
- Survival selection: Truncation

2.1 Baseline MOEA on problem d1

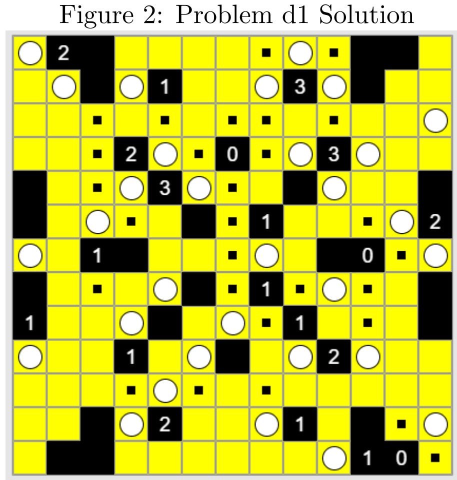
Performance of the average and best of the three objectives over 10000 runs:

Figure 1: Problem d1 Average and Best Fitness



Visual depiction of problem and solution (solution found using a Light Up puzzle website[3], NOT found by the MOEA):

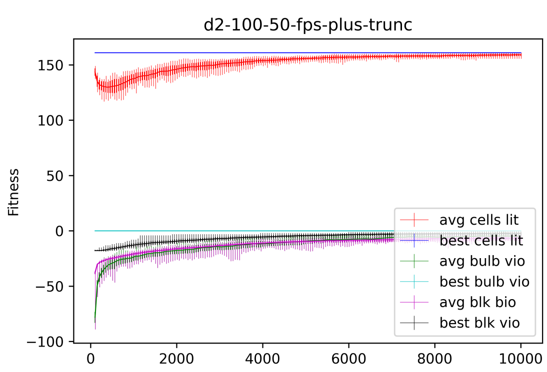
Figure 2: Problem d1 Solution



2.2 Baseline MOEA on problem d2

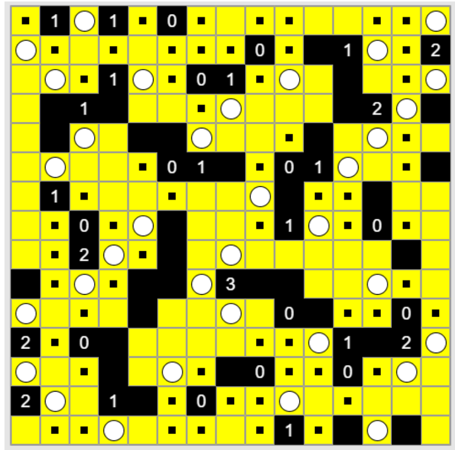
Performance of the average and best of the three objectives over 10000 runs:

Figure 3: Problem d2 Average and Best Fitness



Visual depiction of problem and solution (solution found using a Light Up puzzle website[3], NOT found by the MOEA):

Figure 4: Problem d2 Solution



3 GREEN 3

3.1 Detailed description of the MOEA

The MOEA is a variation of the Nondominated Sorting Genetic Algorithm II (NSGA-II) by Deb et.al.[1] outlined as follows in the function "execute_one_run" of the MOEAStrategy class:

- Create initial population P_0 with or without validity forced (user-configurable)
- Sort P_0 on the basis of non-domination using the NSGA-II "fast-non-dominated-sort" adapted to the data structures of this MOEA implementation
- Best level is level 0
- Fitness is set to level number; lower number, lower fitness (all fitness values are 0 or negative)
- Primary generational loop:
  - Housekeeping (write logs, check for termination)
  - User-configurable parent selection then single-point crossover combination and bitwise mutation make the next Q (first is Q_0)
  - User configurable survival strategy determines R
    - Plus: $R = P + Q$
    - Comma: $R = Q$
  - Sort R on the basis of non-domination
  - User-configurable survival selection determines next P from R

3.2 Detailed descriptions of experiments

A baseline MOEA configuration was identified as described in section GREEN 2. This baseline and several variations of it were run against problems d1 and d2. The variations changes one element at a time, such as parent selection method, plus vs. comma, survival selection, and validity-forced initialization or not. Each variation will be identified in the description of its performance as follows: The graph and table titles match the configuration files used for the experiments. The titles can be generally interpreted as

```
<puzzle>-<mu>-<lambda>-<parent_selection>-<survival_strategy>-<survival_selection>-<special>
```

e.g. the baseline configuration `d1-100-50-fps-plus-trunc` uses all the settings as described in GREEN 2, while `d1-100-50-fps-plus-rand` uses the same settings except uniform random for survivor selection.

The statistical analysis for any comparison of two alternatives consisted of performing an F-Test, determining equal or unequal variances, then performing a two-tailed t-Test. In all cases, $\alpha = 0.5$ and the null hypothesis is that the two samples are NOT statistically different.

These variations were compared to the baseline:

- Survivor selection: Truncation vs. Uniform Random
- Survivor selection: Truncation vs. Fitness Proportional
- Parent selection: Fitness Proportional vs. k-Tournament
- Survival strategy: Plus vs. Comma
- Initialization: Validity Forced vs. not

3.2.1 Distance

For the distance comparisons, solutions for d1 and d2 were found using a Light Up puzzle website[3]. The optimal solution for d1 can be stated as 127 cells lit, 0 bulb cell violations, 0 black cell violations, and 31 bulbs placed. The optimal solution for d2 can be stated as 161 cells lit, 0 bulb cell violations, 0 black cell violations, and 34 bulbs placed. For each run in each experiment, the distances from the best results to the optimal result were computed. The previously-described statistical analysis was performed on the distances from two experimental configurations to compare the two configurations.

3.2.1.1 Survivor selection: Truncation vs. Uniform Random

For both d1 and d2, Truncation survivor selection is statistically significantly better than Uniform Random.

Figure 5: d1-100-50-fps-plus-trunc vs. d1-100-50-fps-plus-rand – Distance from best values to optimal over 30 runs

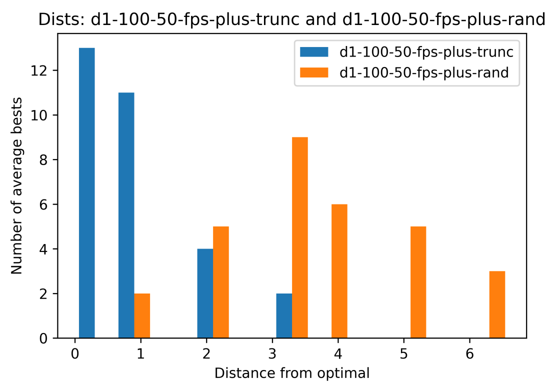


Table 1: F-Test for d1-100-50-fps-plus-trunc vs. d1-100-50-fps-plus-rand with $\alpha = 0.05$

	d1-100-50-fps-plus-trunc	d1-100-50-fps-plus-rand
Mean	0.8333333333333334	3.5333333333333333
Variance	0.8333333333333331	1.981609195402299
Observations	30	30
df	29	29
F	0.4205336426914152	
P(F≤f) one-tail	0.011363686412386787	
F Critical one-tail	0.5373999648406917	

abs(mean 1) < abs(mean 2) and F < F Critical implies unequal variances.

Table 2: t-Test for d1-100-50-fps-plus-trunc vs. d1-100-50-fps-plus-rand with Unequal Variances

	d1-100-50-fps-plus-trunc	d1-100-50-fps-plus-rand
Mean	0.8333333333333334	3.533333333333333
Variance	0.8333333333333331	1.981609195402299
Observations	30	30
df	58	29
t Stat	-8.814336478843448	
P(T≤t) two-tail	9.85038038283198e-12	
t Critical two-tail	2.0017174830120923	

abs(t Stat) > abs(t Critical two-tail) so we reject the null hypothesis – the two samples are statistically different. The average improvement of d1-100-50-fps-plus-trunc over d1-100-50-fps-plus-rand is 2.6999999999999997 closer.

Figure 6: d2-100-50-fps-plus-trunc vs. d2-100-50-fps-plus-rand – Distance from best values to optimal over 30 runs

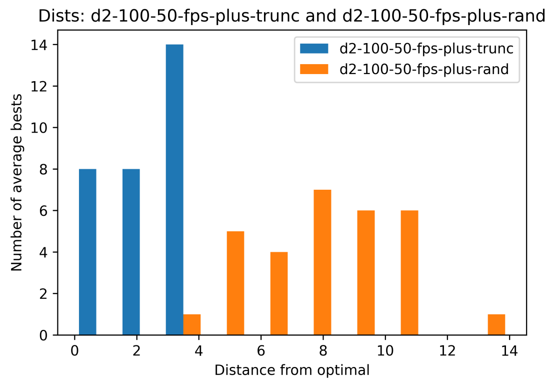


Table 3: F-Test for d2-100-50-fps-plus-trunc vs. d2-100-50-fps-plus-rand with $\alpha = 0.05$

	d2-100-50-fps-plus-trunc	d2-100-50-fps-plus-rand
Mean	2.3	7.833333333333333
Variance	1.3896551724137933	5.8678160919540225
Observations	30	30
df	29	29
F	0.23682664054848193	
P(F≤f) one-tail	0.00010897118536399184	
F Critical one-tail	0.5373999648406917	

abs(mean 1) < abs(mean 2) and F < F Critical implies unequal variances.

Table 4: t-Test for d2-100-50-fps-plus-trunc vs. d2-100-50-fps-plus-rand with Unequal Variances

	d2-100-50-fps-plus-trunc	d2-100-50-fps-plus-rand
Mean	2.3	7.833333333333333
Variance	1.3896551724137933	5.8678160919540225
Observations	30	30
df	58	29
t Stat	-11.250058951733761	
P(T≤t) two-tail	2.9648475338932954e-14	
t Critical two-tail	2.0017174830120923	

abs(t Stat) > abs(t Critical two-tail) so we reject the null hypothesis – the two samples are statistically different. The average improvement of d2-100-50-fps-plus-trunc over d2-100-50-fps-plus-rand is 5.533333333333333 closer.

3.2.1.2 Survivor selection: Truncation vs. Fitness Proportional

For d1, Truncation survivor selection is statistically significantly better than FPS; however for d2 they are not statistically different.

Figure 7: d1-100-50-fps-plus-fps vs. d1-100-50-fps-plus-fps – Distance from best values to optimal over 30 runs

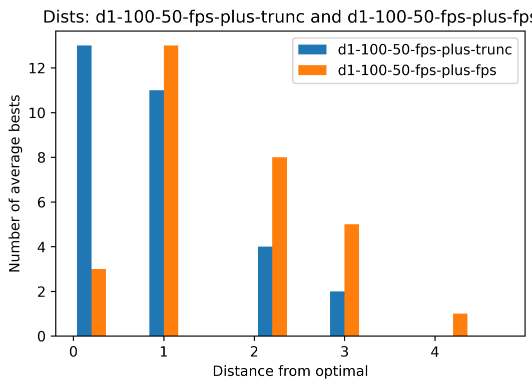


Table 5: F-Test for d1-100-50-fps-plus-trunc vs. d1-100-50-fps-plus-fps with $\alpha = 0.05$

	d1-100-50-fps-plus-trunc	d1-100-50-fps-plus-fps
Mean	0.8333333333333334	1.6
Variance	0.8333333333333331	1.006896551724138
Observations	30	30
df	29	29
F	0.8276255707762555	
P(F≤f) one-tail	0.30689278685954097	
F Critical one-tail	0.5373999648406917	

abs(mean 1) < abs(mean 2) and F > F Critical implies equal variances.

Table 6: t-Test for d1-100-50-fps-plus-trunc vs. d1-100-50-fps-plus-fps with Equal Variances

	d1-100-50-fps-plus-trunc	d1-100-50-fps-plus-fps
Mean	0.8333333333333334	1.6
Variance	0.8333333333333331	1.006896551724138
Observations	30	30
df	58	29
t Stat	-3.0955025706493835	
P(T≤t) two-tail	0.0030237281383004313	
t Critical two-tail	2.0017174830120923	

abs(t Stat) > abs(t Critical two-tail) so we reject the null hypothesis – the two samples are statistically different. The average improvement of d1-100-50-fps-plus-trunc over d1-100-50-fps-plus-fps is 0.7666666666666667 closer.

Figure 8: d2-100-50-fps-plus-trunc vs. d2-100-50-fps-plus-fps – Distance from best values to optimal over 30 runs

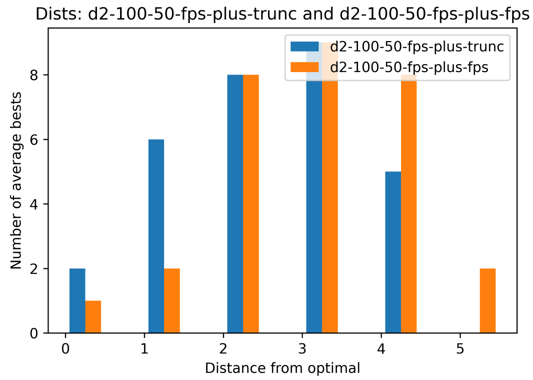


Table 7: F-Test for d2-100-50-fps-plus-trunc vs. d2-100-50-fps-plus-fps with $\alpha = 0.05$

	d2-100-50-fps-plus-trunc	d2-100-50-fps-plus-fps
Mean	2.3	2.9
Variance	1.3896551724137933	1.403448275862069
Observations	30	30
df	29	29
F	0.9901719901719903	
P(F≤f) one-tail	0.48948296782142464	
F Critical one-tail	0.5373999648406917	

abs(mean 1) < abs(mean 2) and F > F Critical implies equal variances.

Table 8: t-Test for d2-100-50-fps-plus-trunc vs. d2-100-50-fps-plus-fps with Equal Variances

	d2-100-50-fps-plus-trunc	d2-100-50-fps-plus-fps
Mean	2.3	2.9
Variance	1.3896551724137933	1.403448275862069
Observations	30	30
df	58	29
t Stat	-1.9663841605003505	
P(T≤t) two-tail	0.054045970311819035	
t Critical two-tail	2.0017174830120923	

abs(t Stat) < abs(t Critical two-tail) so we accept the null hypothesis – the two samples are NOT statistically different.

3.2.1.3 Parent selection: Fitness Proportional vs k-Tournament

For both d1 and d2, FPS is not statistically different than k-Tournament (tournament size 10).

Figure 9: d1-100-50-fps-plus-trunc vs. d1-100-50-k10-plus-trunc – Distance from best values to optimal over 30 runs

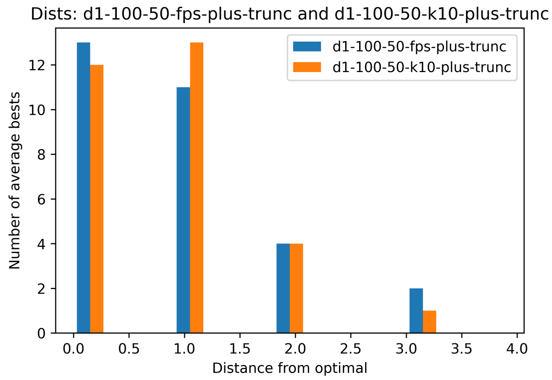


Table 9: F-Test for d1-100-50-fps-plus-trunc vs. d1-100-50-k10-plus-trunc with $\alpha = 0.05$

	d1-100-50-fps-plus-trunc	d1-100-50-k10-plus-trunc
Mean	0.8333333333333334	0.8
Variance	0.8333333333333331	0.6482758620689657
Observations	30	30
df	29	29
F	1.2854609929078007	
P(F≤f) one-tail	0.7484058770694293	
F Critical one-tail	0.5373999648406917	

abs(mean 1) > abs(mean 2) and F > F Critical implies unequal variances.

Table 10: t-Test for d1-100-50-fps-plus-trunc vs. d1-100-50-k10-plus-trunc with Unequal Variances

	d1-100-50-fps-plus-trunc	d1-100-50-k10-plus-trunc
Mean	0.8333333333333334	0.8
Variance	0.8333333333333331	0.6482758620689657
Observations	30	30
df	58	29
t Stat	0.14999353490075777	
P(T≤t) two-tail	0.8812982070504818	
t Critical two-tail	2.0017174830120923	

abs(t Stat) < abs(t Critical two-tail) so we accept the null hypothesis – the two samples are NOT statistically different.

Figure 10: d2-100-50-fps-plus-trunc vs. d2-100-50-k10-plus-trunc – Distance from best values to optimal over 30 runs

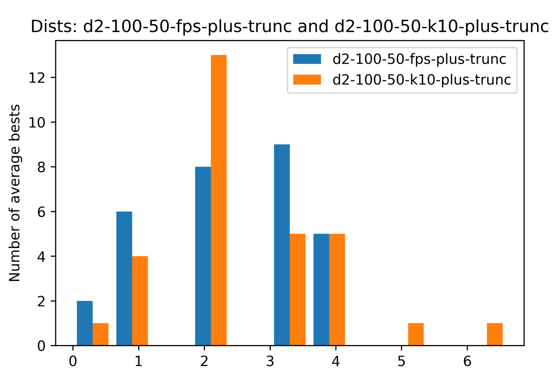


Table 11: F-Test for d2-100-50-fps-plus-trunc vs. d2-100-50-k10-plus-trunc with $\alpha = 0.05$

	d2-100-50-fps-plus-trunc	d2-100-50-k10-plus-trunc
Mean	2.3	2.533333333333333
Variance	1.3896551724137933	1.705747126436782
Observations	30	30
df	29	29
F	0.8146900269541778	
P(F≤f) one-tail	0.29231684073571573	
F Critical one-tail	0.5373999648406917	

abs(mean 1) < abs(mean 2) and F > F Critical implies equal variances.

Table 12: t-Test for d2-100-50-fps-plus-trunc vs. d2-100-50-k10-plus-trunc with Equal Variances

	d2-100-50-fps-plus-trunc	d2-100-50-k10-plus-trunc
Mean	2.3	2.533333333333333
Variance	1.3896551724137933	1.705747126436782
Observations	30	30
df	58	29
t Stat	-0.7264050625446226	
P(T≤t) two-tail	0.4705119760926939	
t Critical two-tail	2.0017174830120923	

abs(t Stat) < abs(t Critical two-tail) so we accept the null hypothesis – the two samples are NOT statistically different.

3.2.1.4 Survival strategy: Plus vs. Comma

For d1, Plus is not statistically different than Comma ($\mu = 100, \lambda = 100$). For d2, Plus is statistically significantly better (slightly).

Figure 11: d1-100-50-fps-plus-trunc vs. d1-100-100-fps-comma-trunc – Distance from best values to optimal over 30 runs

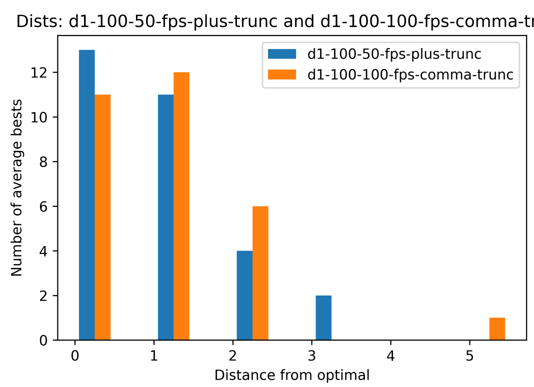


Table 13: F-Test for d1-100-50-fps-plus-trunc vs. d1-100-100-fps-comma-trunc with $\alpha = 0.05$

	d1-100-50-fps-plus-trunc	d1-100-100-fps-comma-trunc
Mean	0.8333333333333334	0.9666666666666667
Variance	0.8333333333333331	1.1367816091954022
Observations	30	30
df	29	29
F	0.7330637007077856	
P(F≤f) one-tail	0.2040361427538532	
F Critical one-tail	0.5373999648406917	

abs(mean 1) < abs(mean 2) and F > F Critical implies equal variances.

Table 14: t-Test for d1-100-50-fps-plus-trunc vs. d1-100-100-fps-comma-trunc with Equal Variances

	d1-100-50-fps-plus-trunc	d1-100-100-fps-comma-trunc
Mean	0.8333333333333334	0.9666666666666667
Variance	0.8333333333333331	1.1367816091954022
Observations	30	30
df	58	29
t Stat	-0.5202997071857234	
P(T≤t) two-tail	0.6048351803709172	
t Critical two-tail	2.0017174830120923	

abs(t Stat) < abs(t Critical two-tail) so we accept the null hypothesis – the two samples are NOT statistically different.

Figure 12: d2-100-50-fps-plus-trunc vs. d2-100-100-fps-comma-trunc – Distance from best values to optimal over 30 runs

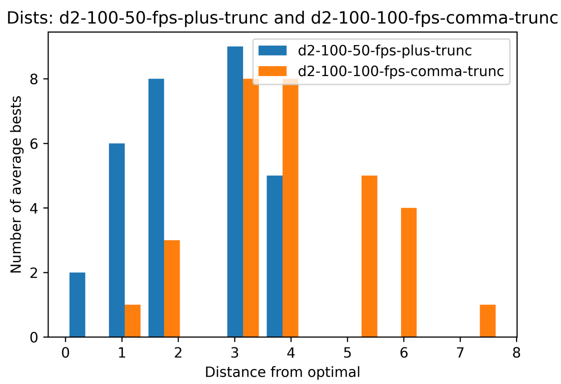


Table 15: F-Test for d2-100-50-fps-plus-trunc vs. d2-100-100-fps-comma-trunc with $\alpha = 0.05$

	d2-100-50-fps-plus-trunc	d2-100-100-fps-comma-trunc
Mean	2.3	3.966666666666667
Variance	1.3896551724137933	2.033333333333333
Observations	30	30
df	29	29
F	0.6834369700395705	
P(F≤f) one-tail	0.15551993137941117	
F Critical one-tail	0.5373999648406917	

abs(mean 1) < abs(mean 2) and F > F Critical implies equal variances.

Table 16: t-Test for d2-100-50-fps-plus-trunc vs. d2-100-100-fps-comma-trunc with Equal Variances

	d2-100-50-fps-plus-trunc	d2-100-100-fps-comma-trunc
Mean	2.3	3.966666666666667
Variance	1.3896551724137933	2.033333333333333
Observations	30	30
df	58	29
t Stat	-4.934085337683503	
P(T≤t) two-tail	7.128995483925265e-06	
t Critical two-tail	2.0017174830120923	

abs(t Stat) > abs(t Critical two-tail) so we reject the null hypothesis – the two samples are statistically different. The average improvement of d2-100-50-fps-plus-trunc over d2-100-100-fps-comma-trunc is 1.666666666666667 closer.

3.2.1.5 Initialization: Validity Forced vs. not

For both d1 and d2, initialization with Validity Forced is not statistically different than uniform random initialization.

Figure 13: d1-100-50-fps-plus-trunc vs. d1-100-50-fps-plus-trunc-noval – Distance from best values to optimal over 30 runs

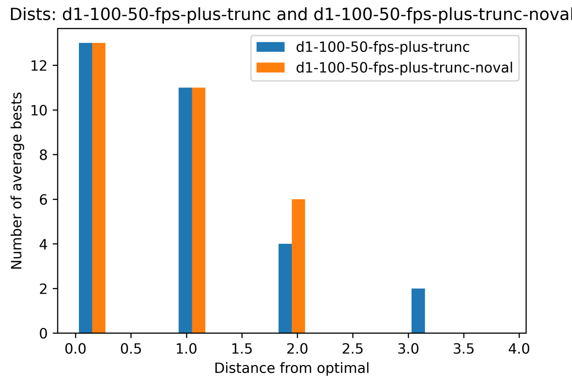


Table 17: F-Test for d1-100-50-fps-plus-trunc vs. d1-100-50-fps-plus-trunc-noval with $\alpha = 0.05$

	d1-100-50-fps-plus-trunc	d1-100-50-fps-plus-trunc-noval
Mean	0.8333333333333334	0.7666666666666667
Variance	0.8333333333333331	0.5988505747126437
Observations	30	30
df	29	29
F	1.3915547024952013	
P(F≤f) one-tail	0.8105908202273817	
F Critical one-tail	0.5373999648406917	

abs(mean 1) > abs(mean 2) and F > F Critical implies unequal variances.

Table 18: t-Test for d1-100-50-fps-plus-trunc vs. d1-100-50-fps-plus-trunc-noval with Unequal Variances

	d1-100-50-fps-plus-trunc	d1-100-50-fps-plus-trunc-noval
Mean	0.8333333333333334	0.7666666666666667
Variance	0.8333333333333331	0.5988505747126437
Observations	30	30
df	58	29
t Stat	0.3051195066242609	
P(T≤t) two-tail	0.7613966361952904	
t Critical two-tail	2.0017174830120923	

abs(t Stat) < abs(t Critical two-tail) so we accept the null hypothesis – the two samples are NOT statistically different.

Figure 14: d2-100-50-fps-plus-trunc vs. d2-100-50-fps-plus-trunc-noval – Distance from best values to optimal over 30 runs

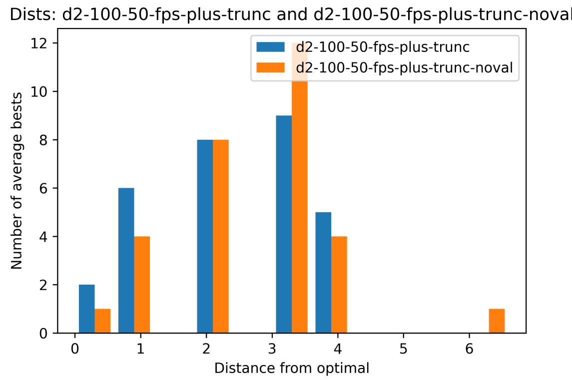


Table 19: F-Test for d2-100-50-fps-plus-trunc vs. d2-100-50-fps-plus-trunc-noval with $\alpha = 0.05$

	d2-100-50-fps-plus-trunc	d2-100-50-fps-plus-trunc-noval
Mean	2.3	2.6
Variance	1.3896551724137933	1.4206896551724133
Observations	30	30
df	29	29
F	0.9781553398058257	
P(F≤f) one-tail	0.4764921717723981	
F Critical one-tail	0.5373999648406917	

abs(mean 1) < abs(mean 2) and F > F Critical implies equal variances.

Table 20: t-Test for d2-100-50-fps-plus-trunc vs. d2-100-50-fps-plus-trunc-noval with Equal Variances

	d2-100-50-fps-plus-trunc	d2-100-50-fps-plus-trunc-noval
Mean	2.3	2.6
Variance	1.3896551724137933	1.4206896551724133
Observations	30	30
df	58	29
t Stat	-0.9801715137255418	
P(T≤t) two-tail	0.33107241410768906	
t Critical two-tail	2.0017174830120923	

abs(t Stat) < abs(t Critical two-tail) so we accept the null hypothesis – the two samples are NOT statistically different.

3.2.2 Diversity

Diversity was calculated by using the crowding distances for the final best front of an experiment. The previously-described statistical analysis was performed on the crowding distances from two experimental configurations to compare the two configurations, even though the number of data points is frequently insufficient for these findings to be conclusive. Readers can draw their own conclusions from the graphs as they see fit.

3.2.2.1 Survivor selection: Truncation vs. Uniform Random

For d1, there is no statistical difference in diversity, though from the small dataset the difference appears stark. For d2, Uniform Random survivor selection is slightly but statistically significantly more diverse than Truncation.

Figure 15: d1-100-50-fps-plus-trunc vs. d1-100-50-fps-plus-rand – Diversity (Crowding distance)

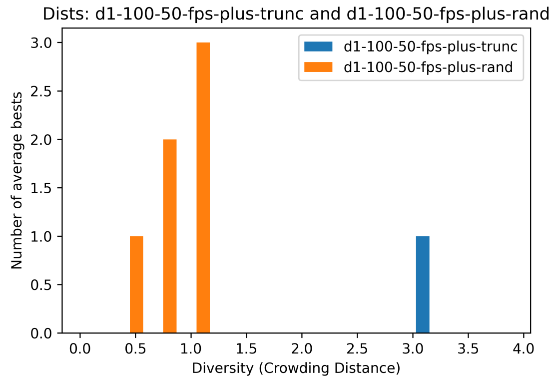


Table 21: F-Test for d1-100-50-fps-plus-trunc vs. d1-100-50-fps-plus-rand with $\alpha = 0.05$

	d1-100-50-fps-plus-trunc	d1-100-50-fps-plus-rand
Mean	3.0	0.8814102564102564
Variance	nan	0.04282955292570677
Observations	1	1
df	0	0
F	nan	
P(F≤f) one-tail	nan	
F Critical one-tail	nan	

Table 22: t-Test for d1-100-50-fps-plus-trunc vs. d1-100-50-fps-plus-rand with Unequal Variances

	d1-100-50-fps-plus-trunc	d1-100-50-fps-plus-rand
Mean	3.0	0.8814102564102564
Variance	nan	0.04282955292570677
Observations	1	1
df	0	0
t Stat	nan	
P(T≤t) two-tail	nan	
t Critical two-tail	nan	

abs(t Stat) < abs(t Critical two-tail) so we accept the null hypothesis – the two samples are NOT statistically different.

Figure 16: d2-100-50-fps-plus-trunc vs. d2-100-50-fps-plus-rand – Diversity (Crowding distance)

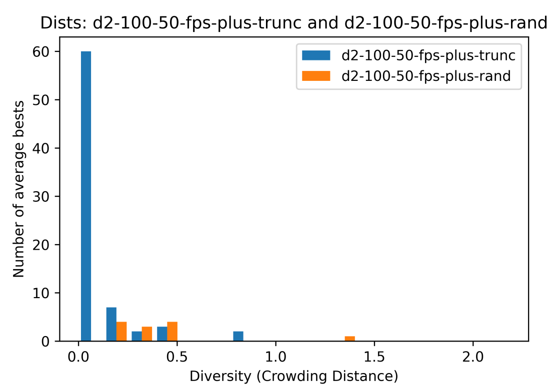


Table 23: F-Test for d2-100-50-fps-plus-trunc vs. d2-100-50-fps-plus-rand with $\alpha = 0.05$

	d2-100-50-fps-plus-trunc	d2-100-50-fps-plus-rand
Mean	0.07094594594594594	0.3944444444444444
Variance	0.029247038134024444	0.09574864289150002
Observations	74	74
df	73	73
F	0.3054564247679882	
P(F≤f) one-tail	4.501794803117597e-07	
F Critical one-tail	0.6787174812558107	

abs(mean 1) < abs(mean 2) and F < F Critical implies unequal variances.

Table 24: t-Test for d2-100-50-fps-plus-trunc vs. d2-100-50-fps-plus-rand with Unequal Variances

	d2-100-50-fps-plus-trunc	d2-100-50-fps-plus-rand
Mean	0.07094594594594594	0.3944444444444444
Variance	0.029247038134024444	0.09574864289150002
Observations	74	74
df	146	73
t Stat	-3.535073365057785	
P(T≤t) two-tail	0.004052422037672706	
t Critical two-tail	1.9763456545827003	

abs(t Stat) > abs(t Critical two-tail) so we reject the null hypothesis – the two samples are statistically different. The average improvement of d2-100-50-fps-plus-trunc over d2-100-50-fps-plus-rand is -0.32349849849849843 more diverse.

3.2.2.2 Survivor selection: Truncation vs. Fitness Proportional

For d1, there is no statistical difference in diversity. For d2, FPS survivor selection is slightly but statistically significantly more diverse than truncation.

Figure 17: d1-100-50-fps-plus-trunc vs. d1-100-50-fps-plus-fps – Diversity (Crowding distance)

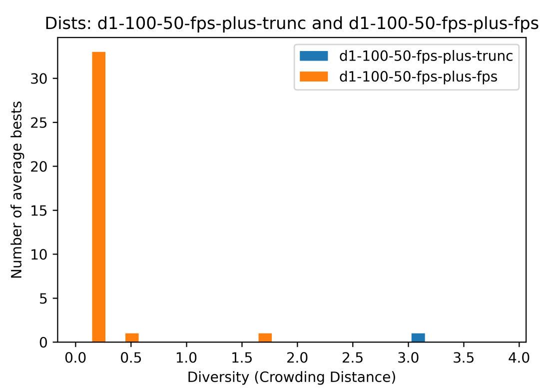


Table 25: F-Test for d1-100-50-fps-plus-trunc vs. d1-100-50-fps-plus-fps with $\alpha = 0.05$

	d1-100-50-fps-plus-trunc	d1-100-50-fps-plus-fps
Mean	3.0	0.11020408163265305
Variance	nan	0.0854870519636426
Observations	1	1
df	0	0
F	nan	
P(F≤f) one-tail	nan	
F Critical one-tail	nan	

Table 26: t-Test for d1-100-50-fps-plus-trunc vs. d1-100-50-fps-plus-fps with Unequal Variances

	d1-100-50-fps-plus-trunc	d1-100-50-fps-plus-fps
Mean	3.0	0.11020408163265305
Variance	nan	0.0854870519636426
Observations	1	1
df	0	0
t Stat	nan	
P(T≤t) two-tail	nan	
t Critical two-tail	nan	

abs(t Stat) < abs(t Critical two-tail) so we accept the null hypothesis – the two samples are NOT statistically different.

Figure 18: d2-100-50-fps-plus-trunc vs. d2-100-50-fps-plus-fps – Diversity (Crowding distance)

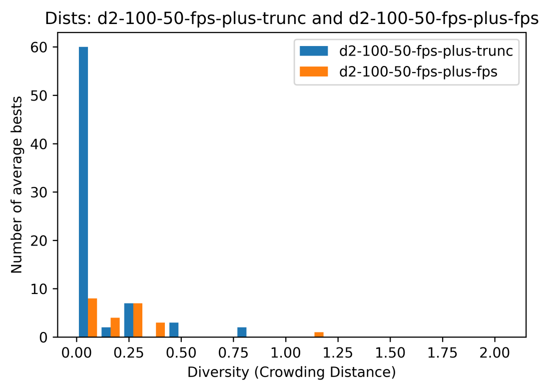


Table 27: F-Test for d2-100-50-fps-plus-trunc vs. d2-100-50-fps-plus-fps with $\alpha = 0.05$

	d2-100-50-fps-plus-trunc	d2-100-50-fps-plus-fps
Mean	0.07094594594594594	0.2065217391304348
Variance	0.029247038134024444	0.05484189723320159
Observations	74	74
df	73	73
F	0.5332973439933826	
P(F≤f) one-tail	0.003964990481930329	
F Critical one-tail	0.6787174812558107	

abs(mean 1) < abs(mean 2) and F < F Critical implies unequal variances.

Table 28: t-Test for d2-100-50-fps-plus-trunc vs. d2-100-50-fps-plus-fps with Unequal Variances

	d2-100-50-fps-plus-trunc	d2-100-50-fps-plus-fps
Mean	0.07094594594594594	0.2065217391304348
Variance	0.029247038134024444	0.05484189723320159
Observations	74	74
df	146	73
t Stat	-2.5714984772256693	
P(T≤t) two-tail	0.015390410055745223	
t Critical two-tail	1.9763456545827003	

abs(t Stat) > abs(t Critical two-tail) so we reject the null hypothesis – the two samples are statistically different. The average improvement of d2-100-50-fps-plus-trunc over d2-100-50-fps-plus-fps is -0.13557579318448887 more diverse.

3.2.2.3 Parent selection: Fitness Proportional vs k-Tournament

For both d1 and d2, there is no statistical difference in diversity between FPS and k-Tournament (k=10) parent selection.

Figure 19: d1-100-50-fps-plus-trunc vs. d1-100-50-k10-plus-trunc – Diversity (Crowding distance)

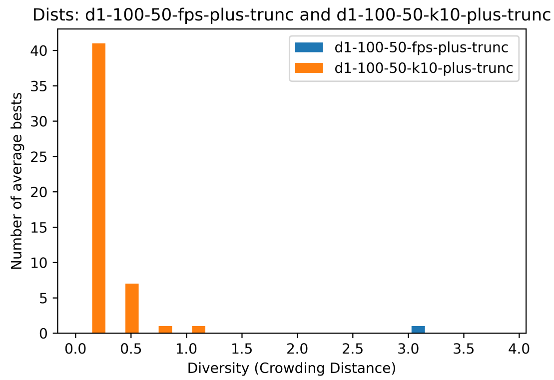


Table 29: F-Test for d1-100-50-fps-plus-trunc vs. d1-100-50-k10-plus-trunc with $\alpha = 0.05$

	d1-100-50-fps-plus-trunc	d1-100-50-k10-plus-trunc
Mean	3.0	0.1
Variance	nan	0.04295792391030486
Observations	1	1
df	0	0
F	nan	
P(F≤f) one-tail	nan	
F Critical one-tail	nan	

Table 30: t-Test for d1-100-50-fps-plus-trunc vs. d1-100-50-k10-plus-trunc with Unequal Variances

	d1-100-50-fps-plus-trunc	d1-100-50-k10-plus-trunc
Mean	3.0	0.1
Variance	nan	0.04295792391030486
Observations	1	1
df	0	0
t Stat	nan	
P(T≤t) two-tail	nan	
t Critical two-tail	nan	

abs(t Stat) < abs(t Critical two-tail) so we accept the null hypothesis – the two samples are NOT statistically different.

Figure 20: d2-100-50-fps-plus-trunc vs. d2-100-50-k10-plus-trunc – Diversity (Crowding distance)

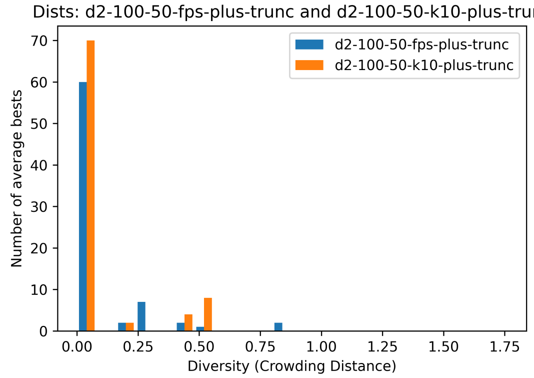


Table 31: F-Test for d2-100-50-fps-plus-trunc vs. d2-100-50-k10-plus-trunc with $\alpha = 0.05$

	d2-100-50-fps-plus-trunc	d2-100-50-k10-plus-trunc
Mean	0.07094594594594594	0.07142857142857142
Variance	0.029247038134024444	0.02760757314974184
Observations	74	74
df	73	73
F	1.0593846107149745	
P(F≤f) one-tail	0.5969980731769605	
F Critical one-tail	0.6787174812558107	

abs(mean 1) < abs(mean 2) and F > F Critical implies equal variances.

Table 32: t-Test for d2-100-50-fps-plus-trunc vs. d2-100-50-k10-plus-trunc with Equal Variances

	d2-100-50-fps-plus-trunc	d2-100-50-k10-plus-trunc
Mean	0.07094594594594594	0.07142857142857142
Variance	0.029247038134024444	0.02760757314974184
Observations	74	74
df	146	73
t Stat	-0.017970953479941 37	
P(T≤t) two-tail	0.9856849893365522	
t Critical two-tail	1.9763456545827003	

abs(t Stat) < abs(t Critical two-tail) so we accept the null hypothesis – the two samples are NOT statistically different.

3.2.2.4 Survival strategy: Plus vs. Comma

For d1, there is no statistical difference in diversity. For d2, Comma survival strategy is slightly but statistically significantly more diverse than Plus.

Figure 21: d1-100-50-fps-plus-trunc vs. d1-100-100-fps-comma-trunc – Diversity (Crowding distance)

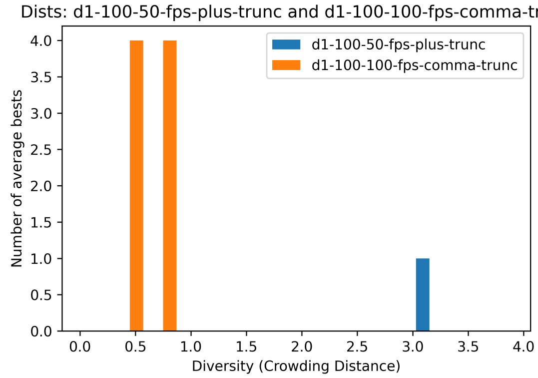


Table 33: F-Test for d1-100-50-fps-plus-trunc vs. d1-100-100-fps-comma-trunc with $\alpha = 0.05$

	d1-100-50-fps-plus-trunc	d1-100-100-fps-comma-trunc
Mean	3.0	0.6772727272727272
Variance	nan	0.04035419126328217
Observations	1	1
df	0	0
F	nan	
P(F≤f) one-tail	nan	
F Critical one-tail	nan	

Table 34: t-Test for d1-100-50-fps-plus-trunc vs. d1-100-100-fps-comma-trunc with Unequal Variances

	d1-100-50-fps-plus-trunc	d1-100-100-fps-comma-trunc
Mean	3.0	0.6772727272727272
Variance	nan	0.04035419126328217
Observations	1	1
df	0	0
t Stat	nan	
P(T≤t) two-tail	nan	
t Critical two-tail	nan	

abs(t Stat) < abs(t Critical two-tail) so we accept the null hypothesis – the two samples are NOT statistically different.

Figure 22: d2-100-50-fps-plus-trunc vs. d2-100-100-fps-comma-trunc – Diversity (Crowding distance)

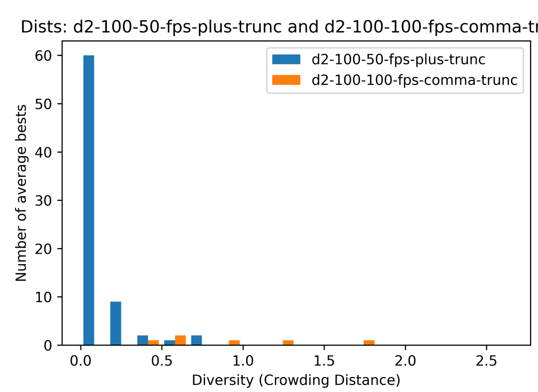


Table 35: F-Test for d2-100-50-fps-plus-trunc vs. d2-100-100-fps-comma-trunc with $\alpha = 0.05$

	d2-100-50-fps-plus-trunc	d2-100-100-fps-comma-trunc
Mean	0.07094594594594594	0.8983333333333334
Variance	0.029247038134024444	0.2036166666666667
Observations	74	74
df	73	73
F	0.1436377415111293	
P(F≤f) one-tail	5.615632142042441e-15	
F Critical one-tail	0.6787174812558107	

abs(mean 1) < abs(mean 2) and F < F Critical implies unequal variances.

Table 36: t-Test for d2-100-50-fps-plus-trunc vs. d2-100-100-fps-comma-trunc with Unequal Variances

	d2-100-50-fps-plus-trunc	d2-100-100-fps-comma-trunc
Mean	0.07094594594594594	0.8983333333333334
Variance	0.029247038134024444	0.2036166666666667
Observations	74	74
df	146	73
t Stat	-4.465432308200549	
P(T≤t) two-tail	0.006252595433800118	
t Critical two-tail	1.9763456545827003	

abs(t Stat) > abs(t Critical two-tail) so we reject the null hypothesis – the two samples are statistically different. The average improvement of d2-100-50-fps-plus-trunc over d2-100-100-fps-comma-trunc is -0.8273873873873875 more diverse.

3.2.2.5 Initialization: Validity Forced vs. not

For both d1 and d2, there is no statistical difference in diversity between initialization with and without validity forced.

Figure 23: d1-100-50-fps-plus-trunc vs. d1-100-50-fps-plus-noval – Diversity (Crowding distance)

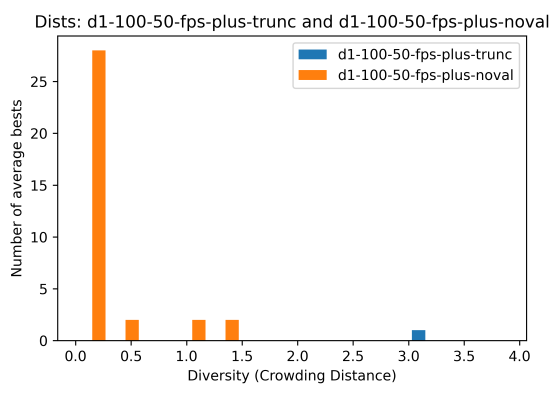


Table 37: F-Test for d1-100-50-fps-plus-trunc vs. d1-100-50-fps-plus-noval with $\alpha = 0.05$

	d1-100-50-fps-plus-trunc	d1-100-50-fps-plus-noval
Mean	3.0	0.17647058823529413
Variance	nan	0.15594474153297683
Observations	1	1
df	0	0
F	nan	
P(F≤f) one-tail	nan	
F Critical one-tail	nan	

Table 38: t-Test for d1-100-50-fps-plus-trunc vs. d1-100-50-fps-plus-noval with Unequal Variances

	d1-100-50-fps-plus-trunc	d1-100-50-fps-plus-noval
Mean	3.0	0.17647058823529413
Variance	nan	0.15594474153297683
Observations	1	1
df	0	0
t Stat	nan	
P(T≤t) two-tail	nan	
t Critical two-tail	nan	

abs(t Stat) < abs(t Critical two-tail) so we accept the null hypothesis – the two samples are NOT statistically different.

Figure 24: d2-100-50-fps-plus-trunc vs. d2-100-50-fps-plus-noval – Diversity (Crowding distance)

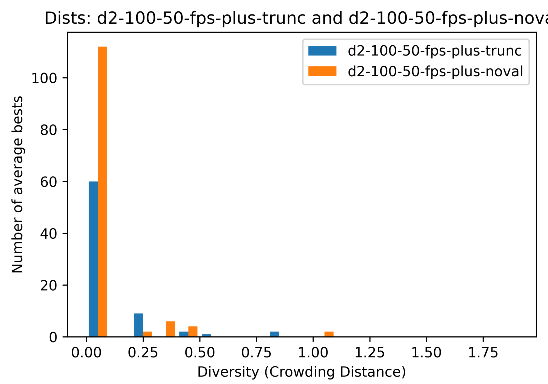


Table 39: F-Test for d2-100-50-fps-plus-trunc vs. d2-100-50-fps-plus-noval with $\alpha = 0.05$

	d2-100-50-fps-plus-trunc	d2-100-50-fps-plus-noval
Mean	0.07094594594594594	0.04761904761904761
Variance	0.029247038134024444	0.024807619047619033
Observations	74	74
df	73	73
F	1.1789538559860904	
P(F≤f) one-tail	0.7581895784359142	
F Critical one-tail	0.6787174812558107	

abs(mean 1) > abs(mean 2) and F > F Critical implies unequal variances.

Table 40: t-Test for d2-100-50-fps-plus-trunc vs. d2-100-50-fps-plus-noval with Unequal Variances

	d2-100-50-fps-plus-trunc	d2-100-50-fps-plus-noval
Mean	0.07094594594594594	0.04761904761904761
Variance	0.029247038134024444	0.024807619047619033
Observations	74	74
df	146	73
t Stat	0.9586356102309811	
P(T≤t) two-tail	0.3393591321621934	
t Critical two-tail	1.9763456545827003	

abs(t Stat) < abs(t Critical two-tail) so we accept the null hypothesis – the two samples are NOT statistically different.

3.2.3 Conclusion

The author started with a baseline configuration determined to be best through ad-hoc experimentation, and then created several variations of it to see if those variations would improve it. By the author's measurements, the baseline's Truncation survivor selection outperformed or matched the performance of Uniform Random and Fitness Proportional selection; the baseline's Fitness Proportional parent selection matched the performance of k-Tournament (k=10); the baseline's Plus survival strategy outperformed or matches Comma; and the baseline's Validity Forced initialization did not seem to impact performance compared to not using it.

Regarding differences in datasets (problems), when considering distance from optimal, d1 showed a significant difference for Truncation survivor selection over Uniform Random and FPS; while d2 showed a significant difference for Truncation survivor selection over Uniform Random and Plus survival strategy over Comma. When considering diversity (accepting the measure is flawed), d1 showed no statistically significant differences in the variations, while d2 showed statistically significant differences in that Uniform Random and FPS survival selection and the Comma survival strategy showed more diversity than their alternatives, which is the opposite of the performance of those options for distance. For d2, diversity (again, as measured here) would seem to have a negative correlation with distance from optimal, even though diversity is generally desirable.

The diversity (crowding distance) analysis is unreliable. The author struggled with what to measure, and landed on the final best front of each experiment. However, the sheer shortage of members in the data sets in some cases make the statistical analysis meaningless (however, claims of "no statistically significant difference" are still technically true for the given data!). This caveat is true for diversity measurements throughout this report.

4 YELLOW 1: Fitness Sharing

4.1 Implementation

Fitness sharing was implemented closely following a textbook description in section 5.5.3 of Eiben and Smith[2], where new fitness values were calculated following this formula, where σ is user-configurable and defaults to 10.

Figure 25: Fitness Sharing

$$F'(i) = \frac{F(i)}{\sum_j sh(d(i,j))},$$

where the sharing function $sh(d)$ is a function of the distance d, given by

$$sh(d) = \begin{cases} 1 - (d/\sigma_{share})^{\alpha} & \text{if } d \leq \sigma_{share}, \\ 0 & \text{otherwise .} \end{cases}$$

The new fitness values were calculated as indicated in this updated MOEA outline:

- Create initial population P_0 with or without validity forced (user-configurable)
- Sort P_0 on the basis of non-domination using the NSGA-II "fast-non-dominated-sort" adapted to the data structures of this MOEA implementation
- Best level is level 0
- Fitness is set to level number; lower number, lower fitness (all fitness values are 0 or negative)
- Primary generational loop:
  - Housekeeping (write logs, check for termination)
  - User-configurable parent selection then single-point crossover combination and bitwise mutation make the next Q (first is Q_0)
  - User configurable survival strategy determines R

    * Plus: $R = P + Q$
    * Comma: $R = Q$
  - Sort R on the basis of non-domination
  - **Modify fitness values with Fitness Sharing**
  - User-configurable survival selection determines next P from R

4.2 Experiments

Experiments considered distance and diversity for performance on problems d1 and d2 for the baseline configuration with and without Fitness Sharing.

4.2.1 Distance

4.2.1.1 With and Without Fitness Sharing; $\sigma = 10$

For both d1 and d2, using Fitness Sharing with $\sigma = 10$ is not statistically different than not using it.

Figure 26: d1-100-50-fps-plus-trunc vs. d1-100-50-fps-plus-trunc-share10 – Distance from best values to optimal over 30 runs

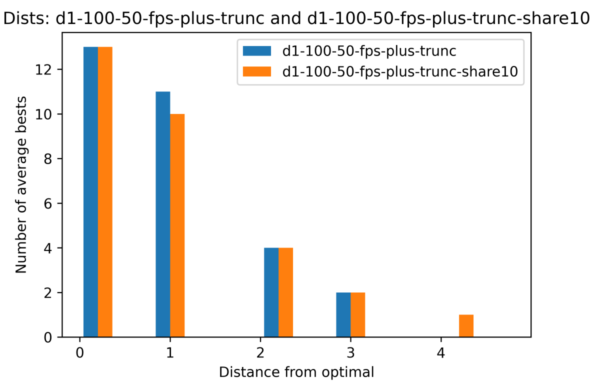


Table 41: F-Test for d1-100-50-fps-plus-trunc vs. d1-100-50-fps-plus-trunc-share10 with $\alpha = 0.05$

	d1-100-50-fps-plus-trunc	d1-100-50-fps-plus-trunc-share10
Mean	0.8333333333333334	0.9333333333333333
Variance	0.8333333333333331	1.167816091954023
Observations	30	30
df	29	29
F	0.7135826771653542	
P(F≤f) one-tail	0.18439880669247	
F Critical one-tail	0.5373999648406917	

abs(mean 1) < abs(mean 2) and F > F Critical implies equal variances.

Table 42: t-Test for d1-100-50-fps-plus-trunc vs. d1-100-50-fps-plus-trunc-share10 with Equal Variances

	d1-100-50-fps-plus-trunc	d1-100-50-fps-plus-trunc-share10
Mean	0.8333333333333334	0.9333333333333333
Variance	0.8333333333333331	1.167816091954023
Observations	30	30
df	58	29
t Stat	-0.38718708994384654	
P(T≤t) two-tail	0.7000348330877215	
t Critical two-tail	2.0017174830120923	

abs(t Stat) < abs(t Critical two-tail) so we accept the null hypothesis – the two samples are NOT statistically different.

Figure 27: d2-100-50-fps-plus-trunc vs. d2-100-50-fps-plus-trunc-share10 – Distance from best values to optimal over 30 runs

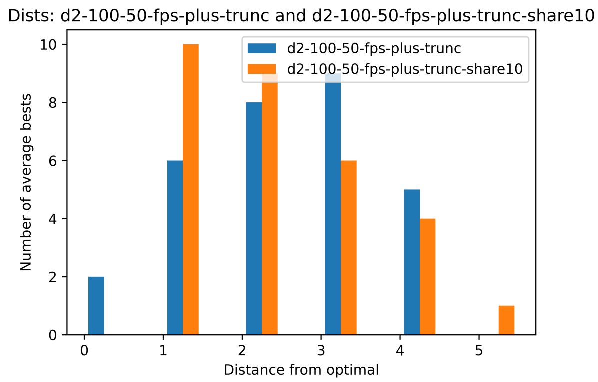


Table 43: F-Test for d2-100-50-fps-plus-trunc vs. d2-100-50-fps-plus-trunc-share10 with $\alpha = 0.05$

	d2-100-50-fps-plus-trunc	d2-100-50-fps-plus-trunc-share10
Mean	2.3	2.2333333333333334
Variance	1.3896551724137933	1.3574712643678162
Observations	30	30
df	29	29
F	1.0237087214225233	
P(F≤f) one-tail	0.5249377827583704	
F Critical one-tail	0.5373999648406917	

abs(mean 1) > abs(mean 2) and F > F Critical implies unequal variances.

Table 44: t-Test for d2-100-50-fps-plus-trunc vs. d2-100-50-fps-plus-trunc-share10 with Unequal Variances

	d2-100-50-fps-plus-trunc	d2-100-50-fps-plus-trunc-share10
Mean	2.3	2.2333333333333334
Variance	1.3896551724137933	1.3574712643678162
Observations	30	30
df	58	29
t Stat	0.22030788649877278	
P(T≤t) two-tail	0.8264051189296409	
t Critical two-tail	2.0017174830120923	

abs(t Stat) < abs(t Critical two-tail) so we accept the null hypothesis – the two samples are NOT statistically different.

4.2.1.2 With and Without Fitness Sharing; $\sigma = 50$

For both d1 and d2, using Fitness Sharing with $\sigma = 50$ is not statistically different than not using it.

Figure 28: d1-100-50-fps-plus-trunc vs. d1-100-50-fps-plus-trunc-share50 – Distance from best values to optimal over 30 runs

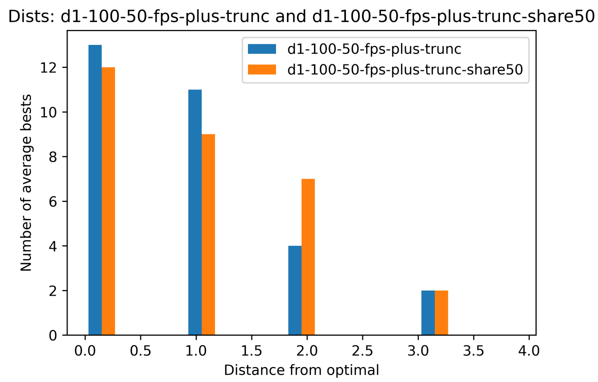


Table 45: F-Test for d1-100-50-fps-plus-trunc vs. d1-100-50-fps-plus-trunc-share50 with $\alpha = 0.05$

	d1-100-50-fps-plus-trunc	d1-100-50-fps-plus-trunc-share50
Mean	0.8333333333333334	0.9666666666666667
Variance	0.8333333333333331	0.9298850574712645
Observations	30	30
df	29	29
F	0.8961681087762667	
P(F≤f) one-tail	0.3849236527047533	
F Critical one-tail	0.5373999648406917	

abs(mean 1) < abs(mean 2) and F > F Critical implies equal variances.

Table 46: t-Test for d1-100-50-fps-plus-trunc vs. d1-100-50-fps-plus-trunc-share50 with Equal Variances

	d1-100-50-fps-plus-trunc	d1-100-50-fps-plus-trunc-share50
Mean	0.8333333333333334	0.9666666666666667
Variance	0.8333333333333331	0.9298850574712645
Observations	30	30
df	58	29
t Stat	-0.549979257638973	
P(T≤t) two-tail	0.584445410431059	
t Critical two-tail	2.0017174830120923	

abs(t Stat) < abs(t Critical two-tail) so we accept the null hypothesis – the two samples are NOT statistically different.

Figure 29: d2-100-50-fps-plus-trunc vs. d2-100-50-fps-plus-trunc-share50 – Distance from best values to optimal over 30 runs

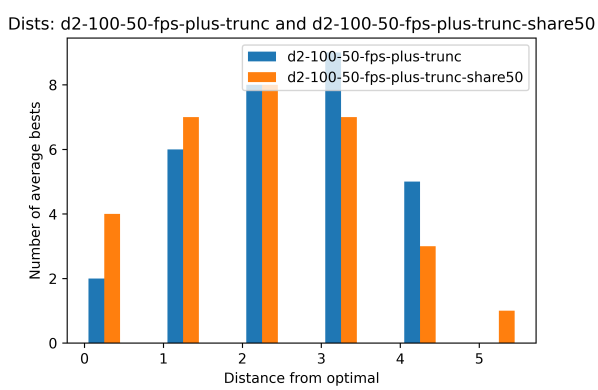


Table 47: F-Test for d2-100-50-fps-plus-trunc vs. d2-100-50-fps-plus-trunc-share50 with $\alpha = 0.05$

	d2-100-50-fps-plus-trunc	d2-100-50-fps-plus-trunc-share50
Mean	2.3	2.0333333333333333
Variance	1.3896551724137933	1.757471264367816
Observations	30	30
df	29	29
F	0.7907128842380642	
P(F≤f) one-tail	0.26561305445616135	
F Critical one-tail	0.5373999648406917	

abs(mean 1) > abs(mean 2) and F > F Critical implies unequal variances.

Table 48: t-Test for d2-100-50-fps-plus-trunc vs. d2-100-50-fps-plus-trunc-share50 with Unequal Variances

	d2-100-50-fps-plus-trunc	d2-100-50-fps-plus-trunc-share50
Mean	2.3	2.0333333333333333
Variance	1.3896551724137933	1.757471264367816
Observations	30	30
df	58	29
t Stat	0.8233268222555574	
P(T≤t) two-tail	0.4137406170020437	
t Critical two-tail	2.0017174830120923	

abs(t Stat) < abs(t Critical two-tail) so we accept the null hypothesis – the two samples are NOT statistically different.

4.2.2 Diversity

4.2.2.1 With and Without Fitness Sharing; $\sigma = 10$

For both d1 and d2, using Fitness Sharing with $\sigma = 10$ is not statistically different than not using it.

Figure 30: d1-100-50-fps-plus-trunc vs. d1-100-50-fps-plus-trunc-share10 – Diversity (Crowding distance)

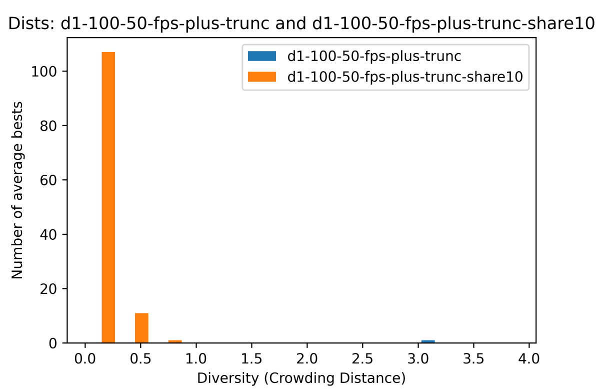


Table 49: F-Test for d1-100-50-fps-plus-trunc vs. d1-100-50-fps-plus-trunc-share10 with $\alpha = 0.05$

	d1-100-50-fps-plus-trunc	d1-100-50-fps-plus-trunc-share10
Mean	3.0	0.05042016806722689
Variance	nan	0.020035132697146657
Observations	1	1
df	0	0
F	nan	
P(F≤f) one-tail	nan	
F Critical one-tail	nan	

Table 50: t-Test for d1-100-50-fps-plus-trunc vs. d1-100-50-fps-plus-trunc-share10 with Unequal Variances

	d1-100-50-fps-plus-trunc	d1-100-50-fps-plus-trunc-share10
Mean	3.0	0.05042016806722689
Variance	nan	0.020035132697146657
Observations	1	1
df	0	0
t Stat	nan	
P(T≤t) two-tail	nan	
t Critical two-tail	nan	

abs(t Stat) < abs(t Critical two-tail) so we accept the null hypothesis – the two samples are NOT statistically different.

Figure 31: d2-100-50-fps-plus-trunc vs. d2-100-50-fps-plus-trunc-share10 – Diversity (Crowding distance)

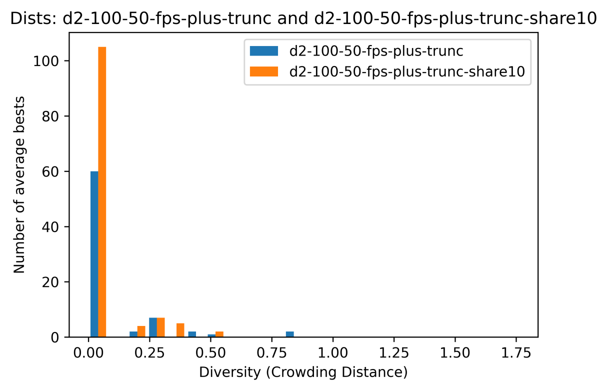


Table 51: F-Test for d2-100-50-fps-plus-trunc vs. d2-100-50-fps-plus-trunc-share10 with $\alpha = 0.05$

	d2-100-50-fps-plus-trunc	d2-100-50-fps-plus-trunc-share10
Mean	0.07094594594594594	0.0440379403794038
Variance	0.029247038134024444	0.012863373169250817
Observations	74	74
df	73	73
F	2.273667859060318	
P(F≤f) one-tail	0.9997194920598562	
F Critical one-tail	0.6787174812558107	

abs(mean 1) > abs(mean 2) and F > F Critical implies unequal variances.

Table 52: t-Test for d2-100-50-fps-plus-trunc vs. d2-100-50-fps-plus-trunc-share10 with Unequal Variances

	d2-100-50-fps-plus-trunc	d2-100-50-fps-plus-trunc-share10
Mean	0.07094594594594594	0.0440379403794038
Variance	0.029247038134024444	0.012863373169250817
Observations	74	74
df	146	73
t Stat	1.203590676038513	
P(T≤t) two-tail	0.2312844811199268	
t Critical two-tail	1.9763456545827003	

abs(t Stat) < abs(t Critical two-tail) so we accept the null hypothesis – the two samples are NOT statistically different.

4.2.2.2 With and Without Fitness Sharing; $\sigma = 50$

For both d1 and d2, using Fitness Sharing with $\sigma = 50$ is not statistically different than not using it.

Figure 32: d1-100-50-fps-plus-trunc vs. d1-100-50-fps-plus-trunc-share50 – Diversity (Crowding distance)

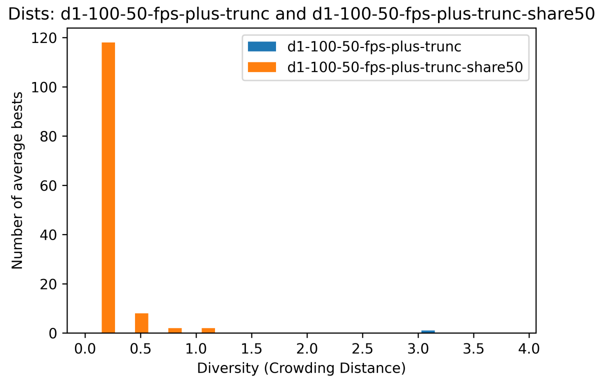


Table 53: F-Test for d1-100-50-fps-plus-trunc vs. d1-100-50-fps-plus-trunc-share50 with $\alpha = 0.05$

	d1-100-50-fps-plus-trunc	d1-100-50-fps-plus-trunc-share50
Mean	3.0	0.046153846153846156
Variance	nan	0.02713840853375737
Observations	1	1
df	0	0
F	nan	
P(F≤f) one-tail	nan	
F Critical one-tail	nan	

Table 54: t-Test for d1-100-50-fps-plus-trunc vs. d1-100-50-fps-plus-trunc-share50 with Unequal Variances

	d1-100-50-fps-plus-trunc	d1-100-50-fps-plus-trunc-share50
Mean	3.0	0.046153846153846156
Variance	nan	0.02713840853375737
Observations	1	1
df	0	0
t Stat	nan	
P(T≤t) two-tail	nan	
t Critical two-tail	nan	

abs(t Stat) < abs(t Critical two-tail) so we accept the null hypothesis – the two samples are NOT statistically different.

Figure 33: d2-100-50-fps-plus-trunc vs. d2-100-50-fps-plus-trunc-share50 – Diversity (Crowding distance)

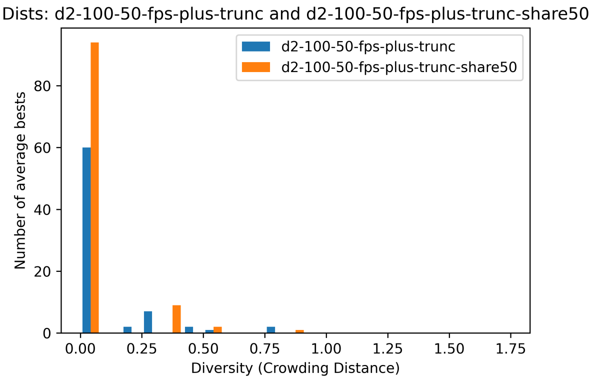


Table 55: F-Test for d2-100-50-fps-plus-trunc vs. d2-100-50-fps-plus-trunc-share50 with $\alpha = 0.05$

	d2-100-50-fps-plus-trunc	d2-100-50-fps-plus-trunc-share50
Mean	0.07094594594594594	0.04559748427672955
Variance	0.029247038134024444	0.018800539083557957
Observations	74	74
df	73	73
F	1.555648910067823	
P(F≤f) one-tail	0.9695188404001582	
F Critical one-tail	0.6787174812558107	

abs(mean 1) > abs(mean 2) and F > F Critical implies unequal variances.

Table 56: t-Test for d2-100-50-fps-plus-trunc vs. d2-100-50-fps-plus-trunc-share50 with Unequal Variances

	d2-100-50-fps-plus-trunc	d2-100-50-fps-plus-trunc-share50
Mean	0.07094594594594594	0.04559748427672955
Variance	0.029247038134024444	0.018800539083557957
Observations	74	74
df	146	73
t Stat	1.059322701997895	
P(T≤t) two-tail	0.29135290014690185	
t Critical two-tail	1.9763456545827003	

abs(t Stat) < abs(t Critical two-tail) so we accept the null hypothesis – the two samples are NOT statistically different.

4.3 Conclusion

The author believes the implementation matches that presented in the literature, however, not only does it not produce statistically significant results in these experiments, it increases computation time by very significant amounts.

5 YELLOW 2: Crowding

5.1 Implementation

Crowding was implemented closely following the pseudocode presented by Deb et. al.[1]. Following this definition, Crowding incorporates survivor selection–so when using this option, the user-specific survivor selection choice is ignored.

The new fitness values were calculated as indicated in this updated MOEA outline:

- Create initial population P_0 with or without validity forced (user-configurable)
- Sort P_0 on the basis of non-domination using the NSGA-II "fast-non-dominated-sort" adapted to the data structures of this MOEA implementation
- Best level is level 0
- Fitness is set to level number; lower number, lower fitness (all fitness values are 0 or negative)
- Primary generational loop:
  - Housekeeping (write logs, check for termination)
  - User-configurable parent selection then single-point crossover combination and bitwise mutation make the next Q (first is Q_0)
  - User configurable survival strategy determines R
    - Plus: $R = P + Q$
    - Comma: $R = Q$
  - Sort R on the basis of non-domination
  - **Calculate crowding and select survivors to determine next P from R**

5.2 Experiments

Experiments considered distance and diversity for performance on problems d1 and d2 for the baseline configuration with and without Crowding.

5.2.1 Distance

5.2.1.1 With and Without Crowding

For d1, using Crowding as implemented produces a small but statistically-significant improvement. For d2, the same holds true, but with a larger improvement.

Figure 34: d1-100-50-fps-plus-trunc vs. d1-100-50-fps-plus-trunc-crowd – Distance from best values to optimal over 30 runs

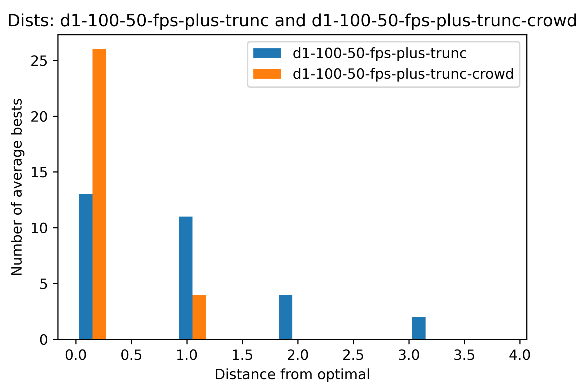


Table 57: F-Test for d1-100-50-fps-plus-trunc vs. d1-100-50-fps-plus-trunc-crowd with $\alpha = 0.05$

	d1-100-50-fps-plus-trunc	d1-100-50-fps-plus-trunc-crowd
Mean	0.8333333333333334	0.13333333333333333
Variance	0.8333333333333331	0.11954022988505751
Observations	30	30
df	29	29
F	6.971153846153842	
P(F≤f) one-tail	0.9999993766567417	
F Critical one-tail	0.5373999648406917	

abs(mean 1) > abs(mean 2) and F > F Critical implies unequal variances.

Table 58: t-Test for d1-100-50-fps-plus-trunc vs. d1-100-50-fps-plus-trunc-crowd with Unequal Variances

	d1-100-50-fps-plus-trunc	d1-100-50-fps-plus-trunc-crowd
Mean	0.8333333333333334	0.13333333333333333
Variance	0.8333333333333331	0.11954022988505751
Observations	30	30
df	58	29
t Stat	3.927724596586999	
P(T≤t) two-tail	0.0003587916194724198	
t Critical two-tail	2.0017174830120923	

abs(t Stat) > abs(t Critical two-tail) so we reject the null hypothesis – the two samples are statistically different. The average improvement of d1-100-50-fps-plus-trunc over d1-100-50-fps-plus-trunc-crowd is -0.7000000000000001 closer.

Figure 35: d2-100-50-fps-plus-trunc vs. d2-100-50-fps-plus-trunc-crowd – Distance from best values to optimal over 30 runs

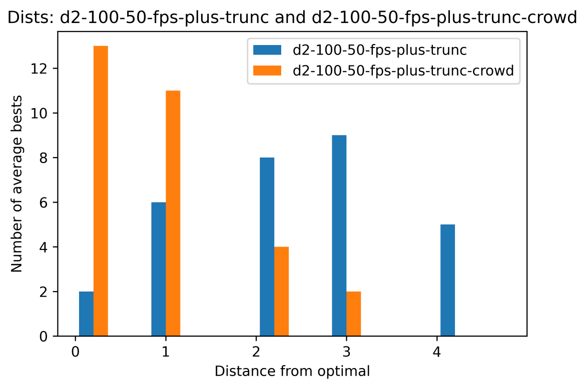


Table 59: F-Test for d2-100-50-fps-plus-trunc vs. d2-100-50-fps-plus-trunc-crowd with $\alpha = 0.05$

	d2-100-50-fps-plus-trunc	d2-100-50-fps-plus-trunc-crowd
Mean	2.3	0.8333333333333334
Variance	1.3896551724137933	0.8333333333333331
Observations	30	30
df	29	29
F	1.6675862068965523	
P(F≤f) one-tail	0.9127432385438341	
F Critical one-tail	0.5373999648406917	

abs(mean 1) > abs(mean 2) and F > F Critical implies unequal variances.

Table 60: t-Test for d2-100-50-fps-plus-trunc vs. d2-100-50-fps-plus-trunc-crowd with Unequal Variances

	d2-100-50-fps-plus-trunc	d2-100-50-fps-plus-trunc-crowd
Mean	2.3	0.8333333333333334
Variance	1.3896551724137933	0.8333333333333331
Observations	30	30
df	58	29
t Stat	5.387948557543803	
P(T≤t) two-tail	1.559204791216041e-06	
t Critical two-tail	2.0017174830120923	

abs(t Stat) > abs(t Critical two-tail) so we reject the null hypothesis – the two samples are statistically different. The average improvement of d2-100-50-fps-plus-trunc over d2-100-50-fps-plus-trunc-crowd is -1.4666666666666663 closer.

5.2.2 Diversity

5.2.2.1 With and Without Crowding

For d1, there is no statistical difference in diversity. For d2, using Crowding is slightly but statistically significantly more diverse than not using it.

Figure 36: d1-100-50-fps-plus-trunc vs. d1-100-50-fps-plus-trunc-crowd – Diversity (Crowding distance)

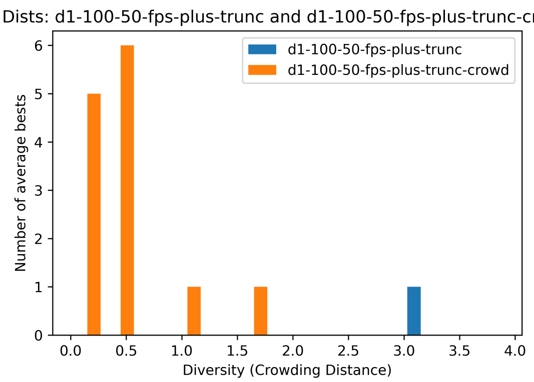


Table 61: F-Test for d1-100-50-fps-plus-trunc vs. d1-100-50-fps-plus-trunc-crowd with $\alpha = 0.05$

	d1-100-50-fps-plus-trunc	d1-100-50-fps-plus-trunc-crowd
Mean	3.0	0.4230769230769231
Variance	nan	0.2019230769230769
Observations	1	1
df	0	0
F	nan	
P(F≤f) one-tail	nan	
F Critical one-tail	nan	

Table 62: t-Test for d1-100-50-fps-plus-trunc vs. d1-100-50-fps-plus-trunc-crowd with Unequal Variances

	d1-100-50-fps-plus-trunc	d1-100-50-fps-plus-trunc-crowd
Mean	3.0	0.4230769230769231
Variance	nan	0.2019230769230769
Observations	1	1
df	0	0
t Stat	nan	
P(T≤t) two-tail	nan	
t Critical two-tail	nan	

abs(t Stat) < abs(t Critical two-tail) so we accept the null hypothesis – the two samples are NOT statistically different.

Figure 37: d2-100-50-fps-plus-trunc vs. d2-100-50-fps-plus-trunc-crowd – Diversity (Crowding distance)

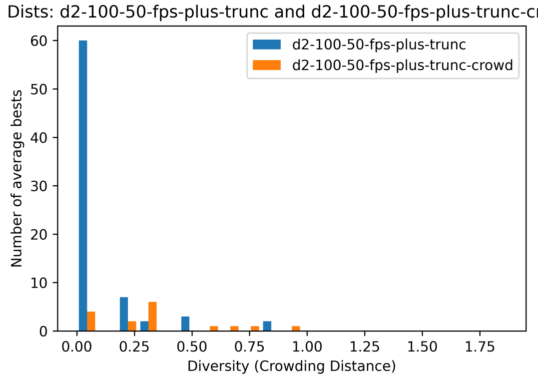


Table 63: F-Test for d2-100-50-fps-plus-trunc vs. d2-100-50-fps-plus-trunc-crowd with $\alpha = 0.05$

	d2-100-50-fps-plus-trunc	d2-100-50-fps-plus-trunc-crowd
Mean	0.07094594594594594	0.33333333333333337
Variance	0.029247038134024444	0.0757201646090535
Observations	74	74
df	73	73
F	0.3862516449221707	
P(F≤f) one-tail	3.4825318270081154e-05	
F Critical one-tail	0.6787174812558107	

abs(mean 1) < abs(mean 2) and F < F Critical implies unequal variances.

Table 64: t-Test for d2-100-50-fps-plus-trunc vs. d2-100-50-fps-plus-trunc-crowd with Unequal Variances

	d2-100-50-fps-plus-trunc	d2-100-50-fps-plus-trunc-crowd
Mean	0.07094594594594594	0.33333333333333337
Variance	0.029247038134024444	0.0757201646090535
Observations	74	74
df	146	73
t Stat	-3.6642064469941213	
P(T≤t) two-tail	0.0018327598806554257	
t Critical two-tail	1.9763456545827003	

abs(t Stat) > abs(t Critical two-tail) so we reject the null hypothesis – the two samples are statistically different. The average improvement of d2-100-50-fps-plus-trunc over d2-100-50-fps-plus-trunc-crowd is -0.2623873873873874 more diverse.

5.3 Conclusion

Crowding allows tie-breaking in between genotypes that have the same level in that it favors diversity–if two individuals don't dominate each other, pick the one that increases diversity in order to favor exploration of the search space. In this experiment, the approach produced improvements over the baseline configuration.

6 RED 1: Fourth Objective to Minimize Bulbs Placed

6.1 Implementation

The MOEA code natively handles all four objectives in its data structures. A configuration option forces calculations to use or ignore the fourth objective (i.e. in the default three-objective mode, the fourth objective value is an ignored placeholder).

6.2 Experiments

Experiments considered distance and diversity for performance on problems d1 and d2 for the baseline configuration with and without the fourth objective.

6.2.1 Distance

6.2.1.1 With and Without the Fourth Objective

For both d1 and d2, including the fourth objective performed statistically significantly worse than excluding it.

Figure 38: d1-100-50-fps-plus-trunc vs. d1-100-50-fps-plus-trunc-4obj – Distance from best values to optimal over 30 runs

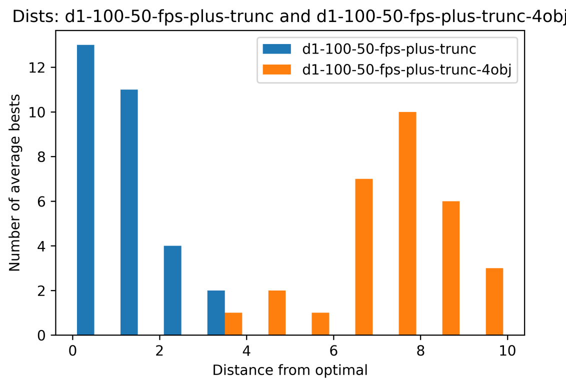


Table 65: F-Test for d1-100-50-fps-plus-trunc vs. d1-100-50-fps-plus-trunc-4obj with $\alpha = 0.05$

	d1-100-50-fps-plus-trunc	d1-100-50-fps-plus-trunc-4obj
Mean	0.8333333333333334	6.8
Variance	0.8333333333333331	2.3034482758620696
Observations	30	30
df	29	29
F	0.36177644710578827	
P(F≤f) one-tail	0.003914756921736153	
F Critical one-tail	0.5373999648406917	

abs(mean 1) < abs(mean 2) and F < F Critical implies unequal variances.

Table 66: t-Test for d1-100-50-fps-plus-trunc vs. d1-100-50-fps-plus-trunc-4obj with Unequal Variances

	d1-100-50-fps-plus-trunc	d1-100-50-fps-plus-trunc-4obj
Mean	0.8333333333333334	6.8
Variance	0.8333333333333331	2.3034482758620696
Observations	30	30
df	58	29
t Stat	-18.452289602431573	
P(T≤t) two-tail	2.575230927925143e-23	
t Critical two-tail	2.0017174830120923	

abs(t Stat) > abs(t Critical two-tail) so we reject the null hypothesis – the two samples are statistically different. The average improvement of d1-100-50-fps-plus-trunc over d1-100-50-fps-plus-trunc-4obj is 5.966666666666667 closer.

Figure 39: d2-100-50-fps-plus-trunc vs. d2-100-50-fps-plus-trunc-4obj – Distance from best values to optimal over 30 runs

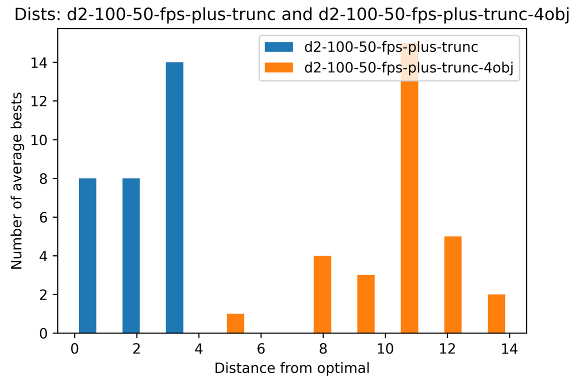


Table 67: F-Test for d2-100-50-fps-plus-trunc vs. d2-100-50-fps-plus-trunc-4obj with $\alpha = 0.05$

	d2-100-50-fps-plus-trunc	d2-100-50-fps-plus-trunc-4obj
Mean	2.3	10.2
Variance	1.3896551724137933	3.1999999999999997
Observations	30	30
df	29	29
F	0.43426724137931044	
P(F≤f) one-tail	0.01405313284385904	
F Critical one-tail	0.5373999648406917	

abs(mean 1) < abs(mean 2) and F < F Critical implies unequal variances.

Table 68: t-Test for d2-100-50-fps-plus-trunc vs. d2-100-50-fps-plus-trunc-4obj with Unequal Variances

	d2-100-50-fps-plus-trunc	d2-100-50-fps-plus-trunc-4obj
Mean	2.3	10.2
Variance	1.3896551724137933	3.1999999999999997
Observations	30	30
df	58	29
t Stat	-20.197503397725345	
P(T≤t) two-tail	9.393350996031728e-26	
t Critical two-tail	2.0017174830120923	

abs(t Stat) > abs(t Critical two-tail) so we reject the null hypothesis – the two samples are statistically different. The average improvement of d2-100-50-fps-plus-trunc over d2-100-50-fps-plus-trunc-4obj is 7.8999999999999995 closer.

6.2.2 Diversity

6.2.2.1 With and Without the Fourth Objective

For both d1 and d2, adding the fourth objective is not statistically different than excluding it with respect to diversity.

Figure 40: d1-100-50-fps-plus-trunc vs. d1-100-50-fps-plus-trunc-4obj – Diversity (Crowding distance)

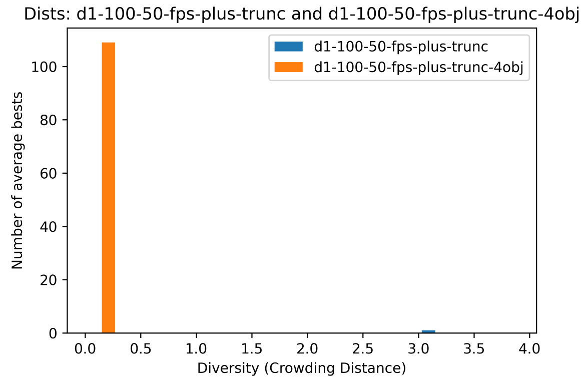


Table 69: F-Test for d1-100-50-fps-plus-trunc vs. d1-100-50-fps-plus-trunc-4obj with $\alpha = 0.05$

	d1-100-50-fps-plus-trunc	d1-100-50-fps-plus-trunc-4obj
Mean	3.0	0.05329365940033428
Variance	nan	0.0037825817525590916
Observations	1	1
df	0	0
F	nan	
P(F≤f) one-tail	nan	
F Critical one-tail	nan	

Table 70: t-Test for d1-100-50-fps-plus-trunc vs. d1-100-50-fps-plus-trunc-4obj with Unequal Variances

	d1-100-50-fps-plus-trunc	d1-100-50-fps-plus-trunc-4obj
Mean	3.0	0.05329365940033428
Variance	nan	0.0037825817525590916
Observations	1	1
df	0	0
t Stat	nan	
P(T≤t) two-tail	nan	
t Critical two-tail	nan	

abs(t Stat) < abs(t Critical two-tail) so we accept the null hypothesis – the two samples are NOT statistically different.

Figure 41: d2-100-50-fps-plus-trunc vs. d2-100-50-fps-plus-trunc-4obj – Diversity (Crowding distance)

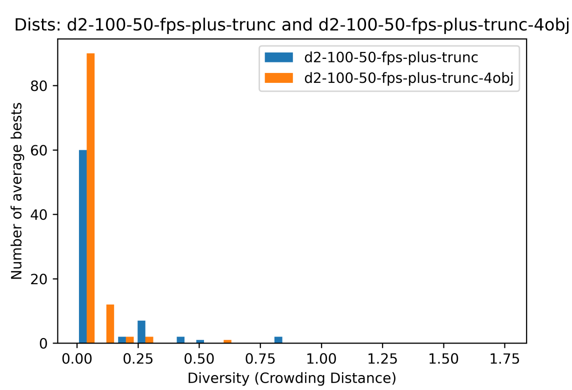


Table 71: F-Test for d2-100-50-fps-plus-trunc vs. d2-100-50-fps-plus-trunc-4obj with $\alpha = 0.05$

	d2-100-50-fps-plus-trunc	d2-100-50-fps-plus-trunc-4obj
Mean	0.07094594594594594	0.05218236474628645
Variance	0.029247038134024444	0.006049057231198919
Observations	74	74
df	73	73
F	4.83497461111435	
P(F≤f) one-tail	0.9999999999245853	
F Critical one-tail	0.6787174812558107	

abs(mean 1) > abs(mean 2) and F > F Critical implies unequal variances.

Table 72: t-Test for d2-100-50-fps-plus-trunc vs. d2-100-50-fps-plus-trunc-4obj with Unequal Variances

	d2-100-50-fps-plus-trunc	d2-100-50-fps-plus-trunc-4obj
Mean	0.07094594594594594	0.05218236474628645
Variance	0.029247038134024444	0.006049057231198919
Observations	74	74
df	146	73
t Stat	0.8827956124049291	
P(T≤t) two-tail	0.37959754280781577	
t Critical two-tail	1.9763456545827003	

abs(t Stat) < abs(t Critical two-tail) so we accept the null hypothesis – the two samples are NOT statistically different.

6.3 Conclusion

The author speculates that adding a fourth competing objective of minimizing bulbs placed fights against the the fact that as many bulbs as possible need to be placed (without violating Light Up rules) to maximize the number of cells lit up. The histograms for distance from optimal show a stark difference in performance (for the worse) when it is added.

References

[1] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan. A fast and elitist multiobjective genetic algorithm: Nsga-ii. *IEEE Transactions on Evolutionary Computation*, 6(2):182–197, April 2002.

[2] A. E. Eiben and James E. Smith. *Introduction to Evolutionary Computing, Second Edition.* Natural Computing Series. Springer, 2015.

[3] Simon Tatham. *Light Up, from Simon Tatham's Portable Puzzle Collection*, 2020 (accessed October 11, 2020). https://www.chiark.greenend.org.uk/ sgtatham/puzzles/js/lightup.html.