CS5401 FS2018 Assignment 1c with bonus

William Lorey
wwlytc@mst.edu

Contents

Introduction

Assignment 1c involved implementing an EA leveraging constraint satisfaction to more effectively solve Light Up puzzles. This report compares the performance of penalty function and repair function constraint satisfaction techniques as well as plain-vanilla EA performance. It also outlines a comparison between Validity Forced plus Uniform Random versus plain Uniform Random initialization for a plain-vanilla EA, a constraint satisfaction EA employing a penalty function, and a constraint satisfaction EA employing a repair function. The extent to which EA performance is affected by the penalty coefficient used in penalty function constraint satisfaction EAs will also be examined.

Constraint Satisfaction EA Performance Comparisons

For the baseline assignment, an EA leveraging a penalty function was implemented. The penalty function subtracted from a given genotype's fitness the number of constraints violated multiplied by a penalty coefficient. Violated constraints involved bulbs shining on other bulbs and black cell constraints not being met.

This EA was compared to an EA implemented for the bonus assignment that used the repair function constraint satisfaction technique. Figure 1 and Figure 2 show evaluations versus fitness for the penalty function EA and the repair function EA respectively tested against the Light Up puzzle provided on the course website. Figure 7 and Figure 8 show evaluations versus fitness for the aforementioned EAs tested against randomly generated puzzles. Note that for situations where the best fitness for a solution plateaus, there is possibility of encountering a decreasing average fitness as higher mutation factors are introduced to combat a stagnant population.

Through visually examining each configuration's fitness plots, it appears that the penalty function EA outperforms the repair function EA. However, statistical analysis seen in Table 1 and Table 2 proves that for randomly generated puzzles, neither constraint satisfaction method yields a more effective EA. Note that the table headers correspond to the configuration files used to generate the data upon which statistical analysis was performed and that all statistical analysis consisted of the F-Test Two-Sample for Variances followed by the t-test Two-Sample Assuming either Unequal or Equal Variances depending on the F-Test output. A performance boost was proven when examining the provided puzzle tests, proving that the penalty function EA was in fact optimal when compared to the repair function EA for that problem.

The vanilla EA tested on a validity enforced initialization (Figure 3 and Figure 9) performed very poorly compared to the constraint satisfaction EAs. This is because with black cell adjacency constraints enforced, invalid solutions produce fitness values of zero. With many invalid solutions dominating the population, no search gradient is provided and the EA struggles to find any solution.

Initialization Comparisons

The effect of Validity Forced plus Uniform Random versus plain Uniform Random was examined regarding the plain-vanilla EA, the penalty function EA, and the repair function EA.

The comparison for the plain-vanilla EA tested against the randomly generated puzzle (graphed in Figure 12 and Figure 9, and analyzed in Table 3) proved that neither initialization method led to quantifiable improvements. Granted, neither plain-vanilla EA configuration managed to reliably find meaningful solutions.

Figure 7 and Figure 10 illustrate the penalty function EA configuration tested with both Uniform Random and Validity Enforced plus Uniform Random for randomly generated puzzles. Figure 1 and Figure 4 illustrate the same comparison for the provided puzzle. The statistical analysis comparing these two configurations can be found in Table 4 and Table 5.

This statistical analysis showed that for randomly generated puzzles, neither initialization method made a quantifiable improvement on the EA's results. Randomly generated puzzles are generally less complicated to solve, so the Validity Enforced initialization's lack of impact in this case is justified.

The analysis also showed that the EA was more sensitive to initialization configuration for the provided puzzle, an objectively difficult puzzle to solve. The Validity Enforced Uniform Random initialization configured EA proved to be statistically better than the Uniform Random initialization configured EA for this problem instance. Since enforcing validity upon initialization shrinks the search space by ensuring the validity of conditions that can only be fulfilled one way (such as a '3' valued black cell placed along an edge leads to one single way to ensure that constraint is valid), and since the search space is more difficult to traverse for highly difficult puzzles, Validity Enforced Uniform Random initialization produces a higher performing EA for the provided puzzle.

The Penalty Coefficient and its Effect on Solution Quality

The penalty function EA, tested against the provided puzzle, was configured using three different penalty coefficients to test their effect on the EA's performance. The first coefficient tested was a small coefficient, equating to 0.25. The second was the medium coefficient,

Table 1: Statistical Analysis performed on the Validity Enforced Penalty Function and Repair function, Randomly Generated Puzzle, EA configurations

	random_gen_validity_enforced	random_gen_validity_enforced_bonus
mean	0.9081384890909392	0.9188060294943169
variance	0.002217915192875166	0.060460786982912955
standard deviation	0.047094746977504466	0.24588775281195474
observations	30	30
df	29	29
F	0.036683531650059054	
F critical	0.5373999648406917	
Unequal variances assumed		
observations	30	
df	31	
t Stat	-0.2294580495633768	
P two-tail	0.8200141447652132	
t Critical two-tail	2.0395	
Nether random_gen_validity_enforced_bonus nor random_gen_validity_enforced is statistically better		

equating to 3. The third was the large coefficient, equating to 10. The fitness graphs for each of these coefficients may be found in Figure 13, Figure 14, and Figure 15 respectively.

In order to properly compare each configuration, pairwise statistical analysis was performed between each pair of coefficient types (small, medium, and large). Analysis comparing the small and medium coefficients (Table 8) showed that the small penalty coefficient of 0.25 lead to an ea that outperformed the EA configured with the medium coefficient of 3. The comparison between the medium and large coefficients (Table 9) led to the conclusion that, again, the smaller coefficient configured EA outperformed the larger penalty coefficient configured EA. This conclusion was also drawn from the last comparison, comparing the small penalty coefficient to the large penalty coefficient (Table 10).

In each comparison case, the smaller penalty function led to a higher best fitness and a higher average population fitness when all run data was averaged. This is because when invalid individuals that do not meet all constraints are penalized, their fitness decreases by a small fraction for smaller penalty coefficients and by a larger fraction for larger penalty coefficients. The final population of a penalty function configured EA using a small penalty function may actually have the same number of invalid solutions as the population of an EA configured with a large penalty coefficient, but the small coefficient EA will read as having higher fitness values. With this fact in mind, it is also important to consider the reason a penalty function constraint satisfaction method is used. Employing a penalty function allows for invalid solutions to be examined with the hope that those invalid solutions lead to a valid solution after some exploration. This can be equated to traversing the 'sea' of validity on the way to a more optimal solution. Keeping the penalty coefficient relatively small incentivises the traversal of invalid areas of the search space, possibly allowing for a more optimal solution to be discovered.

Appendix: Additional Figures and Tables

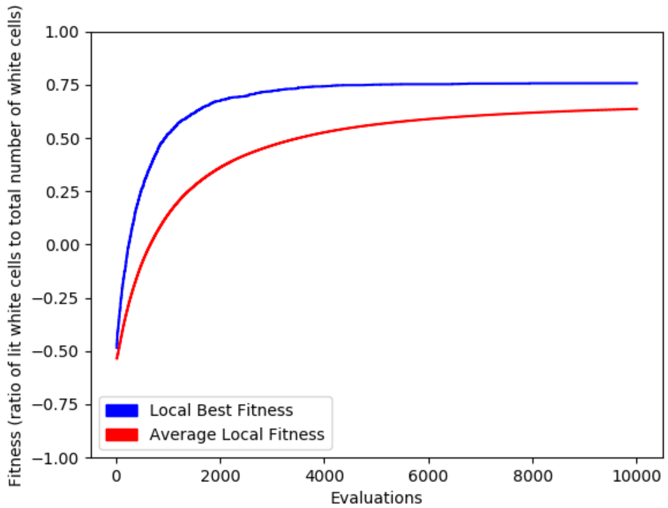

Figure 1: Evaluations versus Average Local Fitness and Evaluations versus Local Best Fitness for the **Penalty Function EA with the Validity Enforced, Provided Puzzle**, Averaged Over All Runs

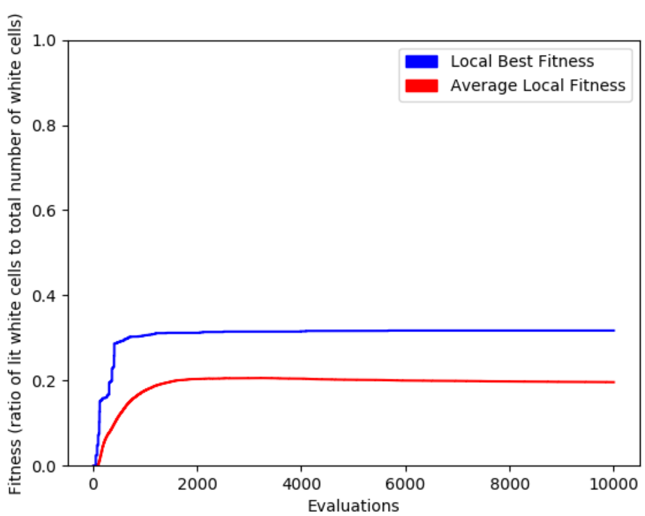

Figure 2: Evaluations versus Average Local Fitness and Evaluations versus Local Best Fitness for the **Repair Function EA with the Validity Enforced, Provided Puzzle**, Averaged Over All Runs

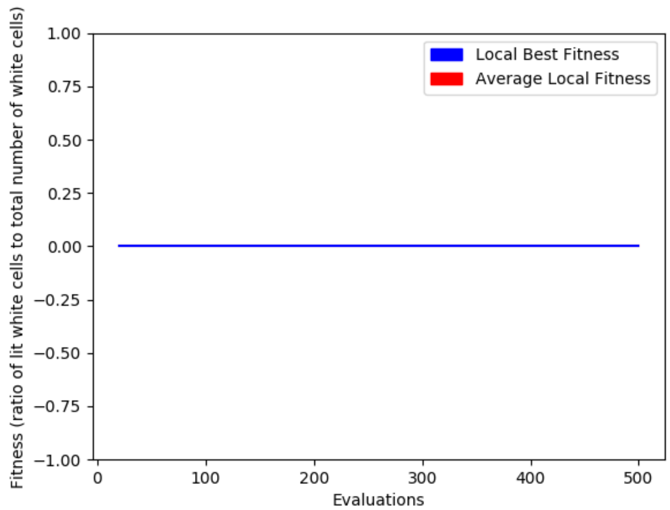

Figure 3: Evaluations versus Average Local Fitness and Evaluations versus Local Best Fitness for the **Plain-Vanilla EA with the Validity Enforced, Provided Puzzle**, Averaged Over All Runs

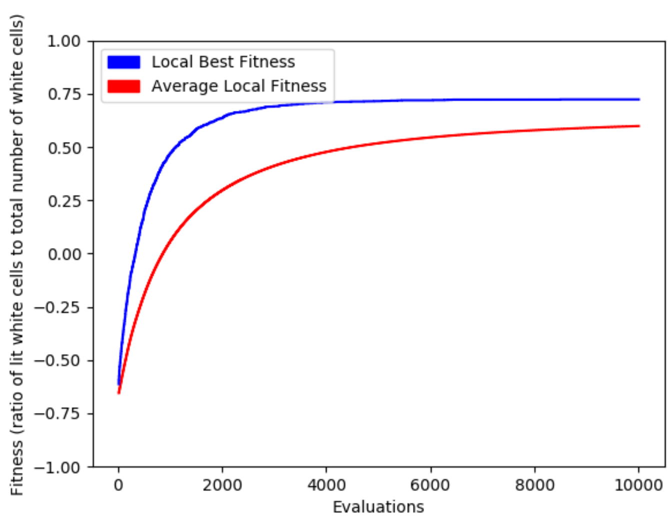

Figure 4: Evaluations versus Average Local Fitness and Evaluations versus Local Best Fitness for the **Penalty Function EA with the Uniform Random Initialized, Provided Puzzle**, Averaged Over All Runs

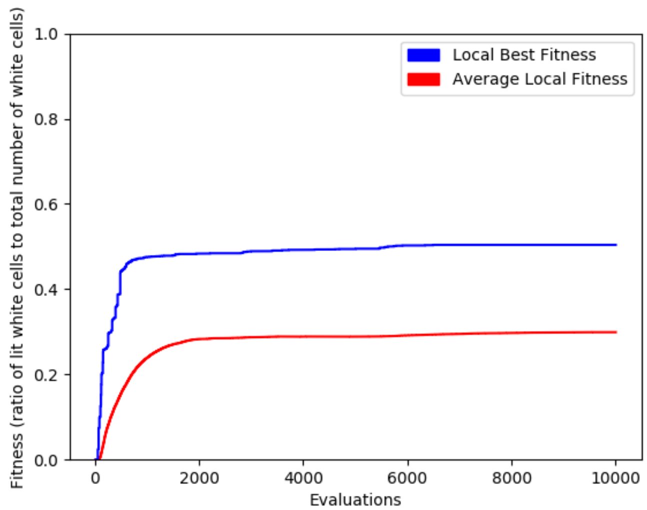

Figure 5: Evaluations versus Average Local Fitness and Evaluations versus Local Best Fitness for the **Repair Function EA with the Uniform Random Initialized, Provided Puzzle**, Averaged Over All Runs

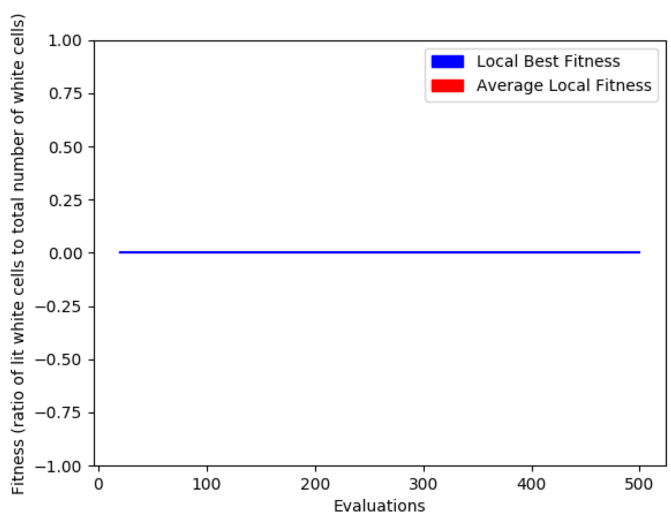

Figure 6: Evaluations versus Average Local Fitness and Evaluations versus Local Best Fitness for the **Plain-Vanilla EA with the Uniform Random Initialized, Provided Puzzle**, Averaged Over All Runs

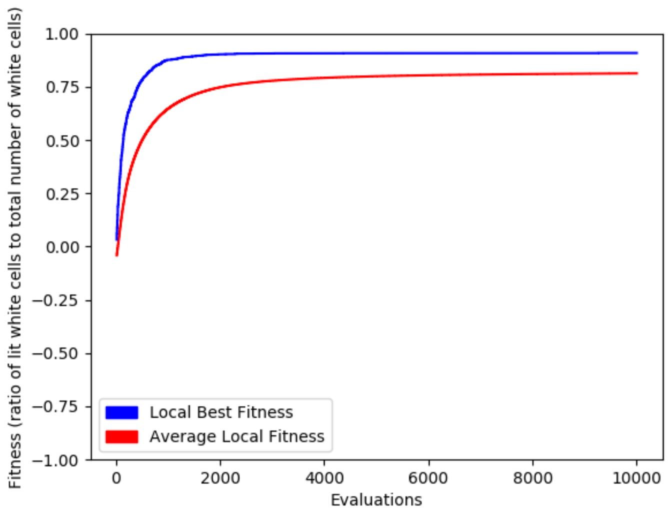

Figure 7: Evaluations versus Average Local Fitness and Evaluations versus Local Best Fitness for the **Penalty Function EA with the Validity Enforced, Randomly Generated Puzzle**, Averaged Over All Runs

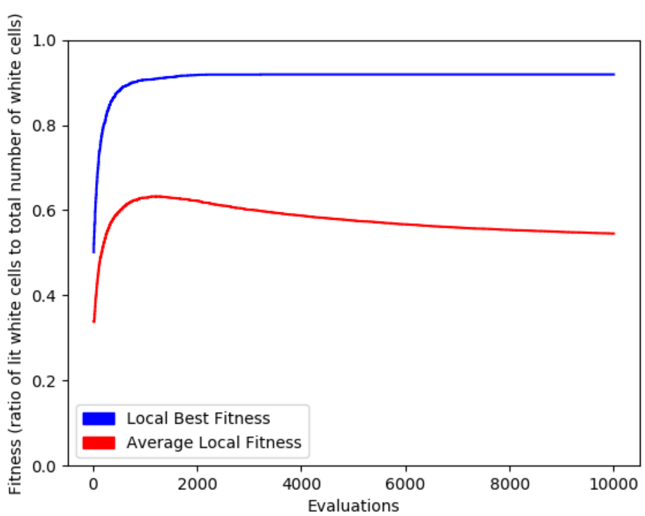

Figure 8: Evaluations versus Average Local Fitness and Evaluations versus Local Best Fitness for the **Repair Function EA with the Validity Enforced, Randomly Generated Puzzle**, Averaged Over All Runs

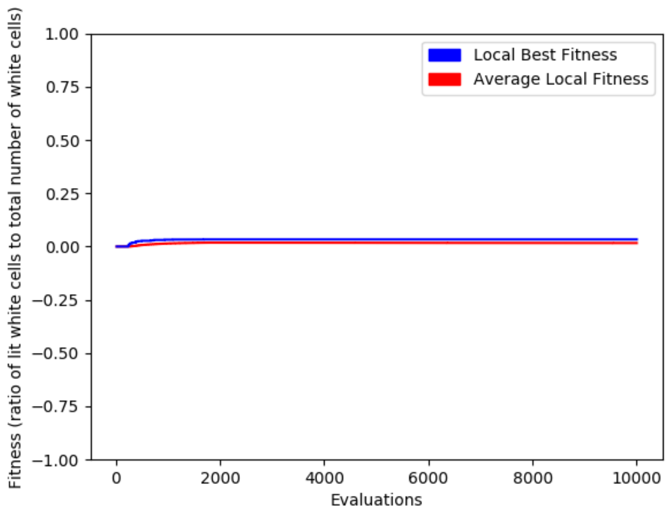

Figure 9: Evaluations versus Average Local Fitness and Evaluations versus Local Best Fitness for the **Plain-Vanilla EA with the Validity Enforced, Randomly Generated Puzzle**, Averaged Over All Runs

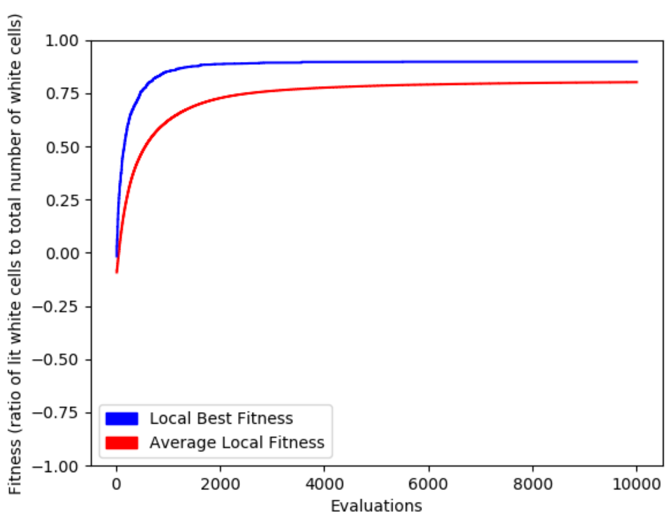

Figure 10: Evaluations versus Average Local Fitness and Evaluations versus Local Best Fitness for the **Penalty Function EA with the Uniform Random Initialized, Randomly Generated Puzzle**, Averaged Over All Runs

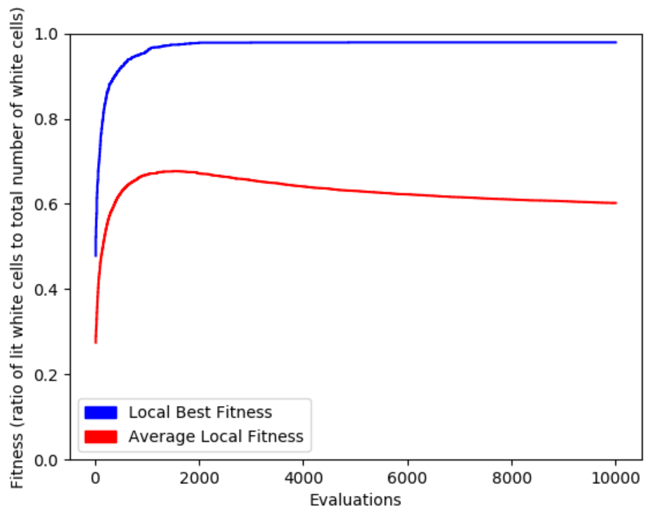

Figure 11: Evaluations versus Average Local Fitness and Evaluations versus Local Best Fitness for the **Repair Function EA with the Uniform Random Initialized, Randomly Generated Puzzle**, Averaged Over All Runs

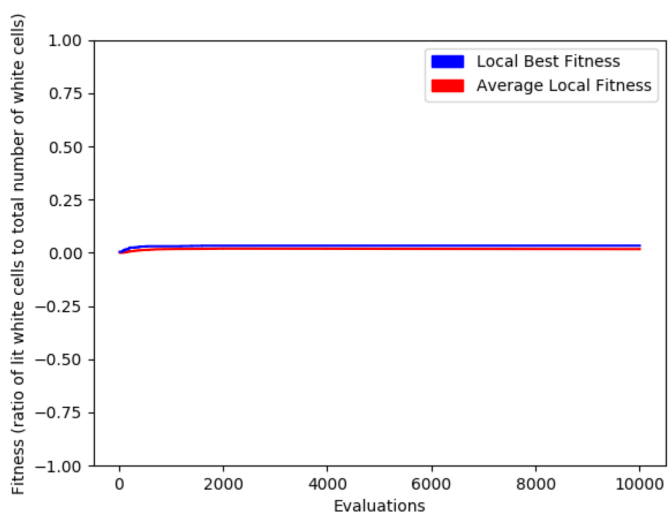

Figure 12: Evaluations versus Average Local Fitness and Evaluations versus Local Best Fitness for the **Plain-Vanilla EA with the Uniform Random Initialized, Randomly Generated Puzzle**, Averaged Over All Runs

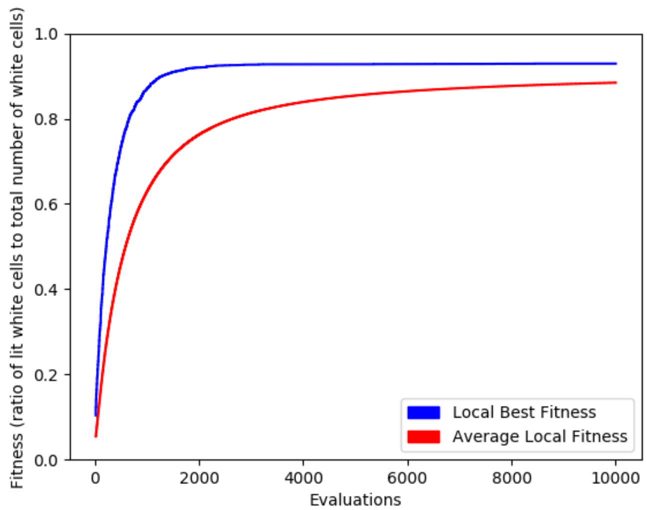

Figure 13: Evaluations versus Average Local Fitness and Evaluations versus Local Best Fitness for the **Penalty Function (with Small Penalty Coefficient = 0.25) EA with the Validity Enforced, Provided Puzzle**, Averaged Over All Runs

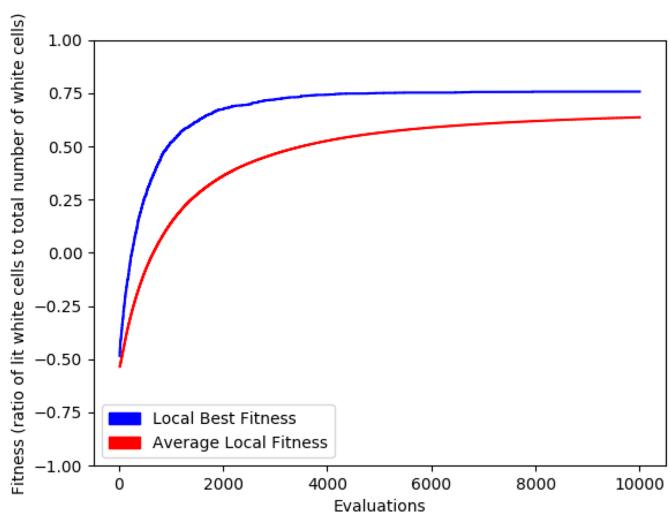

Figure 14: Evaluations versus Average Local Fitness and Evaluations versus Local Best Fitness for the **Penalty Function (with Large Penalty Coefficient = 3) EA with the Validity Enforced, Provided Puzzle**, Averaged Over All Runs

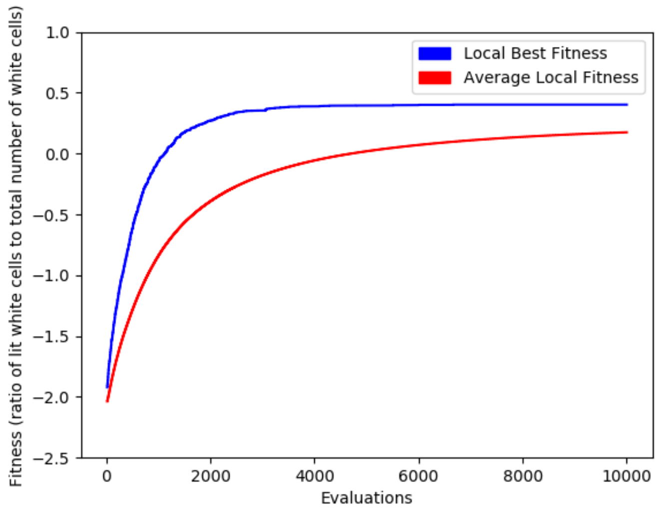

Figure 15: Evaluations versus Average Local Fitness and Evaluations versus Local Best Fitness for the **Penalty Function (with Large Penalty Coefficient = 10) EA with the Validity Enforced, Provided Puzzle**, Averaged Over All Runs

Table 2: Statistical Analysis performed on the Validity Enforced Penalty Function and Repair function, Provided Puzzle, EA configurations

	website_puzzle_validity_enforced	website_puzzle_validity_enforced_bonus
mean	0.7569819819819819	0.3173423423423423
variance	0.0033889801964126286	0.2015810912263615
standard deviation	0.05821494822133426	0.4489778293260832
observations	30	30
df	29	29
F	0.016811994497078302	
F critical	0.5373999648406917	
Equal variances assumed		
observations	30	
df	58	
t Stat	5.229384989438686	
P two-tail	2.4339170289675772e-06	
t Critical two-tail	2.0017	
website_puzzle_validity_enforced is statistically better than website_puzzle_validity_enforced_bonus		

Table 3: Statistical Analysis performed on the Uniform Random Initialized Vanilla and Validity Enforced Initalized Vanilla, Randomly Generated Puzzle, EA configurations

	random_gen_uniform_random_vanilla	random_gen_validity_enforced_vanilla
mean	0.03333333333333333	0.03333333333333333
variance	0.03222222222222215	0.03222222222222222
standard deviation	0.17950549357115012	0.17950549357115014
observations	30	30
df	29	29
F	0.9999999999999998	
F critical	0.5373999648406917	
Unequal variances assumed		
observations	30	
df	31	
t Stat	0.0	
P two-tail	1.0	
t Critical two-tail	2.0395	
Nether random_gen_validity_enforced_vanilla nor random_gen_uniform_random_vanilla is statistically better		

Table 4: Statistical Analysis performed on the Uniform Random Initialized Penalty Function and Validity Enforced Initalized Penalty Function, Randomly Generated Puzzle, EA configurations

	random_gen_uniform_random	random_gen_validity_enforced
mean	0.8971591167744307	0.9081384890909392
variance	0.0048536750468879545	0.002217915192875166
standard deviation	0.06966832168846868	0.047094746977504466
observations	30	30
df	29	29
F	2.1883952382308878	
F critical	0.5373999648406917	
Equal variances assumed		
observations	30	
df	58	
t Stat	-0.7031014079094314	
P two-tail	0.48480534283767107	
t Critical two-tail	2.0017	
Nether random_gen_validity_enforced nor random_gen_uniform_random is statistically better		

Table 5: Statistical Analysis performed on the Uniform Random Initialized Penalty Function and Validity Enforced Initalized Penalty Function, Provided Puzzle, EA configurations

	website_puzzle_uniform_random	website_puzzle_validity_enforced
mean	0.7231981981981981	0.7569819819819819
variance	0.0018428394610827035	0.0033889801964126286
standard deviation	0.042928306058854726	0.05821494822133426
observations	30	30
df	29	29
F	0.5437740424194343	
F critical	0.5373999648406917	
Equal variances assumed		
observations	30	
df	58	
t Stat	-2.515248538012428	
P two-tail	0.014687698277300898	
t Critical two-tail	2.0017	
website_puzzle_validity_enforced is statistically better than website_puzzle_uniform_random		

Table 6: Statistical Analysis performed on the Uniform Random Initialized Repair Function and Validity Enforced Initalized Repair Function, Randomly Generated Puzzle, EA configurations

	random_gen_uniform_random_bonus	random_gen_validity_enforced_bonus
mean	0.9790060265443229	0.9188060294943169
variance	0.0002560072950964829	0.060460786982912955
standard deviation	0.016000227970141015	0.24588775281195474
observations	30	30
df	29	29
F	0.004234269976817107	
F critical	0.5373999648406917	
Equal variances assumed		
observations	30	
df	58	
t Stat	1.315652066180613	
P two-tail	0.19346782059815182	
t Critical two-tail	2.0017	
Nether random_gen_validity_enforced_bonus nor random_gen_uniform_random_bonus is statistically better		

Table 7: Statistical Analysis performed on the Uniform Random Initialized Repair Function and Validity Enforced Initalized Repair Function, Provided Puzzle, EA configurations

	website_puzzle_uniform_random_bonus	website_puzzle_validity_enforced_bonus
mean	0.5038288288288288	0.3173423423423423
variance	0.2221989388036686	0.2015810912263615
standard deviation	0.47137982434939724	0.4489778293260832
observations	30	30
df	29	29
F	1.1022806625952564	
F critical	0.5373999648406917	
Unequal variances assumed		
observations	30	
df	31	
t Stat	1.5426809688332364	
P two-tail	0.12835995483574056	
t Critical two-tail	2.0395	
Nether website_puzzle_validity_enforced_bonus nor website_puzzle_uniform_random_bonus is statistically better		

Table 8: Statistical Analysis performed on the Valadity Inforced Initialized Penalty Function with Small Penalty and Medium Penalty, Provided Puzzle, EA configurations

	website_puzzle_validity_enforced_small_penalty	website_puzzle_validity_enforced
mean	0.9291666666666667	0.7569819819819819
variance	0.00010251805859913959	0.0033889801964126286
standard deviation	0.010125120177022077	0.05821494822133426
observations	30	30
df	29	29
F	0.030250415363199546	
F critical	0.5373999648406917	
Equal variances assumed		
observations	30	
df	58	
t Stat	15.69233619401264	
P two-tail	1.5429129447240864e-22	
t Critical two-tail	2.0017	
website_puzzle_validity_enforced_small_penalty is statistically better than website_puzzle_validity_enforced		

Table 9: Statistical Analysis performed on the Valadity Inforced Initialized Penalty Function with Medium Penalty and Large Penalty, Provided Puzzle, EA configurations

	website_puzzle_validity_enforced	website_puzzle_validity_enforced_large_penalty
mean	0.7569819819819819	0.4020270270270271
variance	0.0033889801964126286	0.024783144631117603
standard deviation	0.05821494822133426	0.15742663253438918
observations	30	30
df	29	29
F	0.13674536653260058	
F critical	0.5373999648406917	
Equal variances assumed		
observations	30	
df	58	
t Stat	11.388392849154927	
P two-tail	2.0169179415154988e-16	
t Critical two-tail	2.0017	
website_puzzle_validity_enforced is statistically better than website_puzzle_validity_enforced_large_penalty		

Table 10: Statistical Analysis performed on the Valadity Inforced Initialized Penalty Function with Small Penalty and Large Penalty, Provided Puzzle, EA configurations

	website_puzzle_validity_enforced_small_penalty	website_puzzle_validity_enforced_large_penalty
mean	0.9291666666666667	0.4020270270270271
variance	0.00010251805859913959	0.024783144631117603
standard deviation	0.010125120177022077	0.15742663253438918
observations	30	30
df	29	29
F	0.004136604136604133	
F critical	0.5373999648406917	
Equal variances assumed		
observations	30	
df	58	
t Stat	17.994926171087094	
P two-tail	2.0818133467335364e-25	
t Critical two-tail	2.0017	
website_puzzle_validity_enforced_small_penalty is statistically better than website_puzzle_validity_enforced_large_penalty		