Partial Functions in Fitness-Shared Genetic Programming

R. I. (Bob) McKay
School of Computer Science, Australian Defence Force Academy,
Northcott Drive, Campbell, ACT,
Canberra, Australia
rim@cs.adfa.edu.au

Abstract- This paper investigates the use of partial functions and fitness sharing in genetic programming. Fitness sharing is applied to populations of either partial or total functions and the results compared. Applications to two classes of problem are investigated: learning multiplexer definitions, and learning (recursive) list membership functions. In both cases, fitness sharing approaches outperform the use of raw fitness, by generating more accurate solutions with the same population parameters. On the list membership problem, variants using fitness sharing on populations of partial functions outperform variants using total functions, whereas populations of total functions give better performance on some variants of multiplexer problems.

1 Introduction

Genetic programming, like other forms of evolutionary computation, can suffer from premature convergence, where variation is eliminated from a population before the desired solution is achieved. Previous work (McKay 2000) demonstrated that fitness sharing, used in other forms of evolutionary computation to increase diversity and delay convergence, can lead to better performance - measured in terms of error rate - in genetic programming as well.

Fitness sharing was introduced by Deb and Goldberg (1989), in a form (explicit fitness sharing) which relied on a distance metric to define the similarity between individuals. Similar individuals are punished for that similarity by being required to share the raw fitness they receive. However there are disadvantages to this approach, because it requires a priori definition of the distance function, before knowledge of the shape of the search space has been acquired.

Smith, Forrest and Perlson (1992) noted that the requirement to define a distance metric could be avoided in problems where the fitness of an individual is built up from the payoff from discrete sub-problems. In this case, the payoff for a sub-problem could simply be shared amongst the individuals which perform well on that sub-problem. The resultant approach is known as implicit fitness sharing. Since a high proportion of genetic programming problems share this payoff structure, implicit fitness sharing is highly relevant.

Most work in evolutionary computation makes use of total functions. In many cases this is a matter of simple necessity - for example, the genome structure in standard genetic algorithms does not support partially defined functions ('don't care' symbols simply indicate that a particular input value is ignored, not that the output value is undefined for a given input). But partial functions are rarely, if ever, used in genetic programming, where they do have a sensible meaning.

On the other hand, total functions are under evolutionary pressure to solve all parts of a problem; this pressure tends to reduce diversity, perhaps fatally so if a critical part of an optimal solution is more difficult to find than equivalent parts of a local optimum.

The aim of the work begun here is to investigate the tradeoffs between the greater implicit parallelism of total functions, and the potentially greater diversity provided by partial functions, and their interaction with the diversity-promoting mechanism of implicit fitness sharing.

2 Details of Approach

Comparisons have been carried out on two classes of problem: learning (recursive) list membership in a lisp-like language, and learning boolean descriptions for 6- and 11-multiplexers. These problems are described in more detail below. The experiments compare the use of implicit fitness sharing, raw fitness, and a combination of the two. Each treatment is repeated twice, using a population of total functions, and a population of partial functions.

2.1 Partial Functions

A partial function is a function whose value is not defined for some argument values. In these experiments, partial functions are represented by the use of a special symbol, 'undef', which may occur at any point in the program tree. When the program runs, any evaluation for which 'undef' is an argument evaluates to 'undef' unless the value of the function is independent of that argument. For example, boolean 'and' has the truth table shown in table 1:

and	**false**	**undef**	**true**
false	false	false	false

undef	false	undef	undef
true	false	undef	true

Table 1: Truth Table for Boolean and

The system is based on Ross' (1999) DCTG-GP system. DCTG-GP was used because its explicit representation of the syntax and semantics of the program populations provided a simple mechanism to specify the syntax and semantics of the 'undef' symbol. However the grammars used merely encode the typing of the problem space, and so the results apply not only to grammar-guided genetic programming (Whigham 1995), but extend to strongly typed genetic programming (Montana 1995).

2.2 Implicit Fitness Sharing and Partial Functions

Fitness sharing aims to reduce the eagerness of evolutionary search and encourage diversity by providing a reward for diversity. Thus fitness sharing algorithms typically reduce the early performance of an algorithm, but more than compensate by delaying convergence, and producing better asymptotic behaviour.

For populations of total functions, the implementation of implicit fitness sharing is straightforward. The payoff for a particular sub-problem is shared amongst all population members which correctly answer the sub-problem (in the problems used here, values are discrete, so grades of correctness do not need to be considered). Mathematically, we replace the raw fitness function for an individual i

$$f_{raw}(i) = \sum_{c \in cases} reward\left(i(c)\right)$$

with the shared fitness function

$$f_{share}(i) = \sum_{c \in cases} \frac{reward\left(i(c)\right)}{\sum_{i':i'(c)=i(c)} reward\left(i'(c)\right)}$$

With populations of partial functions, another issue arises: if the payoffs from the sub-problems are simply added together (assuming there are no negative payoffs), then there is evolutionary pressure toward totality, because even small rewards for poor predictions are better than no reward at all. This pressure would defeat the intention, which is to permit increased diversity through partial functions which are free to concentrate on particular sub-problems. Hence in this work, the shared fitnesses are divided by the number of sub-problems which the program attempts to solve (that is, the fitness of an individual program is the mean of the shared rewards it receives, averaged over all the sub-problems for which its answer is not 'undef'). For an individual i, let N(i) be the number of cases c for which i(c) is defined, then:

$$f_{part\_share}(i) = \sum_{c \in cases} \frac{reward\left(i(c)\right)}{N(i) * \sum_{i':i'(c)=i(c)} reward\left(i'(c)\right)}$$

Thus the evolutionary pressure on partial functions is toward accuracy on sub-problems. This pressure will not necessarily result in individuals capable of solving the whole problem. There are many possible mechanisms to alleviate this, ensemble learning mechanisms being particularly notable. However a very simple approach has been taken in this paper. In addition to runs using raw fitness throughout, and others using shared fitness throughout, there is a third set of runs, in which the fitness measure changes from shared fitness to raw fitness using a fixed schedule. The particular schedule was chosen a priori. It uses fitness sharing for the first 25% of generations of a run, and raw fitness for the last 25%, with a linear ramp between fitness sharing and raw fitness for the remaining 50% of generations (the raw and shared fitnesses are normalised to have the same mean before being added together)

This schedule is doubtless not optimal - the optimal schedule is almost certainly problem-dependent. But most obvious mechanisms to optimise the schedule lead to problems of fairness of comparison with the other two treatments.

3 Experiments: List Membership

The search space for this experiment is defined by the grammar in table 2 (for total functions, the productions leading to 'undef' are deleted):

```
S -> M
M -> if EXPN EXPN M
M -> ''
M -> undef
EXPN -> atom LST
EXPN -> eq LST LST
EXPN -> member x LST
EXPN -> true
EXPN -> false
EXPN -> undef
LST -> first LST
LST -> rest LST
LST -> x
LST -> y
LST -> undef
```

Table 2: Grammar for List Membership

The recursive call to member allows the possibility of infinite loops. To prevent this, a count of the depth of looping was kept, and a depth greater than 20 caused the function to return the value 'loop'; this was treated in fitness evaluation as an incorrect (but defined) answer.

The examples for learning this function consisted of ten true cases and ten false, and are shown in table 3:

TRUE CASES	FALSE CASES
member(1 [1])	member(1 [6])
member(1 [2 1])	member(1 [3 6])
member(1 [2 3 1])	member(1 [2 3 6])
member(1 [2 3 4 1])	member(1 [2 3 4 6])
member(1 [2 3 4 5 1])	member(1 [2 3 4 5 6])
member(1 [2 3 4 5 6 1])	member(1 [2 3 4 5 6 7])
member(1 [2 3 4 5 6 7 1])	member(1 [2 3 4 5 6 7 8])
member(1 [2 3 4 5 6 7 8 1])	member(1 [2 3 4 5 6 7 8 9])
member(1 [2 3 4 5 6 7 8 9 1])	member(1 [2 3 4 5 6 7 8 9 2])
member(1 [2 3 4 5 6 7 8 9 2 1])	member(1 [2 3 4 5 6 7 8 9 2 3])

Table 3: List Membership Cases

The aim of the experiment was to find a program which correctly computes membership. An example solution is:

```
(if (eq x (first y))
    true
    (if (member x (rest y))
        true
        false))
```

The experimental setup used tournament selection and half-ramped initialisation as implemented in DCTG-GP (Ross, 1999); table 4 shows experimental parameters:

PARAMETER	SPECIFICATION
Number of Runs	100
Max Generations	200
Population Size	1000
Max depth (initial pop)	8
Max depth (subsequent)	10
Tournament size	5
Crossover Probability	0.9
Mutation Probability	0.1

Table 4: Run Parameters (List Membership)

The raw fitness function was the proportion of the 20 cases correctly solved. In principle, it would be possible for a non-recursive program to solve all 20 cases, but not within the maximum depth imposed on the population. Each run terminated at 200 generations, earlier if it found a correct solution to the problem.

4 Results: List Membership

The percentage of runs which terminated in a correct solution are shown in table 5:

	Total Funcs	Partial Funcs
Raw Fitness	63	43
Ramped Fitness	79	84
Shared Fitness	79	88

Table 5: Percentage of 100 Runs Generating Correct Solution

The differences between the raw fitness and other treatments are significant at the 1% level (maximum likelihood ratio test) for both the total and partial function cases. The other differences are of lower significance, but the difference between the total and partial functions, when the ramped and shared fitness cases are aggregated together, gives a probability value of 6.5% for the null hypothesis. For confirmation, trials involving fitness sharing were replicated a further 200 runs. Results are shown in table 6:

	Total Funcs	Partial Funcs
Ramped Fitness	79.3	84.7
Shared Fitness	82.7	88

Table 6: Percentage of 300 Runs Generating Correct Solution

With a total of 300 runs, the null hypothesis for the difference between partial and total functions has a probability value of 1.2%, when ramped and shared fitness are considered together. Considered separately, the null hypothesis probabilities are 8.8% and 6.4% respectively. The differences between ramped and shared fitness are of low significance, even when total and partial results are considered together.

Figure 1 shows the percentage of runs incomplete plotted against generation. In the legend, 'total' and 'partial' refer to populations of partial and total functions respectively, and 'raw', 'share' and 'ramp' refer to the use of raw fitness throughout a run, fitness sharing throughout a run, and the ramped approach described above (to improve readability, the partial functions/raw fitness treatment is omitted in all figures since it is of little independent interest).

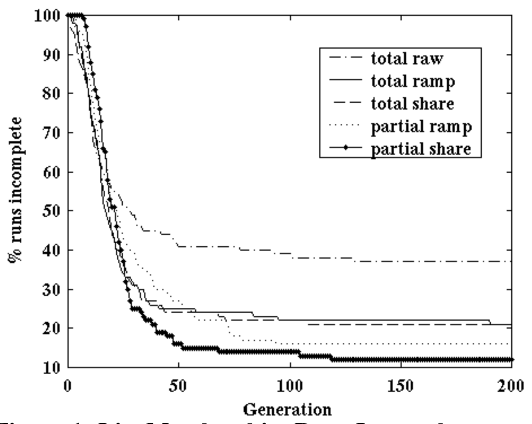

Figure 1: List Membership, Runs Incomplete

Figure 2 plots cases not covered by fittest individual against generation (first 20 generations omitted for clarity)

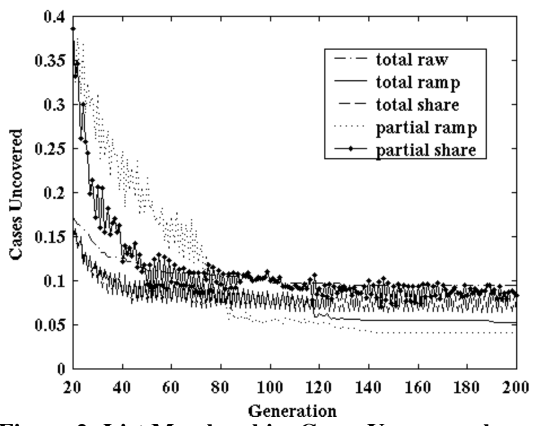

Figure 2: List Membership, Cases Un-covered, Fittest Individual

One interesting aspect of figure 2 is the 2-generation oscillation in coverage rate exhibited by fitness-sharing treatments. On detailed examination, the populations in many runs contain a group of individuals with low error rate (and hence high raw fitness), and another group of very different individuals with higher error rate, but covering a different subset of the test cases. Small fluctuations in size of the second set are reflected in larger changes in the shared fitness of its members relative to the first set, and hence to the selection pressure toward it, resulting in the oscillations seen. These fluctuations can also be seen in figure 3, showing the standard deviation of the number of individuals in the population covering each test case, a proxy measure of the phenotypic diversity of the population (if the population is of low diversity, then some test cases will be well covered, and others poorly, so the variance in cover will be high).

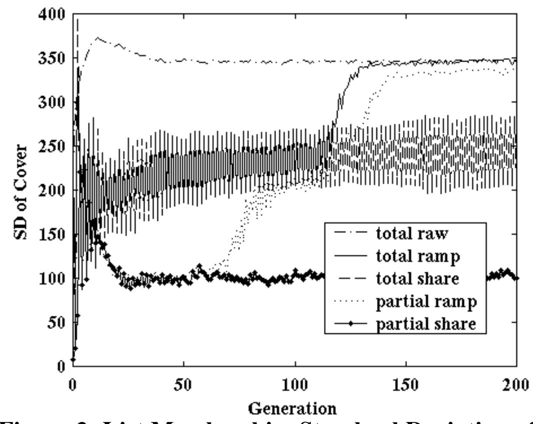

Figure 3: List Membership, Standard Deviation of Cover of Test Cases

Figure 4 shows the number of cases which the best individual in the population attempts to predict (ie gives an output other than 'undef'). The effect of ramped fitness sharing is clearly visible.

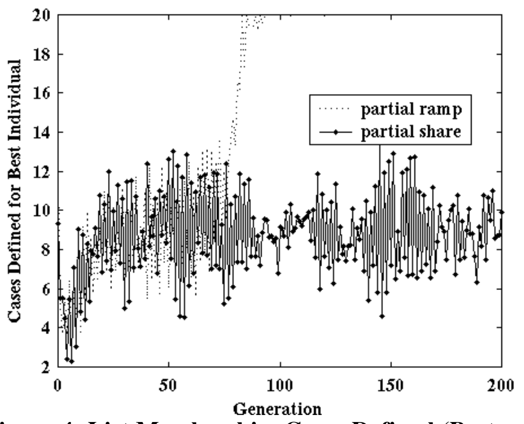

Figure 4: List Membership, Cases Defined (Best Individual)

5 Experiments: Multiplexers

Two sets of experiments were conducted, using the 6-multiplexer and 11-multiplexer problems. The former seeks a boolean expression for a multiplexer with two address and four data lines, the latter with three address and eight data lines. The search space for the 6-multiplexer is defined by the grammar in table 7 (for total functions, the productions leading to 'undef' are deleted).

```
EXPR → BOOL
BOOL → TERM
BOOL → and BOOL BOOL
BOOL → or BOOL BOOL
BOOL → not BOOL
BOOL → if BOOL BOOL BOOL
BOOL → undef
TERM → a0
TERM → a1
TERM → d0
TERM → d1
TERM → d2
TERM → d3
TERM → undef
```

Table 7: Grammar for 6-Multiplexer

The search space for the 11 multiplexer extends this by adding TERM productions for address line a2 and data lines d4 through d7.

The examples for learning the 6 multiplexer consisted of the 64 possible input/output pairs - see table 8. For the 11 multiplexer, computational cost precluded evaluation over the 2048 input/output pairs in each generation. Instead, for each generation, 64 of these pairs were randomly selected and used to evaluate that generation. Since termination was based on a zero error rate for these

64 cases, it is possible that some incorrect solutions were accepted as correct. However this possibility does not affect the comparisons undertaken in this work, since all treatments are affected equally.

Inputs						Output
a0	a1	d0	d1	d2	d3	
0	0	0	0	0	0	0
0	1	0	1	0	0	1
0	1	1	1	0	0	1
1	1	1	1	1	1	1

Table 8: 6 Multiplexer Input/Output Examples

The aim of the experiments was to find a boolean function which correctly defines the multiplexer. An example solution for the 6-multiplexer is:

```
(if a0 (if a1 d3 d2)
       (if a1 d1 d0))
```

The experimental setup used tournament selection and half-ramped initialisation; experimental parameters for the 6 multiplexer runs are given in table 9:

PARAMETER	SPECIFICATION
Number of Runs	100
Max Generations	100 (experiment 1)
	500 (experiment 2)
	1000 (experiment 3)
Population Size	500 (experiments 1,2)
	100 (experiment 3)
Max depth (initial pop)	8
Max depth (subsequent)	8
Tournament size	5
Crossover Probability	0.9
Mutation Probability	0.1

Table 9: Run Parameters (6 Multiplexer)

Each run of the multiplexers was terminated at the nominated generation, or earlier if it found a correct solution to the problem. As with the membership problem, three forms of fitness evaluation were used: raw fitness, implicit fitness sharing, and the ramped approach. Each form was evaluated on populations both of total functions and of partial functions.

PARAMETER	SPECIFICATION
Number of Runs	100
Max Generations	400 (experiment 4)
	1000 (experiment 5)
Population Size	500
Max depth (initial pop)	8
Max depth (subsequent)	10
Tournament size	5
Crossover Probability	0.9
Mutation Probability	0.1

Table 10: Run Parameters (11 Multiplexer)

6 Results: Multiplexers

The percentage of runs which terminated in a correct solution for experiment 1 are shown in tables 11:

	Total Funcs	Partial Funcs
Raw Fitness	84	75
Ramped Fitness	100	88
Shared Fitness	100	82

Table 11: % of Runs Generating Correct Solution (6 Multiplexer, experiment 1)

For the 6 multiplexer, the differences between total and partial functions are of low significance for the raw fitness treatment, but the null hypothesis has a probability of under .001% for both ramped and shared fitness. Similarly, the differences between raw fitness and ramped and shared fitness are highly significant for total functions. For partial functions, the ramped and raw fitness treatments are significantly different (p = 1.2%), but the other differences are of low significance. Thus it appeared that for the 6 multiplexer problem, delayed convergence in the partial functions run did not convey a benefit. The percentage of runs generating a complete solution, and the coverage rate of the best individual, are plotted in figures 5 and 6.

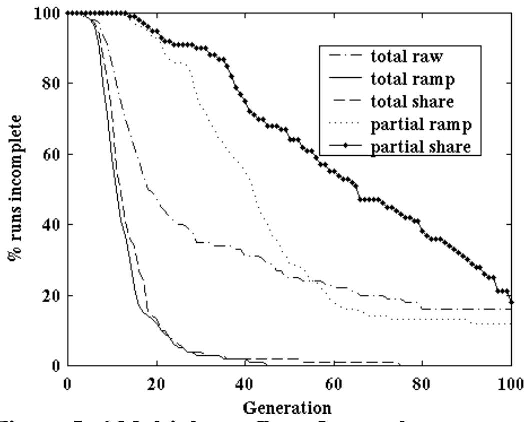

Figure 5: 6 Multiplexer, Runs Incomplete (experiment 1)

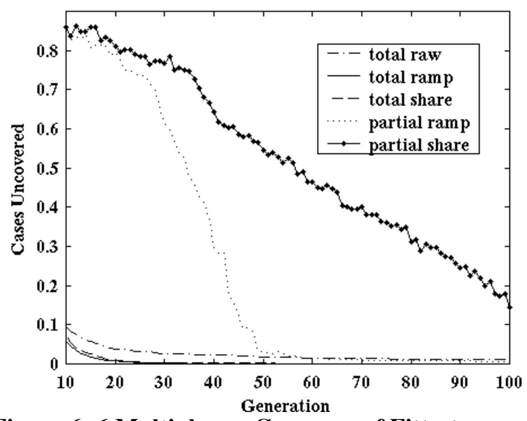

Figure 6: 6 Multiplexer, Coverage of Fittest Individual (experiment 1)

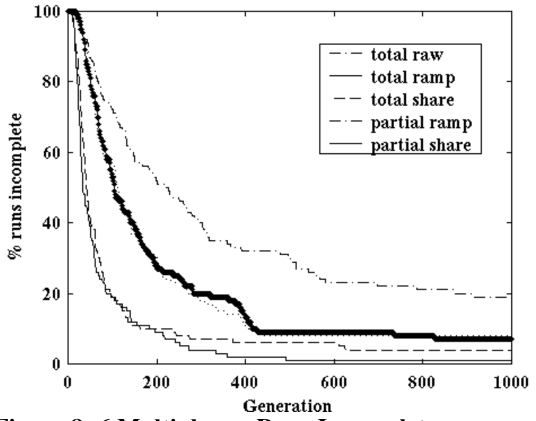

Figure 8: 6 Multiplexer, Runs Incomplete (experiment 3)

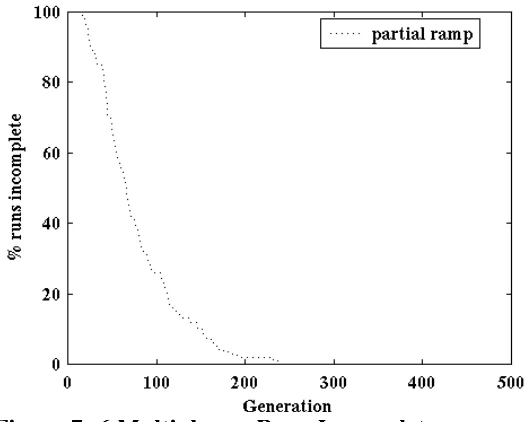

Figure 7: 6 Multiplexer, Runs Incomplete (experiment 2)

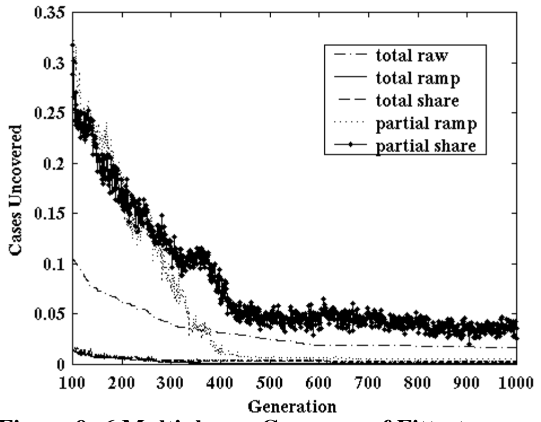

Figure 9: 6 Multiplexer, Coverage of Fittest Individual (experiment 3)

From the results of experiment 1, it appeared possible that the ramped fitness/partial functions treatment had not yet reached convergence, so the experiment was repeated for that treatment, but allowed to continue to generation 500. In that experiment, all runs eventually succeeded, so it became apparent that the 6-multiplexer problem with a population of 500 was insufficiently difficult to differentiate the convergent behaviour of partial and total function population behaviours when combined with the ramped approach to fitness. This is shown in figure 7.

One way to make the problem more difficult is simply to reduce the size of the populations. A further set of experiments was conducted, using a population of only 100 individuals, but extending to generation 1000.

	Total Funcs	Partial Funcs
Raw Fitness	81	33
Ramped Fitness	99	92
Shared Fitness	97	93

Table 12: Percentage of Correct Solution (6 Multiplexer, experiment 3)

The fitness sharing runs are significantly different from the raw fitness runs (probability level infinitesimal), the difference between fitness sharing and the ramped treatment is not significant, and the differences between total and partial functions are significant for ramped fitness (p=.011), but not for shared fitness (p=.18). From the plots, it appears likely that the runs are fully converged, and hence that the differences in converged behaviour are real.

Plots of the standard deviation of case coverage (figure 10) and of the degree of totality of the two partial function treatments (figure 11) are included to permit comparison with the list membership experiments. The effect of ramped fitness sharing is again clearly visible in both plots.

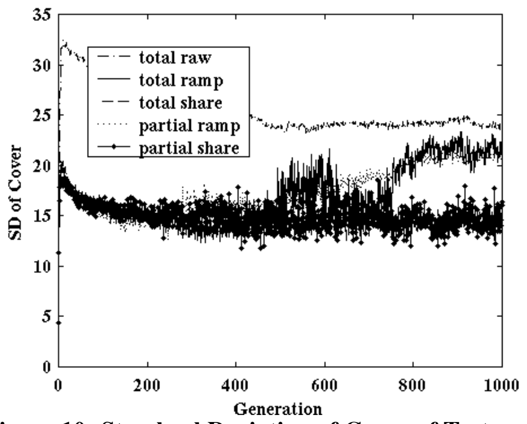

Figure 10: Standard Deviation of Cover of Test Cases (6 Multiplexer, Experiment 3)

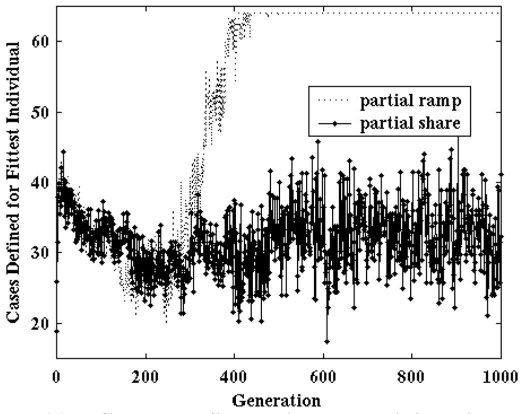

Figure 11: Cases Defined (Best Individual) - 6 Multiplexer, Experiment 3

	Total Funcs	Partial Funcs
Raw Fitness	50	16
Ramped Fitness	99	85
Shared Fitness	99	3

Table 13: Percentage of Correct Solutions (11 Multiplexer, Experiment 4)

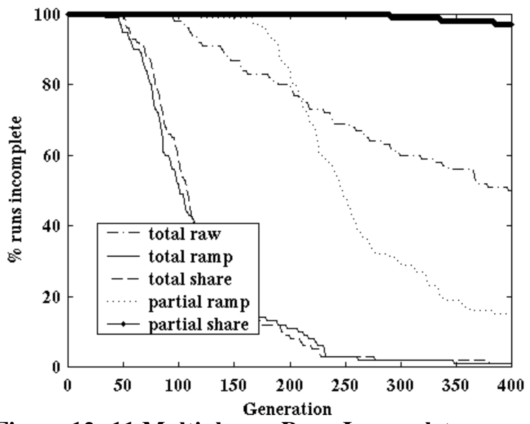

Figure 12: 11 Multiplexer, Runs Incomplete (experiment 4)

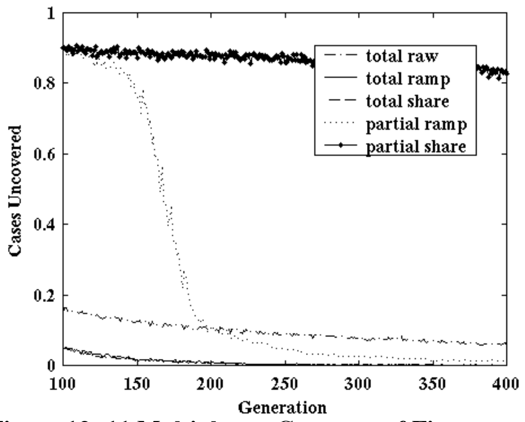

Figure 13: 11 Multiplexer, Coverage of Fittest Individual (experiment 4)

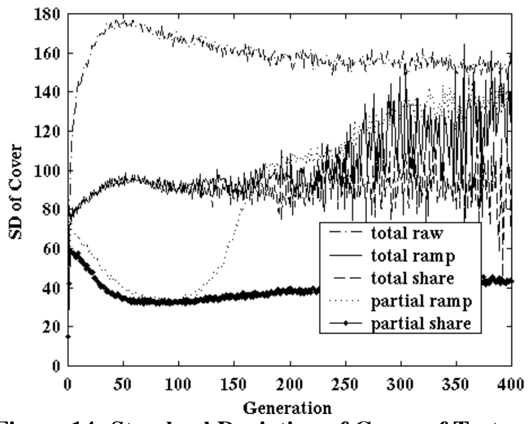

Figure 14: Standard Deviation of Cover of Test Cases (11 Multiplexer, Experiment 4)

For the 11 multiplexer, fitness sharing methods again outperform raw fitness. The inability of the partial functions to find a total solution without the evolutionary pressure of ramped fitness is clear. The difference between the partial and total function populations with ramped fitness is highly significant ($p < .0001$). Figure 14 demonstrates again the ability of fitness sharing to maintain population diversity, and the higher diversity achieved by partial functions, and figure 15 shows the effect on coverage by the partial function populations of the ramped fitness approach.

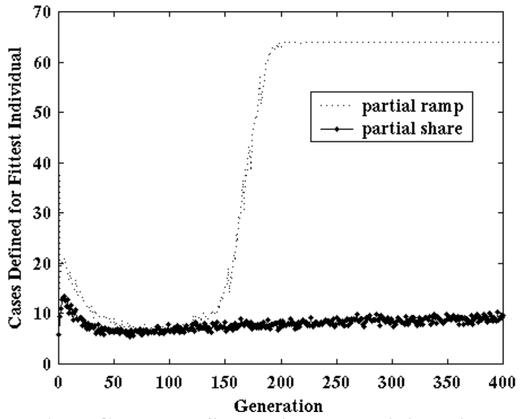

Figure 15: Cases Defined (Best Individual) - 11 Multiplexer, Experiment 4

However figures 12 and 13 again suggest that the partial functions / ramped fitness treatment may not be fully converged. A further run of this case was performed to 1000 generations. This again permitted the partial functions / ramped fitness treatment to reach full convergence, and reliably attain a complete solution, so that again, the difference between the partial and total function searches are one of eagerness and rate of convergence, rather than of converged behaviour.

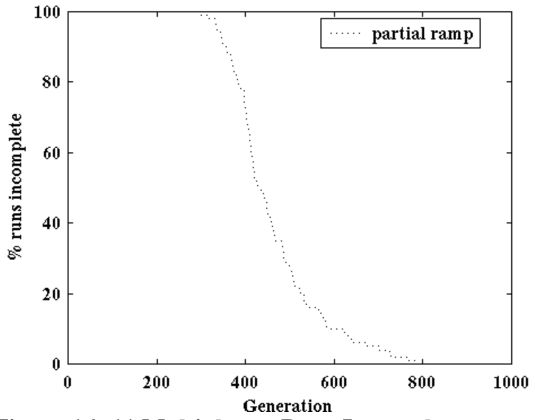

Figure 16: 11 Multiplexer, Runs Incomplete (experiment 5)

7 Discussion and Further Work

The results presented here confirm previous results, that fitness sharing approaches to population diversity carry clear benefits, and result in significantly improved system performance. The comparison between partial and total populations is more equivocal - in one problem (list membership) partial function populations had better converged behaviour, in another (6 multiplexer) total function populations behaved better when the population size was decreased to an extreme level. In the third (11 multiplexer), they could not be distinguished.

Total functions give a more eager search and hence earlier convergence, so the balance between computational cost and converged behaviour might seem to favour populations of total rather than partial functions. However the greater simplicity of partial functions could result in lower evaluation costs, restoring the balance; this requires further exploration, but preliminary results on the three problems detailed above suggest that cpu-time to convergence with ramped fitness does not differ greatly between the total and partial function approaches, even though the total function treatments converge earlier in number of generations.

The experiments reported here use toy problems, and there are important issues of applicability to real-World learning problems. However toy problems do have the virtue of permitting detailed experimentation at acceptable computation cost which would not be feasible with larger problems. Work is currently in progress to apply the techniques described here to large-scale ecological datasets.

The population evaluation mechanism used here - error of the fittest individual - is poortly suited to populations of partial functions. Work is under way using ensemble measures for evaluation.

One possible advantage of partial functions derives from their smaller average size. This may reduce the potential for overfitting, and hence support better generalisation. Experimental investigation of this aspect is planned.

8 Conclusions

In the experiments reported here, covering two very different types of problems, there is strong evidence of the ability of fitness sharing to maintain population diversity and delay convergence. This ability is enhanced when fitness sharing is applied to populations of partial functions rather than populations of total functions.

The delayed convergence led to considerably better performance by approaches based on fitness sharing (when compared with raw fitness) for all three problems considered. This improvement occurred whether performance was measured by percentage of runs finding correct solutions, or by error rate at convergence.

The increased delay in convergence provided by populations of partial functions led to significantly better performance over 200 generations in the recursive list membership problem, under both measures.

For the 6 multiplexer problem, the performance of fitness sharing on populations of total functions was so good, with population sizes normally used in genetic programming that further delaying convergence through the use of partial functions could only serve to delay convergence. When the population size was reduced to an extreme level, total function populations significantly out-performed partial function populations.

In the 11 multiplexer problem, with typical population sizes of 500 individuals, the converged behaviour of partial functions under ramped fitness sharing, and total functions under either pure or ramped fitness sharing, could not be distinguished.

Acknowledgements

The ideas in this paper have benefited greatly from discussions over the years with Paul Darwen, Peter Whigham, Xin Yao and Ko-Hsin Liang. I would also like to thank Ted Catchpole for assistance with statistical analysis of the experiments, and the anonymous reviewers for the improvements to the exposition resulting from their comments. The system has been implemented through modifications to Brian Ross' innovative DCTG-GP system, which contributed greatly to the rapidity with which a wide range of different algorithms could be prototyped in the development of this system, and supported a simple mechanism for defining and evaluating partial functions.

Bibliography

Deb, K and Goldberg, D E: 'An investigation of niche and species formation in genetic function optimization' in J D Schaffer (Ed) *Proceedings of the Third International Conference on Genetic Algorithms*, Pp 42-50, Morgan Kaufmann, 1989

McKay, R I: 'Fitness Sharing in Genetic Programming', Genetic and Evolutionary Computation Conference (GECCO-2000), to appear

Montana, D J: 'Strongly Typed Genetic Programming', *Evolutionary Computation* 3(2), Pp. 199-230, 1995

Ross, B J: 'Logic-based Genetic Programming with Definite Clause Translation Grammars', Technical Report #CS-99-02, Dept of Computer Science, Brock University, St Catharines Ontario, 1999; summary in Banzhaf et al (eds) *Proceedings of the Genetic and Evolutionary Computation Conference*, P1236, Morgan Kaufmann, 1999

Smith, R E, Forrest, S and Perelson, A S: 'Searching for diverse, cooperative populations with genetic algorithms', *Evolutionary Computation* 1(2), Pp 127 - 149, 1992

Whigham, P A: 'Grammatically-biased Genetic Programming' in J Rosca (ed) *Proceedings of the Workshop on Genetic Programming: From Theory to Real-World Applications*, Pp 33-41, Morgan Kaufmann, 1995