A Practical Schema Theorem for Genetic Algorithm Design and Tuning

David E. Goldberg and Kumara Sastry
Illinois Genetic Algorithms Laboratory (IlliGAL)
Department of General Engineering
University of Illinois at Urbana-Champaign
104 S. Mathews Ave, Urbana, IL 61801
{deg,ksastry}@uiuc.edu

Abstract

This paper develops the theory that can enable the design of genetic algorithms and choose the parameters such that the proportion of the best building blocks grow. A practical schema theorem has been used for this purpose and its ramification for the choice of selection operator and parameterization of the algorithm is explored. In particular stochastic universal selection, tournament selection, and truncation selection schemes are employed to verify the results. Results agree with the schema theorem and indicate that it must be obeyed in order to ascertain sustained growth of good building blocks. The analysis suggests that schema theorem alone is insufficient to guarantee the success of a selectorecombinative genetic algorithm.

1 Introduction

The importance of building blocks (BBs) and their role in the workings of GAs have long been recognized (Holland, 1975; Goldberg, 1989). Furthermore, the following six conditions for a GA success have been proposed (Goldberg, Deb, & Clark, 1992): (1) Know what GAs are processing - building blocks (BBs), (2) ensure an adequate initial supply of raw BBs, (3) ensure growth of superior BBs, (4) ensure the mixing of BBs, (5) ensure good decisions among competing BBs, and (6) solve problems with bounded BB difficulty. One of the important conditions is to make sure that the GA is well supplied with a sufficient supply of the BBs required to solve a given problem. It is also equally important that the proportion of the good ones in the population grow. The first task is addressed elsewhere in this proceedings (Goldberg, Sastry, & Latoza, 2001), and the second task, that is, the issue of guaranteeing the increase in market share of good BBs in a population is addressed in the current study.

The usual approach in achieving this is the schema theorem (Holland, 1975; De Jong, 1975). Therefore, the objective of this study is to utilize a practical schema theorem to explore the ramifications of the schema theorem for the choice of selection operator and parameterization of the algorithm. In this study we consider three selection schemes: stochastic universal selection (SUS) (Baker, 1987; Grefenstette & Baker, 1989), s-wise tournament selection (Goldberg, Korb, & Deb, 1989), and truncation selection (Muhlenbein & Schlierkamp-Voosen, 1993). SUS is a proportionate scheme and s-wise tournament selection and truncation selection are ordinal schemes. The performance of each of these selection schemes in both early as well as late in the GA run is analyzed based on the schema theorem.

We start by presenting a brief note on the schema theorem, both its original version, and a generalized version. A simplified version of the generalized schema theorem is then used for the BB growth design. Three different selection schemes are considered in the light of the BB growth design and are analyzed for parameter settings to ensure the growth of best BBs in the population.

2 Generalized Schema Theorem

There have been many studies on schema theorem, and a complete literature review on schema theorem is beyond the scope of this study. Instead, we present a brief overview of schema theorem and refer the reader elsewhere (Goldberg, 1989) for a detailed exposition of the same. Under proportionate selection, single-point crossover, and no mutation, the schema theorem may be written as follows:

$$\langle m(H,t+1)\rangle \geq m(H,t)\frac{f(H,t)}{\overline{f}(t)}\left[1-p_c\frac{\delta(H)}{\ell-1}\right], \quad (1)$$

where $\langle m\,(H,t+1)\rangle$ is the expected number of individuals that represent the schema H at generation $t+1$, $m(H,t)$ is the number of individuals that represent the schema H at generation t, $f(H,t)$ is the average fitness value of the individuals containing schema H at generation t, $\overline{f}(t)$ is the average fitness of the population at generation t, p_c is the crossover probability, $\delta(H)$ is the defining length defined as the distance between the outermost fixed positions of a schema, and ℓ is the string length.

Inspection of the schema theorem and an analysis of proportionate selection and single-point crossover (Goldberg, 1989), indicates that the the term $m\,(H,t)\,\frac{f(H,t)}{\overline{f}(t)}$ accounts for the selection, and the term $\left[1-p_c\frac{\delta(H)}{\ell-1}\right]$ accounts for crossover operation. It should be noted that the term representing the selection operator is exact and the inequality occurs due to the crossover operation. Some factors like crossover between identical individuals (self-crossover) are neglected. The schema theorem tells us that the proportion of schemata increases when they have above average fitness and relatively low crossover disruption.

However, the schema theorem as given by equation 1 is restricted to proportionate selection and one-point crossover. This concern can be eliminated by identifying the characteristic form of schema theorem and substituting appropriate terms for other selection schemes and genetic operators. Recognizing that a selection scheme might allocate the numbers of schemata in a different manner, and a genetic operator might yield a different survival probability when compared to proportionate selection and single point crossover, the following *generalized* schema theorem (Goldberg & Deb, 1991) can be written

$$\langle m(H,t+1)\rangle \geq m(H,t)\gamma(H,m_i,f_i,t), \quad (2)$$

where,

$$\gamma(H,m_i,f_i,t) = \phi(H,m_i,f_i,t)P_s(H,m_i,f_i,t), \quad (3)$$

and ϕ is the selection factor, and is a function of the fitness function f_i, the distribution of structures in the population m_i, at generation t. The value of ϕ for a schema H is calculated by adding the contributions of all the individual strings that are members of the schema H. P_s is a survival probability. The generalized schema theorem can alternatively be written in proportion form by dividing throughout by population size N as,

$$\langle P(H,t+1)\rangle \geq P(H,t)\gamma(H,m_i,f_i,t). \quad (4)$$

This theorem states that desirable schemata grow if $\gamma\,(H,m_i,f_i,t) \geq 1$. Although both the selection factor ϕ and the survival probability P_s are functions of the fitness function and the population, both quantities can be characterized more simply and are explained in the following section.

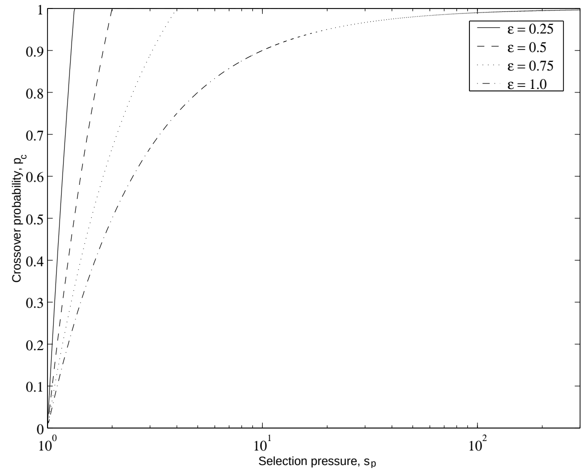

Figure 1: Limiting crossover probability p_c as a function of selection pressure s_p for different values of operator loss ϵ.

3 Designing for BB Growth

To employ the schema theorem in design, we simplify it by replacing ϕ with the selection pressure s_p, and parameterize the survival probability on an operator loss ϵ and the crossover probability p_c. The schema theorem can now be written as

$$\langle m\,(H,t+1)\rangle \geq m\,(H,t)\,s_p\,[1-p_c\epsilon]\,. \quad (5)$$

Then the desirable schema's grow if

$$s_p\,[1-p_c\epsilon] \geq 1. \quad (6)$$

Rearranging in terms of crossover probability p_c gives

$$P_c \leq \frac{1-s_p^{-1}}{\epsilon}. \quad (7)$$

The limiting p_c value is plotted as a function of selection pressure at different crossing losses in figure 1. We can see that even in the case of total loss of schema integrity, BB market share growth can still be ensured with reasonable combinations of sufficient selection pressure and diminished crossover probability. An interesting factor to note is that the schema theorem can always be satisfied with zero exchange — $p_c = 0$. However, in such a case the whole basis of operation principle of the GA with crossover is violated. This suggests that schema theorem must be obeyed, but that does not guarantee even a modicum of building block exchange, which is very important for a successful GA design.

To apply the BB design developed in this section in a real GA requires the consideration of selection pressure exerted by a given selection procedure. This issue is addressed in the next section for three different selection schemes.

4 Selection Schemes and Selection Pressure

We estimate the selection pressure s_p exerted in two phases, one early in the run and the other late in the run. The reason for doing so can be justified as follows: The initial generations are critical to the success of a GA run, because unless a schema (or its components) grow at the outset, its chances for success later on are quite poor. However, even if the conditions early in a GA push the best BBs fairly aggressively, but as the evolution wears on, even schemes with fairly steady drive toward convergence loose some of their initial punch. This might be a deal breaker, because the loss of selection pressure combined with high schema loss and fixed crossover probability might cause the evolution to stall before the best BBs dominate the population.

Here we consider the selection pressure exerted by three selection schemes, s-wise tournament, truncation, and proportionate selection procedures.

4.1 Tournament Selection

In s-wise tournament selection, s individuals are randomly drawn from a population (with or without replacement) and the best individual is selected. Assume that the initial proportion of superior BBs at time $t = 0$ is P_0 and is very small. The proportion of the best individuals under selection alone can be written as (Goldberg & Deb, 1991),

$$P_{t+1} = 1 - (1 - P_t)^s \,. \tag{8}$$

To consider the early performance of tournament selection, we calculate the selection ratio $\phi_t = P_{t+1}/P_t$ when $P_t \approx 0$.

$$\phi_t = P_t^{-1} \left[1 - (1 - P_t)^s\right]. \tag{9}$$

Recognizing that P_0 is small, and that $(1-x)^n \approx 1-nx$ for small x, we obtain

$$\phi_t \approx \frac{1 - (1 - sP_t)}{P_t} = s. \tag{10}$$

Thus early in the run when a good schema exists only in small proportions, the best schemas increase by a factor of s, the tournament size. This result is justified, since each individual participates in an average of s tournaments and the best individuals win all s of them. Equation 7 can be written for the case of tournament selection early in the run as

$$p_c \leq \frac{1 - s^{-1}}{\epsilon}. \tag{11}$$

As the proportion of a good schema becomes significant, the effective loss due to crossover reduces. This occurs due to the fact that for many crossover operators, a schema crossed with itself yields an instance of the same schema (Holland, 1975; Schaffer, 1987). The success probability, P_s, incorporating self-crosses can be written as

$$P_s = 1 - p_c' \left(1 - P_t\right), \tag{12}$$

where $p_c' = p_c\epsilon$, and the term $1 - P_t$ is the result of self crossing — a schema contained in a string crossed with another string containing the schema is not disrupted regardless of the crossover point. The late behavior of the tournament selection is then given by

$$P_{t+1} = \left[1 - (1 - P_t)^s\right]\left[1 - p_c' \left(1 - P_t\right)\right]. \tag{13}$$

To perform a late analysis, we assume that the better structures have taken over more than $(1/s)^{\text{th}}$ of the population. Then the selection term $\left[1 - (1 - P_t)^s\right]$ approaches 1 and the late performance is described by

$$P_{t+1} = \left[1 - p_c' \left(1 - P_t\right)\right]. \tag{14}$$

The selection ratio, ϕ_t is given by

$$\phi_t = P_t^{-1} \left[1 - p_c' \left(1 - P_t\right)\right]. \tag{15}$$

It can be easily seen that $\phi_t \geq 1$ if (1) $0 < P_t \leq 1$, (2) $0 < p_c' \leq 1$, and (3) $s\left[1 - p_c'\right] \geq 1$, based on which the selection pressure and crossover probability are chosen for initial growth. Thus, by accounting for self-crossing we have shown that the subsequent effective slowing of convergence rate due to selection is more than overcome by the effective reduction in crossover probability.

4.2 Truncation Selection

Truncation selection is an ordinal selection scheme in which the top $1/s$ individuals in a population are given s copies each. If the proportion P_t of best individuals in a population is $1/s$ (early in the run), then the growth is geometric,

$$P_{t+1} = sP_t. \tag{16}$$

In other words, the selection rate $\phi_t = s$, which is the same as that for tournament selection. Once the proportion of good structures reaches or exceeds $1/s$ $\left(P_t \geq s^{-1}\right)$, the scheme saturates and the final proportion is,

$$P_{t+1} = 1. \tag{17}$$

Therefore the late performance is influenced only by the crossover operation and the proportion of good structures can be written as

$$P_{t+1} = \left[1 - p'_c\,(1 - P_t)\right]. \tag{18}$$

Thus truncation and tournament selection procedures have essentially similar late performance.

4.3 Proportionate Selection

To understand the early performance of proportionate selection, consider a highly simplified model. Consider two possible alternatives, 1 and 2, represented by objective function values f_1 and f_2 respectively with $f_2 > f_1$. Here the assumption is that alternative 1 is the average individual and alternative 2 is the best individual in the initial population. The proportion of alternative 2 can be tracked with the following difference equation,

$$P_{t+1} = \frac{f_2}{\overline{f}_t} P_t, \tag{19}$$

where $\overline{f}_t$ is the average function value at generation t, and is given by $\overline{f}_t = f_1\,(1 - P_t) + f_2 P_t$. Substituting this value in the above equation we get,

$$P_{t+1} = \frac{s}{(s-1)P_t + 1} P_t, \tag{20}$$

where $s = f_2/f_1$. The early performance, when there is a small proportion of good structures ($P_t \approx 0$), is given by $P_{t+1} \approx sP_t$. In other words, the selection rate for proportionate selection early in the run is given by $\phi_t = s$. This is similar to the early performance of tournament and truncation selection. However, unlike tournament and truncation selection schemes, the value of s will vary from problem to problem. It is very hard to know a priori whether an arbitrarily scaled problem will have adequate selection pressure. It is

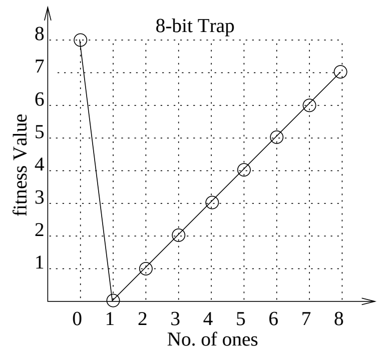

Figure 2: 8-bit trap function

one of the reasons why ordinal selection procedures like tournament and truncation selection are preferred over proportionate selection (Baker, 1985). It is also one of the reasons for using some sort of scaling procedure along with proportionate selection.

Analysis of the late behavior of proportionate selection follows the similar procedure as that of the previous two selection schemes. Modifying equation 20 to incorporate self-crossing we get,

$$P_{t+1} = \frac{sP_t}{(s-1)P_t + 1} \left[1 - p'_c\,(1 - P_t)\right]. \tag{21}$$

Requiring the selection rate to be greater than or equal to one, we get

$$\frac{s}{(s-1)P_t + 1} \left[1 - p'_c\,(1 - P_t)\right] \geq 1. \tag{22}$$

Rearranging the above relation yields

$$p'_c \leq 1 - s^{-1}. \tag{23}$$

This model appears to show that proportionate selection would lead to full convergence in any situation for which the initial situation was favorable. However, it should be noted that s, defined as f_2/f_1, is not a constant during a run. In other words, both f_2 and f_1 can change in every generation, and also that the average fitness f_1 increases as a run goes on. This implies that the average fitness rises and the best BB's progress to dominate the population stalls. This phenomenon can be more accurately modeled by replacing the term s of equation 22 with a term where f_1 or the average fitness rises with increasing P_t. In the next section, we verify the theory with computational experiments.

5 Results

The theory developed in the above two sub-sections is verified with a computational experiment. Similar

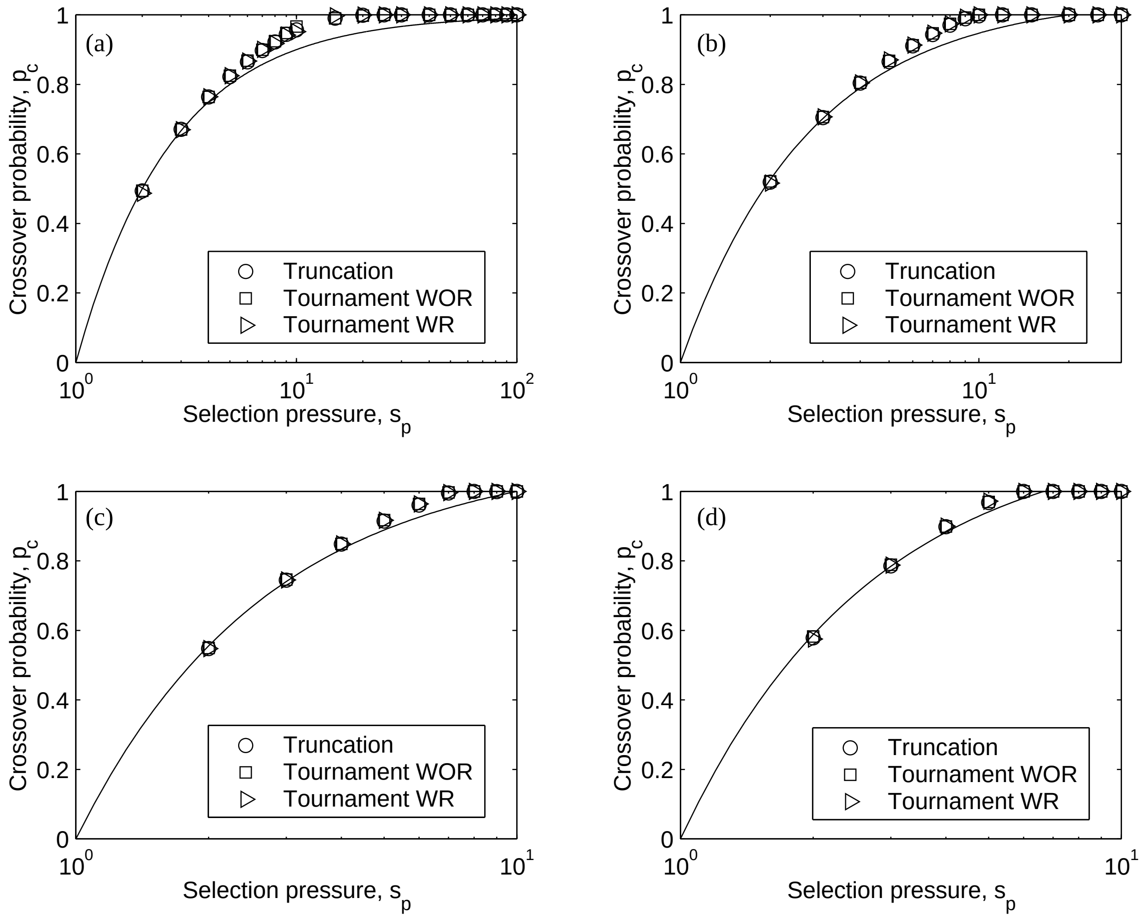

Figure 3: The experimental results showing the required crossover probability at a given selection pressure for tournament with or without replacement, and truncation selection at different disruption rates: (a) $\epsilon = 1.0$, (b) $\epsilon = 0.95$, (c) $\epsilon = 0.9$, and (d) $\epsilon = 0.85$.

control maps are reported elsewhere (Thierens, 1995), although the problem considered in the present study consists of a single building block and has a known schema disruption rate. We optimize a single 8-bit trap function (Goldberg, Deb, & Horn, 1992) (figure 2) with $f_8 = 7$, $f_0 = 8$, and trap break fitness $f_z = 0$ at $z = 1$. Single point crossover is used with either tournament selection (with and without replacement) or truncation selection and the disruption rate is fixed.

A trap function is used to make it difficult to find the best point (all-zeros, 00000000) and easy to find the second best (all-ones, 11111111). Experiments are run for specified s values to determine the crossover probability p_c when 25 independent runs successfully increase the proportion of the best strings for 10 consecutive generations. The population size taken was $N = 5000$. Bisection method was used to determine the p_c value within a tolerance of 10^{-5}. The results shown in figure 3 are average of 10 such independent runs of the bisection method. Both tournament and truncation selection agree with the design equation well at different disruption rates, $\epsilon = 0.85$, 0.9, 0.95, and 1.0.

The theory on proportionate selection is verified using stochastic universal selection. In these runs, neither scaling nor ranking is employed and hence the selection pressure cannot be manipulated independent of the problem scaling. This results in three cases: (1) convergence to the best point (All-zeros, 00000000), (2) stall of the best point, at some proportionate value, and (3) mis-convergence to the second best point (all-ones, 11111111). These three cases are exemplified in figures 4(a)–(d). The results shown are based on 25 successes in 25 trials in a population of 5000.

Figure 4(a) shows a successful takeover of the population by the best point. Initially, the inferior point (all-ones) grows faster, but due to low crossover probability

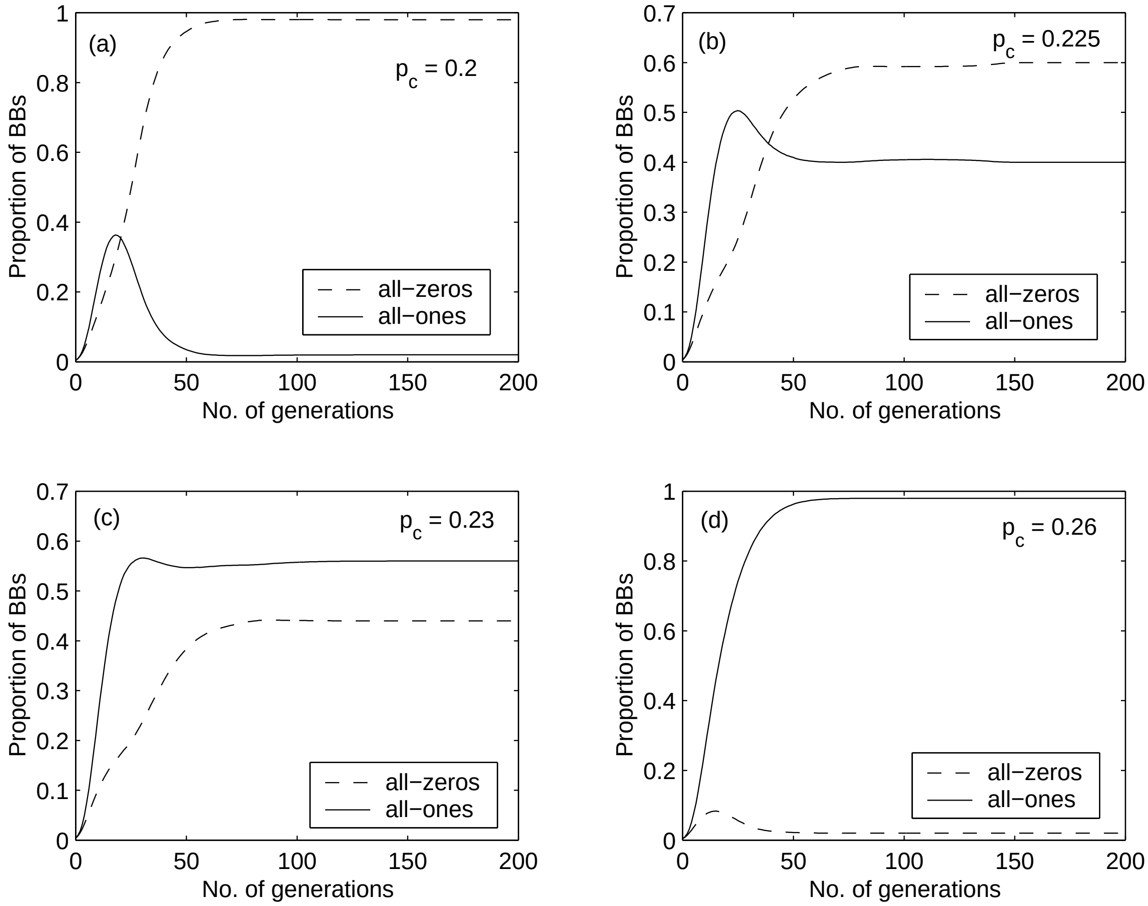

Figure 4: The proportion of all-ones points and all-zeros points is plotted versus generation number at different crossover probabilities. The results are averaged over 25 independent runs.

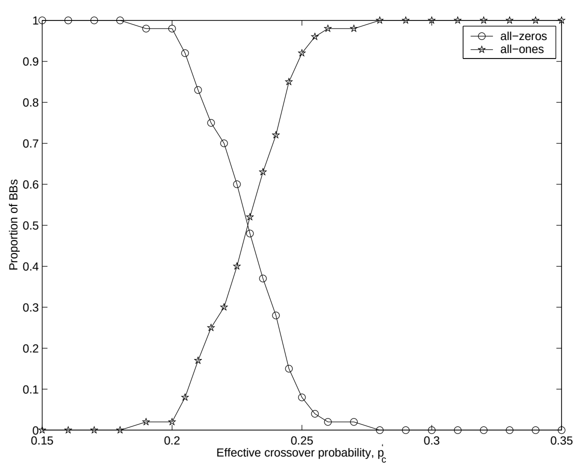

Figure 5: The proportion of all-ones points and all-zeros points when stall occurs is plotted versus crossover probability. The results are averaged over 100 independent runs.

($p_c = 0.2$), the superior point is able to completely take over the population. Figures 4(b)–(c) demonstrates the stall of all-zeros proportion at crossover probabilities, $p_c = 0.225$, and $p_c = 0.23$ respectively. Finally, figure 4(d) indicates failure of the best point to takeover the population at $p_c = 0.26$. The proportion of best and the second best point when stall occurs is plotted against crossover probability in figure 5. The results are averaged over 100 independent trials. The plot shows that at low crossover probabilities the best point successfully takes over the population, at high crossover probabilities the best point fails to sustain the market share increase, and at intermediate crossover probability values the best point stalls at some intermediate proportion value. However, it should also noted that we can have either a total success or total failure in take over of the population by the best point. Examples of the same phenomena are shown in figures 6(a)–(d). In all these cases the proportion of best point stalls for a long time (30–70 generations) and then either it succeeds or fails to takeover

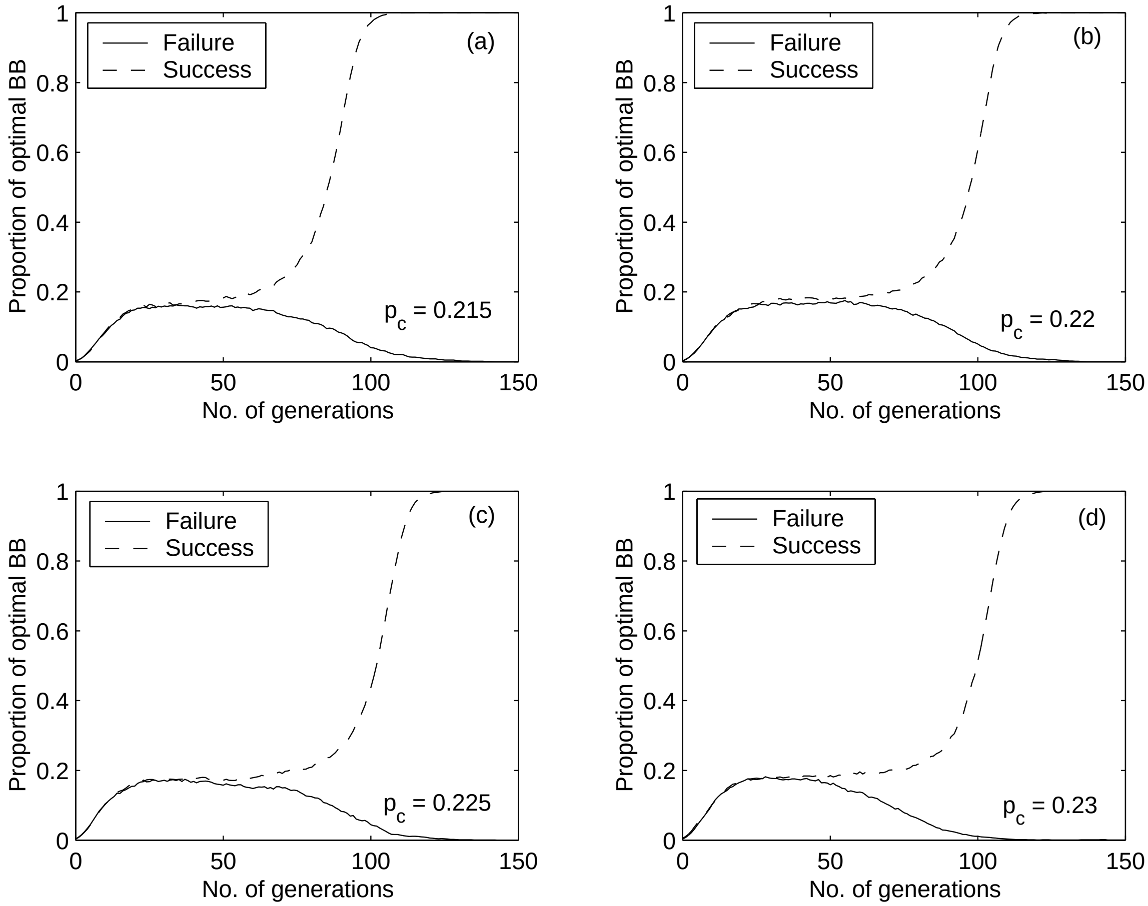

Figure 6: Demonstration of success and failure after intermediate stall at different crossover probabilities (a) $p_c = 0.215$, (b) $p_c = 0.22$, (c) $p_c = 0.225$, and (d) $p_c = 0.23$. The results are of a single run

the population. The results shown are at crossover probabilities of 0.21, 0.215, 0.22, and 0.225. These results indicate the reason for preferring pushier schemes like tournament and truncation selection over proportionate selection.

6 Conclusion

This study clearly shows that the schema theorem works with different selection schemes and genetic operators and that it must be obeyed. The schema theorem provides a good bounding advice on how to assure the growth of good subsolutions and to sustain the growth to takeover the population. Employing early and late analysis, it has been demonstrated that one can set the GA parameters, the selection pressure, s and the crossover probability p_c for tournament and truncation selection schemes to obey schema theorem. Unscaled proportionate schemes have a tendency to stall, indicating that proportionate selection schemes though useful when applied with scaling procedures or niching, should not be generally used without such augmentation. The fact that schema theorem can be satisfied with a crossover probability of zero suggests that schema theorem does not consider the positive effects of crossover, the exchange of BBs that is at the heart of GA mechanics.

Acknowledgments

This work was sponsored by the Air Force Office of Scientific Research, Air Force Materiel Command, USAF, under grant F49620-00-1-0163. Research funding for this work was also provided by the National Science Foundation under grant DMI-9908252. Support was also provided by a grant from the U. S. Army Research Laboratory under the Federated Laboratory Program, Cooperative Agreement DAAL01-96-2-0003. The U. S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation thereon.

The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Air Force Office of Scientific Research, the National Science Foundation, the U. S. Army, or the U. S. Government.

Thanks are also due to Martin Pelikan, Ravi Srivastava and Abhishek Sinha for careful readings of this paper and a number of comments.

References

Baker, J. (1985). Adaptive selection methods for genetic algorithms. In Grefenstette, J. (Ed.), *Proceedings of the International Conference on Genetic Algorithms and Their Applications* (pp. 101–111). Hillsdale, NJ: Lawrence Erlbaum Associates.

Baker, J. (1987). Reducing bias and inefficieny in the selection algorithm. In Grefenstette, J. (Ed.), *Proceedings of the Second International Conference on Genetic Algorithms* (pp. 14–21). Hillsdale, NJ: Lawrence Erlbaum Associates.

De Jong, K. (1975). *An analysis of the behavior of a class of genetic adaptive systems.* (university microfilms no. 76-9381), University of Michigan, Ann Arbor.

Goldberg, D. (1989). *Genetic algorithms in search optimization and machine learning.* Reading, MA: Addison-Wesley.

Goldberg, D., & Deb, K. (1991). A comparitive analysis of selection schemes used in genetic algorithms. In *Foundations of Genetic Algorithms*, Volume 1 (pp. 69–93). San Mateo, CA: Morgan Kaufmann.

Goldberg, D., Deb, K., & Clark, J. (1992). Genetic algorithms, noise, and the sizing of populations. *Complex Systems*, *6*, 333–362.

Goldberg, D., Deb, K., & Horn, J. (1992). Massive multimodality, deception, and genetic algorithms. In Manner, R., & Manderick, B. (Eds.), *Parallel Problem Solving from Nature*, Volume 2 (pp. 37–46). UK: Elsevier Science Publishers.

Goldberg, D., Korb, B., & Deb, K. (1989). Messy genetic algorithms: Motivation, analysis, and first results. *Complex Systems*, *3*(5), 493–530.

Goldberg, D., Sastry, K., & Latoza, T. (2001). On the supply of building blocks. In *Proceedings of Genetic and Evolutionary Computation Conference.* San Francisco, CA: Morgan Kaufmann.

Grefenstette, J., & Baker, J. (1989). How genetic algorithms work: A critical look at implicit parallelism. In Schaffer, J. (Ed.), *Proceedings of the Third International Conference on Genetic Algorithms* (pp. 20–27). San Mateo, CA: Morgan Kaufmann.

Holland, J. (1975). *Adaptation in natural and artificial systems.* Ann Arbor, MI: University of Michigan Press.

Muhlenbein, H., & Schlierkamp-Voosen, D. (1993). Predictive models for the breeder genetic algorithm: I. Continous parameter optimization. *Evolutionary Computation*, *1*(1), 25–49.

Schaffer, J. (1987). Some effects of selection procedures on hyperplane sampling by genetic algorithms. In Davis, L. (Ed.), *Genetic Algorithms and Simulated Annealing* (pp. 89–103). Los Altos, CA: Morgan Kaufmann.

Thierens, D. (1995). *Analysis and Design of Genetic Algorithms.* Doctoral dissertation, Katholieke Universiteit Leuven, Leuven, Belgium.