Gendered Selection Strategies in Genetic Algorithms for Optimization

José Sánchez-Velazco
School of Computer Science
The University of Birmingham
Birmingham, B15 2TT
UK
J.G.Sanchez@cs.bham.ac.uk

John A. Bullinaria
School of Computer Science
The University of Birmingham
Birmingham, B15 2TT
UK
J.A.Bullinaria@cs.bham.ac.uk

Abstract

The selection operator in the standard genetic algorithm (GA) determines which individuals are chosen from a relatively homologous population for mating and crossover. This operator is crucial for the performance of the GA, since it may lead the algorithm to premature convergence and limited search scope (or genetic diversity) by repeatedly choosing very strong individuals with similar genetic code. In the model proposed here, a sexual strategy is introduced by simulating distinct gender groups, with each gender having different partner selection criteria, and a model of sexual selection that allows for competition between individuals in the same group and co-operation when a mating relation is established. As in natural systems, crossover is only permitted between individuals in contrasting gender groups, and the mutation probabilities depend on the individual's gender. Experimental results on some standard optimization problems provides evidence that this is a useful strategy.

1 Introduction

The design of Genetic Algorithms (GAs) has been based on inspiration from Darwinian natural selection [3] and biological genetic evolution. Most GA systems include common relevant features from living organisms, such as genetic recombination (by means of crossover), mutation, and fitness assessment for selection. However, one of the most common biological features is not included in the standard algorithm, namely, gender [6, 11].

With the appearance of gender, organisms became more "adaptable" to their environment by inheriting beneficial traits from both parents favoured by selection. Also, a more efficient process of resource and energy administration became possible by task-specialization and task-separation according to each gender's advantages [8, 7]. This paved the way for the appearance of some well known animal behaviours, such as co-operation between complementary genders, which helped with the ever present need to move towards an optimization of the problem of survival.

A few gendered GA models have already been proposed in the literature [2, 9, 10, 12], but in most cases the inclusion of gender is merely limited to a representation of an objective function (in multi-objective optimization), or as a tag in the chromosome preventing crossover with other individuals bearing the same gender flag. Many of the most fundamental observed principles of sexual selection, such as the derived competition and co-operation patterns within and between gender groups, are not modelled.

The strategies proposed here aim to extend the GA towards a more "natural" approach by the inclusion of gender and sexual selection in the population. A different fitness assessment method is used for each gender, thus modelling natural preferences in partner selection. Male individuals are selected based on their *direct* or *competitive* fitness, which is the output of the standard GA quality function over the chromosome. Females, however, are selected based on their *indirect* or *co-operative* fitness, which includes a measure of the potential of the individual to increase her offspring's direct fitness (compared to their father), and her fertility capabilities according to her age and life-span.

The organisation of this paper is as follows: Section 2 provides a full description of the gendered strategies, Section 3 discusses some experimental results using three different optimisation problems, and Section 4 gives some conclusions and recommendations for future work.

2 The Gender Strategy

Sex is believed to have "evolved" from asexual organisms some 2 billion years ago. It has prevailed as a mechanism for reproduction and task separation since then, and it is perhaps one of the most commonly observed features in nature. Sexual reproduction presented a number of advantages over asexual mechanisms, and made possible a more varied genetic pool [7]. For instance, if a deleterious genetic mutation took place in an asexual individual, it would be transmitted to all descendants unless a back-mutation reverting the gene, or an additional mutation compensating its effects, appeared. With sexual reproduction, a simple genetic recombination with a partner not having this mutation could produce offspring completely free of the change.

Although various sexual attractors are very common in nature, their exact functions are often still unclear. However, we may safely assume that important principles lie behind them, for they have passed nature's test of evolution [2], and in fact, they have led to life's most common and successful strategies for reproduction. Despite the apparent risk for survival that some phenotypic and behavioural traits based on sexual selection represent to the bearer (for example, the brighter color and longer heavier feathers of the male longtailed widow bird, which make him easier to spot and more vulnerable to predators), they endow the individual with better chances for getting a mate for reproduction. In the case of many bird species, the pompous quirks observed in them are an instrument for fitness assessment in the sexual selection process that will reward the fittest with the proliferation of his genes.

The GA strategies proposed here are modelled on the following biological considerations:

- Gender: In most gendered species, males and females have different phenotypic traits (*sexual dimorphism*), and usually these groups are responsible for different subsets of the tasks that as a whole are vital for survival [8, 6, 11].

- Mutation Rates: It is a recently established fact that, in most species, males and females have considerably different rates of certain types of mutations [11, 6, 5].

- Sexual Selection: Males and females tend to have distinct preferences when it comes to selecting a partner for mating.

- Fertility and age influence the individual's fitness and help maintain genetic diversity in the population by allowing new individuals to intervene in crossover.

The rest of this section will specify the various elements of the gender extension to the GA we are proposing.

First, let P be the population of the GA, and let X and Y be two proper subsets of P representing females and males respectively. Following the definition of proper subsets $P = X \cup Y$ and $X \cap Y = \emptyset$. A system parameter γ_Y then denotes the fraction of individuals in set Y, so it follows that at any given time, the probability of an individual $a \in P$ being in set Y or X is:

$$p(a, Y) = \gamma_Y \quad , \quad p(a, X) = 1 - \gamma_Y \qquad (1)$$

Mating takes place in pairs, with one male and one female parent, and using crossover, two new individuals are produced, one of each gender. The male parent is selected first using a standard selection procedure based on his task fitness. After the male parent has been selected, a female parent is then chosen using a sexual selection scheme that includes direct fitness, age and fertility potential.

The *direct* fitness of an individual is the outcome of the task performance quality or fitness function applied to the individual's genotype, and therefore it will also be referred to as the individual's *competitive* fitness. The *indirect* fitness measure that is used to select the female parents can be described as the weighted average of the individual's direct fitness, an age function and her potential to produce fit offspring compared to her partner's direct fitness. A female's indirect fitness will also be referred to as her *cooperative* fitness.

To make all this more precise, let us define our selection function to be $Sel(\phi(a))$, which returns one chosen individual from the set of all competing individuals a with fitness function ϕ. The choice will normally involve some form of randomness, and will always be biased according to the fitness ϕ. The male (y) and female (x) parents are selected according to:

$$\forall y \in Y, \quad \forall x \in X$$

$$y_{sel} = Sel(f(y)) \qquad (2)$$

$$x_{sel} = Sel\left(\frac{w_1 f(x) + w_2 \Delta f(y) + w_3 g(Age(x))}{w_1 + w_2 + w_3}\right) \qquad (3)$$

in which the w_i are a set of fixed weighting parameters that we must choose to specify appropriate relative importance to the three components.

Both male and female selection depends on the direct fitness $f(y)$. The second factor for females is

$$\Delta f(y) = f(y_{son}) - f(y_{sel}) \tag{4}$$

which is a measure of how much the female's contribution is likely to enhance a male offspring's fitness over it's father's. Clearly, Δf will have a value of zero for the initial generation. Thereafter, whenever a male offspring has been produced and compared with it's father, Δf can be computed for that female for the next selection process. This is not a disadvantage for the oestrogenic individuals, because the sexual selection scheme involves no competition between contrasting gender groups.

The third factor, involving the function $g(Age(x))$, is introduced for females to provide an effect of age on the chances of being selected. In biological populations it will depend in a complex manner on perceived likelihoods of producing healthy offspring and being experienced at looking after them. For our purposes, a simple triangular function of width σ around the age of maximum fertility μ proves adequate. Thus for $Age(x) < \mu + \sigma$ we have

$$g(Age(x)) = 1 - \frac{|Age(x) - \mu|)}{\sigma} \tag{5}$$

and beyond that age it is zero. For practical purposes we can take $\mu + \sigma$ to be the given individual's life-span.

The combination of this selection scheme and the crossover operator steadily evolve a symbiotic relationship. The rationale for this is that each gender specialises on a substring of the genome. For example, with a one-point crossover, females may contribute to the trailing part of the genome string. The better a female's contribution is to her offspring's fitness, the better her fitness will also be, and hence the better her chance to get selected again. Such co-operative patterns between opposite genders allow the selection and crossover operators to "adapt" to each other.

As with all GAs, each individual $a \in P$ will be specified by a chromosome that represents the given problem solution as a string of length L in some generalised alphabet. We then have to define appropriate mutation and crossover operators on those chromosomes.

Mutation is implemented at different fixed rates for each gender group, and following biological systems, the male's rate is higher than the female's [11, 6, 5], so

$$\forall y \in Y, \forall x \in X$$

$$m_y > m_x \tag{6}$$

For all the experiments presented in this paper, a standard one-point crossover was used, but there are obviously many other possibilities.

3 Experimental Results

Three different standard optimization problems were selected to test our proposed gendered selection strategy. The first was the exponential NP-hard Travelling Salesman Problem (TSP), the second was one of DeJong's test functions in his GA analysis [4], and the last was a function with many local optima in its landscape [10]. Our sexual selection GA and the standard GA were compared using the results from each test problem performed with all other parameters matched.

For each test function, the mutation rate for the standard GA was computed using

$$m_P = \gamma_Y m_Y + (1 - \gamma_Y) m_X \tag{7}$$

where γ_Y is the proportion of males in the population of the gendered GA, and m_Y is the male mutation rate in that algorithm. From equation 1, we can see that the effective mutation rate for both algorithms should then be approximately equal.

In all our experiments the well known "roulette wheel" selection method was used so as to maintain similarity with the standard GA for performance comparisons, and a one-point crossover was used. All the results presented are averaged over 100 experimental runs.

3.1 Test Function 1 (The TSP)

The TSP is a classic combinatorial problem that has been widely used to study different optimisation algorithms in the past. The task is to find the minimum length route that visits, with no repetition, each member of a set of points/nodes in some metric space. Many evolutionary methods have already been proposed for this problem, and some effective evolutionary operators have been design to boost the algorithms' performance. For more detailed information on the TSP, and references to other proposed approaches, see [1].

To ensure a fair comparison of our sexual selection strategy against the standard GA, the

Nodes	Standard GA		Gendered GA	
	DeJ's X	Optima	DeJ's X	Optima
20	0.7382	96	0.9828	100
50	0.2370	0	0.7189	44
200	0.0350	0	0.1009	0
500	0.0121	0	0.0265	0

Table 1: Comparison of performance for 100 experimental runs of the Standard and Gendered GA for the TSP problem.

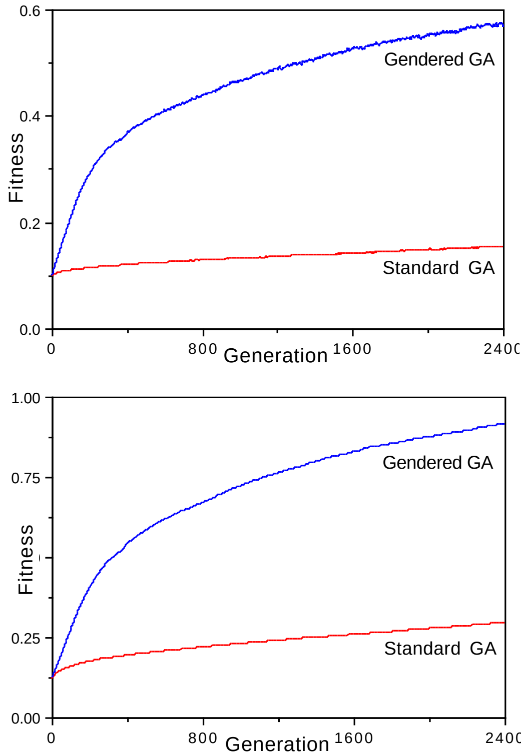

Figure 1: 50-node TSP results for the standard and gendered GA. Fitness averages over the population (above), and average best individual fitnesses (below).

standard genetic operators for the TSP were used. The usual TSP representation was implemented, i.e. the chromosome consisted of the list of points in the order they are visited, so a chromosome (2 4 1 3) represented the tour $2 \rightarrow 4 \rightarrow 1 \rightarrow 3 \rightarrow 2$. Mutation simply consisted of swapping two cities in the list.

The initial parameters were population size $|P| = 200$, crossover probability of 0.75 and elitism proportion of 0.07. The gendered strategy parameters, used in equation 3, were $w_1 = 0.75$, $w_2 = 0.55$ and $w_3 = 0.18$. The perceived fertility function, defined in equation 5, had parameter values $\mu = 2$ and $\sigma = 4$. The mutation rate for males was $m_Y = 0.1$, and for females it was $m_X = 0.001$, and hence, from equation 7, the mutation rate for the standard GA was 0.0505. The proportion of male individuals in the population was 0.5 (i.e. 50%).

The metric used for comparison of the gendered and standard GA performance was De Jong's off-line measure [4], defined as

$$X_e^*(h) = \frac{1}{T}\sum_{t=1}^{T} f_e^*(t) \tag{8}$$

where $f_e^*(t) = best\{f(a_1), f(a_2), \ldots, f(a_{|P|})\}$ at generation t, and T is the total number of generations. This measure is thus the average of the best individuals over all the generations produced by the algorithm. Since all our experiments were based on 100 trials, our off-line measure is then averaged over 100 runs.

In Table 1 the performance comparison of the gendered and standard GA for the TSP is shown. The main indicator given is DeJong's off-line measure and the table also shows the number of known global optima that each algorithm achieved. As mentioned before, no problem-oriented operators were used e.g. edge preservation crossover, etc. since we were aiming to compare the behaviour of the strategy against the standard GA. Figure 1 shows the typical increases in performance with generation we get for the two approaches.

3.2 Test Function 2

The second test function used was proposed by DeJong in his analysis of the behaviour of a GA [4]. This function is defined as

$$f(x, y) = 100(x^2 - y)^2 + (1 - x)^2$$

$$-2.048 \leq x, y \leq 2.048$$

Figure 2 shows the function. The representation used was a binary Gray code chromosome.

Here, different parameters from the TSP case are appropriate. For both GAs, the population size was 80 individuals iterated for 14 generations, and an elitism proportion of 0.07 was used. Then for the gendered GA, the male mutation probability was set at 0.1, the female mutation probability at 0.001, and the proportion of males in the population was 0.5. Again using equation 7, the mutation rate for the standard GA was set at 0.0505. The w_i factors in equation 7 were $w_1 = 0.65$, $w_2 = 0.2$ and $w_3 = 0.15$.

Table 2 shows the performance of the gendered and standard GA for test function 2.

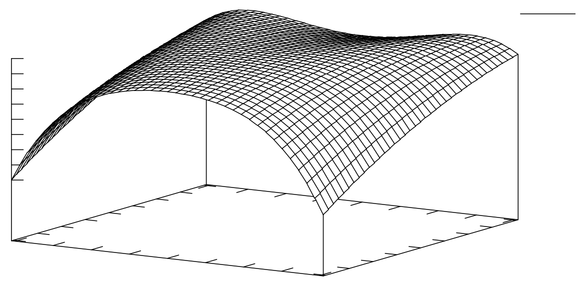

Figure 2: Test Function 2. This was previously used for analysis of the behaviour of genetic algorithms in DeJong's studies [4].

Metric	Standard GA	Gendered GA
Best	3905.89	3905.93
Pop. Fit	3515.99	3529.40
Best Fit	3905.67	3905.65

Table 2: Best solution obtained, population fitness average and best fitness average for the standard and gendered GA on test function 2.

The best solution obtained by each algorithm is shown on the first row, while the average fitness of the population and the average fitness of the best individual over the experimental trials are shown in rows 3 and 4 respectively. Figure 3 shows how the fitness varies with generation.

The differences here are smaller than for the TSP case, so we performed t-tests to provide a rigorous comparison of the results from each algorithm. In this case, the sexual selection GA does perform significantly better than the standard GA in terms of the population average fitnesses ($p < 0.01$), but not in terms of the best individual fitnesses (p = 0.7).

3.3 Test Function 3

For the last test, a function used in [10], containing many local optima in its landscape, was chosen. It is defined as

$$f(x,y) = x^{1/3}sin(x) + y^{1/3}sin(y) + c$$

$$0.0 \leq x, y \leq 32.0$$

in which the constant c is an arbitrary integer value chosen to keep the function with positive values. Figure 4 provides a scene of the function's multi-peak nature. As with test function 2, the chromosome here was a binary Gray coded representation.

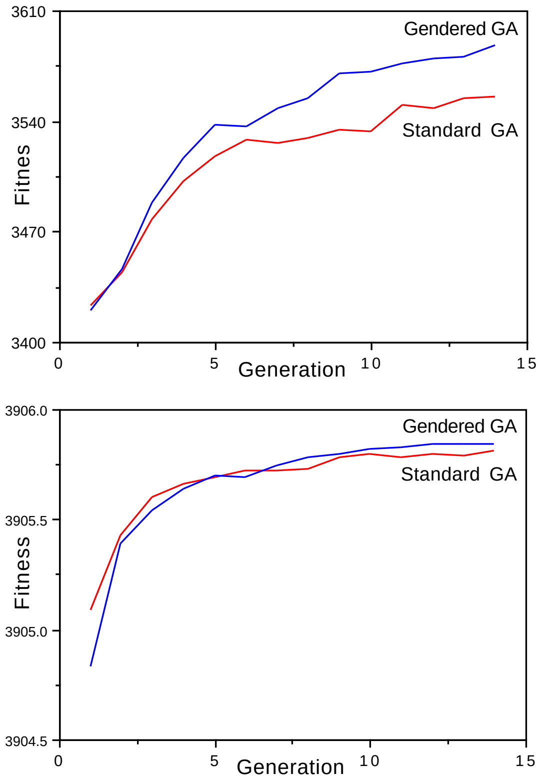

Figure 3: Results for the standard and gendered GA on test function 2. Fitness average of the population (above), and average best individual fitness (below).

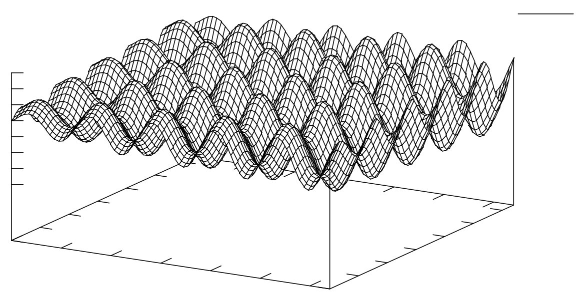

Figure 4: Test Function 3. This test function has a landscape, or problem search space, with many local optima.

Metric	Standard GA	Gendered GA
Best	14.977	14.978
Pop. Fit	9.660	9,667
Best Fit	14.712	14.736

Table 3: Best solution obtained, population fitness average and best fitness average for the standard and gendered GA on test function 3.

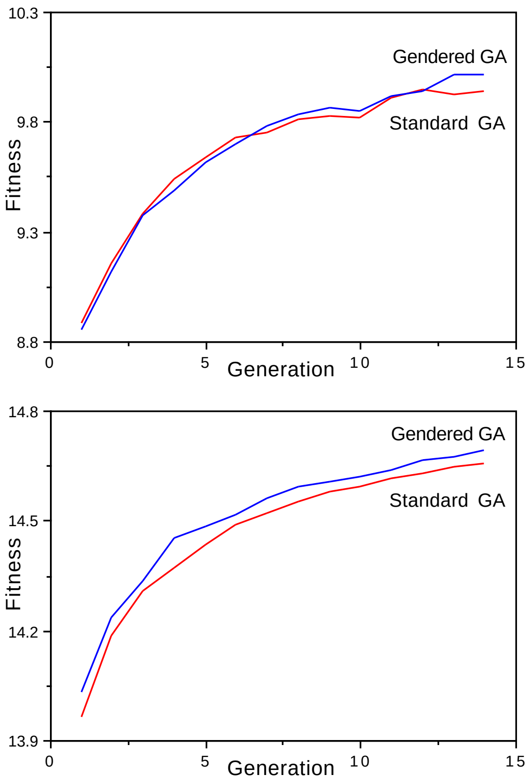

Figure 5: Results for the standard and gendered GA in test function 3. Fitness average of the population (above), and average best individual fitness (below).

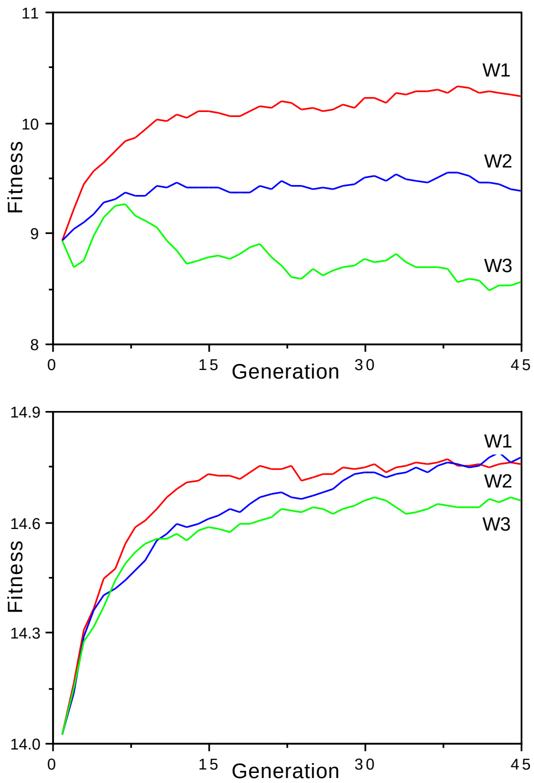

Figure 6: A comparison of the population average fitness behaviour (above), and the best average fitness (below). The Wi indicate the factor which has non-zero weight w_i in equation 3.

The parameters used in this case were a population size of 80 and 14 generations, and an elitism proportion of 0.07. For the gendered GA, a male mutation rate of 0.1 was used, and the females had a rate of 0.01. The male proportion in the population was again 0.5. The w_i factors in equation 7 were $w_1 = 0.5$, $w_2 = 0.3$ and $w_3 = 0.22$.

The results for this test function using the gendered and standard GA are given in table 3. Figure 5 shows how the fitness varies with generation. In this case, there is no significant difference between the sexual selection GA and the standard GA in terms of the population average fitnesses ($p = 0.5$), but the gendered GA is significantly better in terms of the best individuals ($p < 0.01$).

4 Conclusions

We have presented gendered selection strategies for GAs that model the competitive behaviour observed between individuals of the same gender, and the co-operative patterns that emerge between different gender groups. Distinct mate selection patterns, mutation rates and age parameters for the two gender groups are likely to help the algorithm avoid local optima stagnation and possible exponential take-over of strong individuals.

Results from our preliminary experiments suggest that this is a promising approach that can significantly outperform the standard GA, both in terms of the number of generations required and in the quality of the solutions. There is clearly a certain degree of problem dependence in the improvements, and this is something we plan to explore more systematically in the future.

The extra parameters required for the sexual selection scheme certainly add complexity to the algorithm. However, they also provide more flexibility for the designer by allowing freedom to set the weights (w_i) of the underlying factors according to the nature of the problem being dealt with.

Intuitively, we can arrive at a general strategy for understanding and setting the sexual selection weight parameters:

- w_1 (Competitive fitness factor) is the weight given for the quality function of the individual. Setting higher weight on this parameter may result in having better off-line mark on average.

- w_2 (Age and fertility factor) should improve genetic diversity in the population by not allowing very strong individuals take over the genetic pool. In problems where many good solutions are wanted (e.g. design problems), w_2 should have a relatively strong weight

- w_3 (Co-operative fitness factor) is the weight given for the potential of the female to improve her offspring's competitive fitness. On average, if w_1 or w_3 are high, the on-line [4] and off-line measures of the algorithm are likely to improve. However, if the proportion of males in the population is either very high or very low, (i.e. the male:female ratio is very distant from 50:50), then w_2 is likely to make the algorithm find a better solution.

Figure 6 provides an initial indication of the effects on the gendered GA performance results for test function 3 due to each the three weight factors w_i. For the $W1$ plots, the weight factor w_1 was set to 1.0 and the other two were eliminated with a value of 0.0. Similarly, the $W2$ plots correspond to setting the weight factor w_2 to 1.0 with w_1, w_3 set to 0.0, and finally, $W3$ has w_3 set at 1.0 with w_2, w_3 set to 0.0. We see there are clear differences between the effects of the three weights, and also between the results for population averages and best individuals. Plots such as these will provide a valuable source of information for our future investigations of the problem dependences of the gendered GA performances.

Further research is also planned to explore the best general parameter-tuning and parameter control strategies, as well as the nature of the trade-offs of different mutation rates between genders, to build a more solid theoretical basis for the whole gendered GA approach.

Acknowledgements

The first author acknowledges the support of Decanato de Ciencias y Tecnología, Universidad Centroccidental "Lisandro Alvarado", Barquisimeto, Venezuela.

References

[1] Traveling salesman problem - home page, http://www.math.princeton.edu/tsp/, May 2002.

[2] R. Allenson. Genetic algorithms with gender for multi-function optimisation. Technical Report EPCC-SS92-01, Edinburgh Parallel Computing Centre, 1992.

[3] C. Darwin. *The Origin of Species.* Oxford, 1998 (Original edition 1859).

[4] K. A. De Jong. *Analysis of the Behavior of a Class of Genetic Adaptive Systems.* PhD thesis, Department of Computer and Communication Science, University of Michigan, 1975.

[5] C. Dennis and R. Gallagher. *The Human Genome.* Nature Publishing Group, 2001.

[6] S. Freeman and J. Herron. *Evolutionary Analysis.* Prentice-Hall, 2001.

[7] R. Gadagkar. *Survival Strategies Cooperation and Conflict in Animal Societies.* Harvard University Press, 1997.

[8] D. Kimura. *Sex and Cognition.* Cambridge, MA: MIT Press, 2000.

[9] J. Lis and A. Eiben. A multi-sexual genetic algorithm for multiobjective optimization. In *Proceedings of the 1997 IEEE International Conference on Evolutionary Computaiton*, pages 59–64. IEEE, 1997.

[10] J. Rejeb and M. AbuElhaija. New gender genetic algorithm for solving graph partitioning problems. In *Proceedings of the 43rd IEEE Midwest Symposium on Circuits and Systems*, volume 1, pages 444–446. IEEE, 2000.

[11] S. Stearns and R. Hoekstra. *Evolution.* Oxford: Oxford University Press, 2001.

[12] D. Vrajitoru. Simulating gender separation with genetic algorithms. In *Proceedings of the Genetic and Evolutionary Compuatation Conference (GECCO 2002)*, pages 634–641, 2002.