

## Genetic Algorithms and Evolution

B. H. SUMIDA†, A. I. HOUSTON†‡, J. M. MCNAMARA§  
AND W. D. HAMILTON†

† *Department of Zoology, Oxford University, South Parks Road, Oxford OX1 3PS*, ‡ *King's College Research Centre, King's College, Cambridge* and  
§ *School of Mathematics, Bristol University, Bristol, U.K.*

(Received on 4 August 1989, Accepted in revised form on 6 April 1990)

The genetic algorithm (GA) as developed by Holland (1975, *Adaptation in Natural and Artificial Systems*. Ann Arbor: University of Michigan Press) is an optimization technique based on natural selection. We use a modified version of this technique to investigate which aspects of natural selection make it an efficient search procedure. Our main modification to Holland's GA is the subdividing of the population into semi-isolated demes. We consider two examples. One is a fitness landscape with many local optima. The other is a model of singing in birds that has been previously analysed using dynamic programming. Both examples have epistatic interactions. In the first example we show that the GA can find the global optimum and that its success is improved by subdividing the population. In the second example we show that GAs can evolve to the optimal policy found by dynamic programming.

### Introduction

Darwin stressed that the action of natural selection could result in animals that appear to be well-designed. This view of natural selection has been developed into the application of optimality theory to the study of the morphology and behaviour (e.g. Alexander, 1982; Maynard Smith, 1978; Goel & Thompson, 1986; Oster & Wilson, 1978; see also Borstnik *et al.*, 1987; Eigen & Schuster, 1979). Although the use of optimality to study evolution is widespread, it is also criticized (e.g. Gould & Lewontin, 1979).

In this paper we make no attempt to review these arguments. Instead we consider recent attempts to use the evolutionary process to provide optimization techniques (e.g. Holland, 1975; Wang, 1987; Galar, 1989; for some related ideas based on neurobiology and physics see Bounds, 1987). In particular we concentrate on Holland's genetic algorithms, and use them to study biological problems. We begin with a historical background to genetic algorithms. The genetic algorithm kernel (GAK), based upon the original model (Holland, 1975) is described. The basic GAK model is modified to include additional biological features. We are interested in investigating how various aspects of natural selection influence the efficiency of the GA as a search procedure. In as much as we are following the spread of genes through a population, we are doing population genetics. We are, however, looking at fitness landscapes which are highly complex, with many local optima or strong epistatic effects. It is a whole research programme to explore how evolution depends

on all aspects of an organism's biology. Of necessity, we must restrict attention to a few features. A modified GAK is developed to incorporate features of the Shifting-Balance Theory of evolution as proposed by Wright (1977). Two case studies, both involving epistatic effects, are described. The first examines the evolution of an optimal genotype in a multiple-peaked fitness landscape. We investigate the effects of population structure, mutation rate, selection strength and migration frequency on evolution's ability to reach the global optimum. The second case study describes the evolution of daily routines of singing and foraging in a small bird. There has been considerable discussion about the evolutionary advantages of daily routines of singing and foraging (see Kacelnik & Krebs, 1982 for a review). McNamara *et al.* (1987) found optimal routines using dynamic programming. We show that these GAs can evolve to produce these optimal routines with relative ease.

### Background and History

The foundation of genetic algorithms (GAs) is contained in the monograph *Adaptation in Natural and Artificial Systems* by Holland (1975). Holland (1975) is the acknowledged founder of the field of GAs. Other computer simulations with a genetic algorithm flavour (e.g. Fraser, 1962; Bremerman, 1962) had appeared in the literature before Holland's (1975) monograph, but his was the first to formalize the theory and apply GAs to artificial systems. The GA has its roots in the field of artificial intelligence as an optimization technique that generates adaptive responses to changing environments. What then is a GA? GAs are exceptionally powerful heuristic search algorithms which use biological evolution as the framework for their search process, executing a structured yet randomized information exchange to produce fitter individuals through time. The algorithms are *genetic* in the sense that their primary operations, mating, mutation and crossover (see below for details), emulate Mendelian populations (see, e.g. Jacquard, 1974). Individuals or organisms are represented by their chromosomes and are implemented in the GA by vectors of data structures. An ensemble of individuals, called the population, generates a new set of individuals at the end of a generation, by mating the fittest individuals from the current generation. Occasionally, a chromosomal mutation occurs, introducing new information and preventing fixation or loss of genetic information.

GAs, when compared to other optimization methods, offer many advantages. GAs are computationally simple despite their efficiency in tackling difficult problems. Two features that may make an optimization problem difficult are a very large search space of possible solutions and the existence of many local optima. On such problems enumerative procedures such as dynamic programming would require prohibitively large amounts of computer time. GAs are, however, highly efficient techniques for such problems (Brady, 1985; DeJong, 1980). Optimization analyses which rely on gradient techniques (i.e. "hill climbing") assume the existence of derivatives, thus imposing constraints on the characteristics of the search space. Such techniques work well when there is a single optimum, but fail when there are many local optima. The GA does not require derivatives, only a payoff, or fitness measure is necessary.

Even if the fitness function has many local optima the GA can perform well (Schuster, 1989). These factors contributed to the success which has resulted from applying the GA method to a host of problems (e.g. Travelling salesman problem, see Brady, 1985).

The applications which followed Holland's (1975) original work were centred around engineering problems. Since then the method has attracted a wider audience resulting in an increased diversity of applications in other fields (see Davis, 1987; Goldberg, 1989). For example, Axelrod (1987) effectively used genetic algorithms to study the evolution of co-operation in the Prisoner's Dilemma of game theory. In observing the emergence of game strategies through simulated evolution. Axelrod (1987) discovered many interesting and surprising behaviours, including the evolution of a strategy that rivalled the celebrated "Tit for Tat" strategy. He went on to suggest that genetic algorithms may be used profitably to investigate the evolution of other sorts of behaviours. The studies described in this paper affirm his prediction.

### **The Genetic Algorithm Kernel (GAK)**

The GA programme is flexible and many variations have been written according to specific requirements. In general, however, GAs will contain the following essential components:

- (a) A chromosomal representation of possible solutions to the problem.
- (b) A fitness (i.e. objective) function to evaluate the performance of individual solutions.
- (c) A reproduction function to produce a new set of offspring.
- (d) A mutation operator.

A rule for solving the problem is encoded by an ordered sequence of numbers. By analogy to biology, we refer to this ordered sequence as a chromosome. In biology, the chromosomes are sequences of genes that contain the information for producing an individual. In the context of GAs a single (haploid) chromosome contains all the information for specifying a rule. Again by analogy to biology, we refer to the position in the sequence of numbers as a locus and the value of the number at a position as an allele. Holland (1975) used bit-string (i.e. elements from the set {0, 1}) representations in his work. Although much less is known about other forms of encoding in GAs, it will be shown in this paper that alternative schemes can be successful.

The success or failure of any given GA may depend on the encoding scheme. At the most basic level, the encoding scheme will determine the set of possible solutions, and thus imposes an evolutionary constraint. We illustrate this later in the paper in the context of the model for singing vs. foraging. For some problems, there will be a natural coding in which different chromosome segments correspond to functionally independent aspects of the problem. Recombination can then bring together the best segments for each aspect of the problem. In many cases, however, there will be a complex interdependence between the performance of any segment of the chromosome and the values coded on other parts of the chromosome. This is likely

to be true when attempts are made to code daily routines or life histories; the model of singing vs. foraging is a clear example.

Having fixed a code, a chromosome represents a particular attempt at a problem-solving rule and the performance of such a rule is measured by a single number which we refer to as the fitness of the chromosome. One of the strengths of the GAs, as originally envisaged by Holland (1975), is the flexibility that results from the fact that the chromosomes are defined independently of the performance criterion for the problem. Thus, if the parameters of the problem, and hence the performance criterion, changes it is possible to utilize the same chromosomes and encoding for the new circumstances. Holland (1975) saw this ability to adapt to changes as an important attribute of an artificial intelligence system. There is a direct biological analogy in that the same method of encoding (genes) are used when the environment changes and hence the fitness of various phenotypes change.

The GA is based on the evolution of a population of chromosomes. In the simplest version, the population has discrete non-overlapping generations. In each generation, the fitness of each chromosome is determined and a new generation of chromosomes is produced. The composition of this new population is determined by mating, recombination and mutation, where the mating success of each chromosome is determined by its fitness.

In the GA sexual reproduction is the process by which two parent chromosomes give rise to offspring by hybridizing the parental chromosomes via recombination. The recombination procedure begins by choosing a random interlocus interval on the chromosome. This point on the chromosome forms the crossover chiasma, or the point of interchange between the two parent chromosomes. Figure 1 illustrates the crossover operation.

Parents are chosen in pairs from the original population according to some probability distribution. These probabilities are chosen in such a way that the expected number of offspring that a parent has is an increasing function of its fitness. The total number of offspring produced is usually constrained so that the total number of chromosomes in a generation is kept constant.

The offspring, constructed from elements of the fittest chromosomes in the previous generation, form the new population. Such selective pressure, over generations, tends to raise mean fitness. Holland (1975) argues that the power of the GA—and of evolution by natural selection—stems from two separate processes that work in concert to direct progress towards a better solution. Selection isolates and preserves the best building blocks—subsets of coadapted alleles. Crossover effects a highly efficient search, testing subsets of information rather than single points. He made this observation of GAs and applied the term “implicit parallelism” to describe the process. A consequence of subset recombination emerges: crossover enables the bringing together of two or more beneficial mutations which occur on different homologous chromosomes. The probability of a single chromosome acquiring two beneficial and rare mutations by chance would be exceedingly low.

Mutation is a necessary component of GAs. Without mutation the population will lose alleles through fixation or deletion. Mutation rates in GAs, as in nature, are usually quite low. Because mutation is undirected a high mutation rate is more

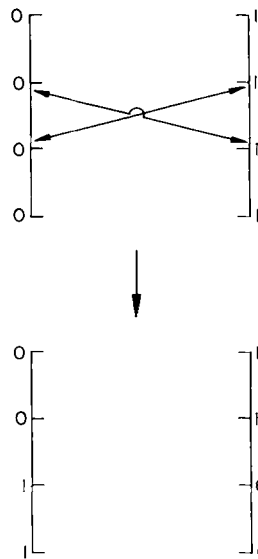


FIG. 1. The crossover operation in the genetic algorithm. Offspring are formed, with a process analogous to sexual recombination, by hybridizing genetic material from two parents. In the upper half of the figure, the crossover point, where genetic material is exchanged between the parents, is indicated by a chiasma. The lower half of the figure illustrates the offspring resulting from a crossover at the indicated position. Although two offspring logically result from recombination of two parents, only one offspring is kept.

likely to disrupt the GA than to improve its performance. Mutation in the GAK is carried out after a new population has been formed and involves a random change at selected loci.

In our version of the GA, each locus on each chromosome has a given small probability of undergoing a mutation. If occurrence of a mutation is determined, the resulting allele is chosen at random from the set of all possible alleles at that locus.

### The Modified Genetic Algorithm Kernel

The GAK outlines the basic plan for any genetic algorithm simulation. In order to be able to investigate which features of natural selection result in an effective search procedure, we modify the GAK to incorporate certain additional features.

Only a single crossover point is used in the GAK. This does not coincide with observations of nuclear division in cells, where crossover may occur repeatedly along the chromosome (Watson *et al.*, 1987). Multiple crossover, with an independent probability of interlocus crossover, is possible in the modified GAK.

Populations in the GAK, like annual plants, are entirely replaced after each generation, forming non-overlapping generations. Our modifications allow overlapping generations, i.e. a fraction of the population up to the entire population can be replaced in the next generation.

Selection was simulated using a form of soft (see Wallace, 1975 for a definition) threshold, or truncation, selection: a fixed percentage of least fit members of the

population was "killed" every generation. Each individual in the current population that "dies" is replaced by an offspring formed from randomly selected parents. Only one offspring is produced from each mating. Offspring are assigned to replace current least fit individuals and the replacement proceeds in an ascending fitness order. The mortality fraction in each generation determines the strength of selection and, in all but one experiment, 16% mortality was used. Truncation selection was used here because this type of selection has been demonstrated (Kimura & Crow, 1978; Crow & Kimura, 1979; Kondrashov, 1984) to be the most effective procedure for bringing about directional gene-frequency changes.

Our most drastic modification of the GAK is to make the GA compatible with Wright's shifting balance model of evolution (e.g. Wright, 1929, and later) as follows.

### **The Search Problem and Wright's Shifting Balance Model**

Deterministic hill-climbing is a method commonly used in optimization procedures. This term is derived from the use of the gradient (i.e. the slope or derivative) to locate the optimum. A positive (uphill) slope indicates progress toward the optimum; a negative (downhill) slope indicates a retreat from the optimum. If movement in any direction results in a negative slope then the optimum has been reached. A hill-climbing algorithm, having discovered an optimum, cannot distinguish whether what has been achieved is a global optimum or a local optimum. Thus, hill-climbing optimization procedures are only appropriate when the solution set is known to contain a single peak.

Unfortunately, many problems in nature may contain multiple peaks (see, e.g. Schaffer & Rosenzweig, 1977; Price *et al.*, 1984; Schuster, 1989), resembling a surface of peaks and valleys. Wright (1969, and earlier) used the metaphor of an adaptive landscape to describe a hypersurface where relative fitness is plotted as a function of the various gene frequencies (see also, Turner, 1971; Conrad, 1979; Provine, 1986). A population subject to selection, like hill-climbing optimization, is attracted towards the nearest peak in the adaptive landscape. This need not be the highest. The population eventually becomes stuck on a peak, through local adaptation. Wright considered what might cause a population to move away from a locally stable equilibrium (*viz* peak), and descend into the surrounding valleys and so reach the domain of attraction of another, higher peak. In particular, Wright studied the consequences of random changes in allele frequencies caused by genetic drift believing it to be an important force in evolution. Fisher (1930) did not agree with Wright

"The views of Fisher and Wright contrast strongly on the significance of random changes in the population. Whereas, to Fisher, random change is essentially noise in the system that renders the deterministic processes somewhat less efficient than they would otherwise be, Wright thinks of such random fluctuations as one aspect whereby evolutionary novelty can come about by permitting novel gene combinations." (Crow & Kimura, 1970: 322)

Whether random genetic drift is a way of creating favourable gene combinations, or is useless, evidence suggests that it is prevalent in nature (Kimura, 1968; King & Jukes, 1969; Crow, 1969). Counterbalancing the rewards gained from random

genetic drift is the loss in mean fitness as the population drifts away from the highest peak. But Wright (1970) argued that this is a necessary cost for letting evolution find new gene combinations.

To enable populations to move from local to global optima, Wright introduced his shifting balance model. The details of the model's processes can be outlined as follows. The population is partitioned into small and somewhat isolated subpopulations. These subpopulations undergo largely random differentiation due to random genetic drift since the effect of genetic drift is greater in small and isolated populations. Now and then a subpopulation acquires a genotype which possesses superior fitness, whereupon mass selection overpowers genetic drift and fixes this genotype in the sub-population. In other words the subpopulation crosses a saddle to another peak. Because of its increased fitness the subpopulation then disperses greater numbers of emigrants than its neighbours thereby attracting surrounding subpopulations to follow across the valley to its (currently) higher peak. This process of "intergroup selection" is repeated, in an ever widening circle to draw the entire population to this peak. In this manner there is a chance that virtually the whole of the adaptive landscape can be explored by continual shifting of control by one adaptive peak to control by a superior one until the global optimum is found.

### **The SBGA Structure**

Motivated by Wright's "continent-island" model, we modify the GAK so that (i) the population is divided into semi-isolated subpopulations and (ii) there is migration between the subpopulations. The subpopulations are referred to as demes, and constitute a locally interbreeding group. Our simulations usually used a deme size of six individuals, or, since the model is haploid, six chromosomes.

Migration between demes provides for the constitution of a single metapopulation. The form and amount of migration in actual populations is important. Our model is an "island" (see Felsenstein, 1976 for definition) model in which migrants are drawn from a random sample of the entire population. In nature, migrants may come primarily from nearby populations. To the extent that nearby populations will have similar allele frequencies, the effects of migration will be smaller than predicted by the island model. The amount of migration sets the limit to how much genetic divergence can occur through random genetic drift within a deme. A high migration rate negates the effect of genetic drift, and the population behaves as a single large unit, its deme structure being irrelevant. Low migration rates increase the strength of local genetic drift but this may (a) obstruct progress to new found peaks, and (b) prevent a deme on a near peak "infecting" others near it with its discovery. It is clear that migration level needs to be carefully chosen. Some guidance about how to achieve this balance can be obtained from studying the quantitative relationship between migration and drift.

The work of Crow & Kimura (1970) following Wright suggests that if the number of migrants is much more than one per deme per generation, there is little local differentiation. In fact remarkably little migration is required to prevent significant genetic divergence among subpopulations due to genetic drift. In line with this

result we usually use a mean migration rate of one migrant per deme every two generations.

Wright's theory also requires (in the absence of deme extinctions) greater numbers of emigrants from subpopulations at higher peaks (Wright, 1945, 1988) this being the expression of interdeme advantage. The SBGA emulated this feature in the following way. Migration occurred every second generation. The total number of migrants in the population was equal to the number of demes. Emigrants were selected from the population until the required number of migrants was obtained. The probability of an emigrant being drawn from a given subpopulation was proportional to that subpopulation's fitness.

The interplay between migration and random genetic drift in the absence of selection was investigated with a one locus two allele SBGA to check both the programming and the correspondence of the discrete model to the limiting continuous distribution of Wright. Small differences are expected for two reasons: the discrete situation vs. the idealized continuous distribution, and the finite size of our population. Figure 2 displays histograms of the equilibrium distribution of allele frequencies that result from three different migration rates with random mating and no selection after 1000 generations. The histograms were compiled by measuring the allele frequency in a population composed of 1000 demes. We can observe from the figure that a low migration rate increases the effect of random genetic drift resulting in fixation and loss of the allele in a large proportion of the population. Intermediate migration rates maintain allelic diversity within demes. High migration rates cause the subpopulations to remain very similar, in effect to fuse into a single population. The figure also shows theoretical curves based on the analyses in Wright (1969) and more specifically on formula 6.15 in Moran (1962) because Moran treats as we do a haploid population. It can be seen that there is very good qualitative agreement between the simulation and theoretical predictions.

### Experiment 1: Optimization in a Multiple-peaked Fitness Landscape

#### REPRESENTATION

The SBGA was studied with a variation of Wright's (1977) *Model with Six Homoeallelic Selective Peaks*. The primary difference is that Wright's (1977) model is diploid whereas the SBGAs model uses haploid chromosomes.

The population consisted of chromosomes with 16 loci. Each locus had two alleles. An allele had either the value of zero (0) or a one (1). This configuration gives  $2^{16}$  possible genotypes. We defined the optimal genotype as a series of eight consecutive zeros followed by eight consecutive ones. A multi-peaked environment was formed using the following algorithm to determine the fitness scores:

- For each chromosome, sum the value,  $v_i$ , at each locus,  $i$ , to give

$$x = \sum_{i=0}^{15} v_i.$$

- Calculate the preliminary fitness function

$$\omega = 64 - (x - 8)^2.$$



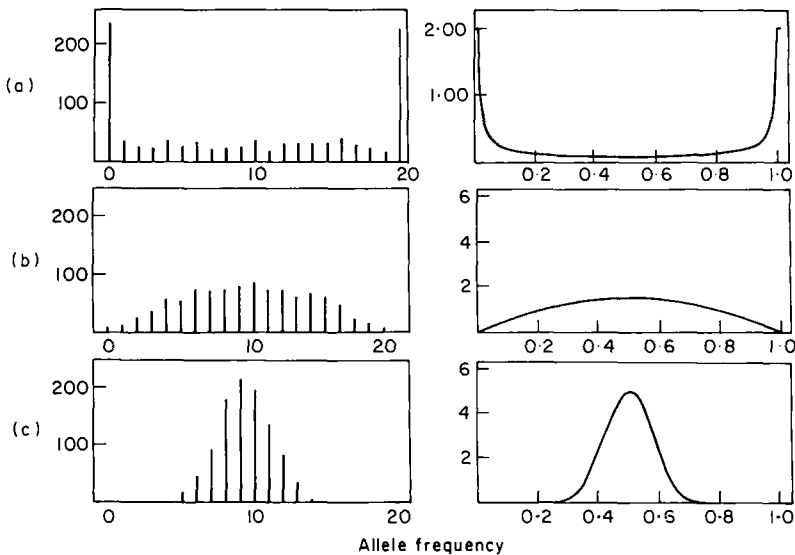


FIG. 2. The equilibrium allelic frequency distributions resulting from simulation drift and migration with the SBGA. The population consists of 20 000 individuals divided into 1000 demes of 20 individuals. Individuals carrying the "A" allele were counted in each deme after 1000 generations and cumulated over the 21 possible counts. Three migration rates are plotted: (a) 0.01, (b) 0.5 and (c) 5 migrants per deme per generation. When migration rate is low, (a), drift is considerable and loss of genetic diversity results from fixation and deletion of the allele (see as accumulations in outer histogram bars). At an intermediate migration rate, (b), the effect of drift is balanced by migration and genetic diversity is preserved, as is seen on the graph, by the wide distribution of the allele frequency in the population. A high migration rate, (c), effectively turns the subpopulations into a single population and the frequency of the allele in the population converges toward the mean value. To the right of each empirical distribution is shown the beta distribution obtained when  $Nm = 0.01$ , 0.5 or 5.0 are applied in formula 6.15 of Moran (1962). Index coefficients of Moran's rather than Wright's are appropriate here because Moran's derivation is for a haploid model like ours. Correspondence between the empirical distributions and their limiting continuous counterparts is good, confirming theory and the programming of our model.

If an optimum is defined as a position in the chromosome state space such that a single step in any direction decreases fitness, the above function specifies  $(8!)^2/16! = 12\,870$  equal optima, which correspond to all chromosomes where numbers of ones and zeros are equal.

- Differentiate peaks to create a global optimum

$$\hat{\omega} = \omega + k \sum_{i=0}^{15} i v_i,$$

where  $k$  is a small constant.

This new fitness function gives moderately higher values to strings with more 1s in the string. The number of peaks produced by this algorithm with 16 loci and  $k$  small remains 12 870, but one, that of eight zeros followed by eight ones, is now a global optimum.

The SBGAs task was to evolve the global optimum genotype. This is a sensible task given a haploid population, but it should be noted that in a diploid population

selection will tend to maximize mean fitness, and this does not necessarily produce any single optimal genotype. For the problem considered there are 65 536 possible genotypes of which 12 869 are local optima. Preliminary studies indicated that the problem was far too simple. Even scaled up versions of the same problem were easily solved. For example the SBGA consistently evolved the optimum genotype within 1000 generations from  $2^{64}$  possible genotypes in an environment with  $1.83 \cdot 10^{18}$  peaks using a population of only 36 chromosomes! Rather than increasing the search space and the number of peaks further, which did not seem to impede the efficiency of the search process, a device inhibiting the building sequences of ones or zeros which could then be recombined, was added to the fitness algorithm to raise the level of difficulty. A sequence of ones flanked at both ends by a zero is called a block of ones. A parameter,  $C$ , giving the length of a standard block is defined. Every block of ones of length  $G$  in the chromosome is given a score

$$b = K(G-1)(G-C)(C-1)^2,$$

where  $K$  is a constant. This score is zero if  $G = 1$  or  $G = C$ , is negative if  $1 < G < C$  and is positive for  $G > C$ . The score,  $b$ , for every sequence of 1s is added to  $\hat{w}$  to give the chromosome fitness. The parameter  $C$  can be imagined as representing the size of the valley that separates the global optimum from most of the local optima—from all those that do not already have blocks larger than  $C$ . Larger values of  $C$  would then represent a wider, more difficult to cross, valley.

To make the search problem difficult, the occurrence of blocks of ones in the initial population was minimized by setting the probability of a "1" being assigned to a chromosomal locus to 0.005. The scoring system used to measure performance was based on the population mean modulus error after 1000 generations. The modulus error lies between 0 (perfect match) and 16 (all bits wrong) with a mean modulus error of eight corresponding to a random choice of zeros and ones. The distribution of mean modulus error that we obtained was bimodal, with modes occurring at 0 and 8 and very few values between 3 and 6. We therefore categorized the data into two discrete classes. A modulus error between 0 and 4.5 inclusive was classed a "hit" and a score greater than 4.5 was classed as a "miss".

The effect of deme structure in the case of a population of total size 72 is explored in Fig. 3, which plots the number of "hits" against deme structure for  $C = 3, 4, 5, 6$  and 7. Taking an arcsine transformation of the proportion of hits, a multiple regression reveals a significant effect of the number of demes ( $P < 0.05$ ,  $df = 11$ ,  $F = 43.59$ ,  $r^2 = 0.9224$ ). As we might expect, the effect of  $C$  is highly significant; as  $C$  increases, fewer hits occur. Although increasing the number of demes makes search more efficient in this case, the effect seems to depend on total population size. Table 1 looks at two examples in which total population size is 72 and two examples which total population size is 36. Subdividing the population of size 36, Table 1(a), does not significantly increase the number of "hits". Increasing the population size from 36 to 72 chromosomes, Table 1(b), resulted in significant differences in the scores obtained from an undivided and divided population when  $C = 4$  ( $\chi^2 = 7.44$ ,  $0.001 < P < 0.01$ ),  $C = 5$  ( $\chi^2 = 6.27$ ,  $0.01 < P < 0.05$ ) and  $C = 6$  ( $\chi^2 = 8.498$ ,  $0.001 < P < 0.01$ ).

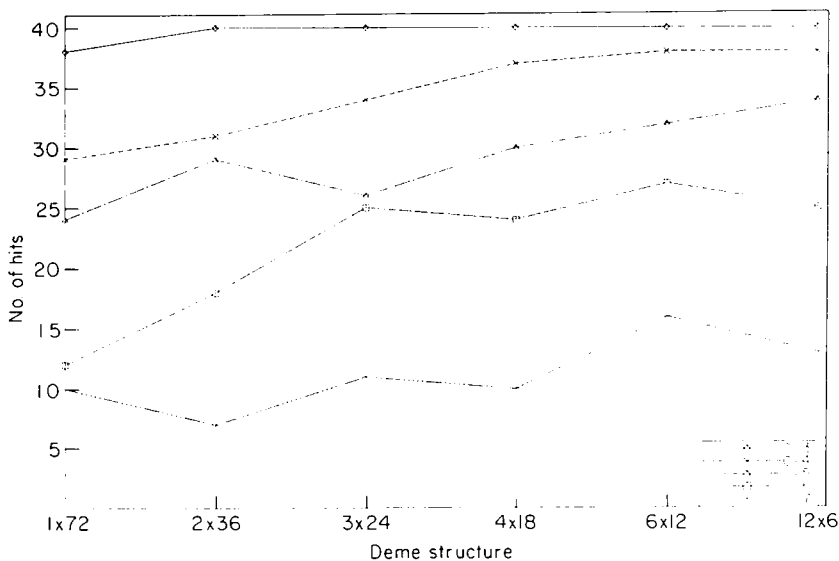


FIG. 3. The effect of varying population structure on performance. A population of 72 individuals was subdivided into one deme of 72, two demes of 36, three demes of 24, four demes of 18, six demes of 12 and 12 demes of 6. The performance, number of "hits", of each population structure is plotted for  $C=3, 4, 5, 6$  and  $7$ . The improvement in performance as the population is partitioned into more numerous demes is statistically significant (see text).

In only one comparison out of the ten does the undivided population score more "hits". However, rather surprisingly that case was more severe than average and overall the difference was not greatest when the problem set is of maximum severity: in neither case was it significant for  $C=7$ .

The overall advantages of both large and subdivided populations, however, are clear. Cumulating the "hit" scores across all values of  $C$ , for a population size of 36, reveals that the overall success of the subdivided population ( $=110$ ) is greater than the undivided population ( $=99$ ). Increasing the population size to 72 results in an increased number of "hits" for both the subdivided ( $=150$ ) and the undivided population ( $=113$ ). But the magnitude of improvement is greater for the subdivided population ( $+40$  hits) as compared to the undivided population ( $+14$  hits). Thus, a larger population size improves performance but the degree of improvement is greater when the population is subdivided.

Tables 2, 3 and 4 investigate the effects of various biological features on the number of "hits" obtained. In each case the population is divided into 12 demes of six individuals, and each datum is based on 40 runs of 1000 generations.

Migrations between subpopulations were proportional to fitness. In each analysis, the proportion of "hits" was subjected to an arcsine transformation in order to normalize the variance.

In Table 2 we investigate the effect of migration frequency. A multiple regression based on  $C^2$  and the number of generations between migrations showed that both these factors were significant ( $P < 0.0005$ ,  $df. = 16$ ,  $F = 44.82$ ,  $r^2 = 0.89$ ).

TABLE 1

*Comparison between performances of subdivided and undivided populations in a multi-peaked environment for various lengths of penalty block (C). Migrations between the subpopulations were proportional to fitness (see text) and the rate of migration was 0.5 migrants per deme per generation. Rank-order soft selection was used for both types of population structure with the mortality rate set to 16% of the population. Each run lasted 1000 generations. The other parameters were  $K=0.8$ ,  $v=0.5$ , mutation = 0.005, probability of crossover = 0.5. Results from a  $2 \times 2$   $\chi^2$  contingency test is given in the far-right column. Significance is indicated by a star (\*)*

(a) Total population size was 36 organized as either a single population of 36 individuals or six subpopulations of six individuals

C	6 × 6		1 × 36		$\chi^2$ (significance)
	Hit	Miss	Hit	Miss	
3	34	6	33	7	0.092 N.S.
4	30	10	23	17	2.739 N.S.
5	24	16	18	22	1.8 N.S.
6	15	25	19	21	0.818 N.S.
7	7	33	6	34	0.092 N.S.

(b) Total population size was 72, organized as either a single population of 72 individuals or 12 sub-populations of six individuals

C	12 × 6		1 × 72		$\chi^2$ (significance)
	Hit	Miss	Hit	Miss	
3	40	0	38	2	2.051 (N.S.)
4	38	2	29	11	7.44 (0.001 < $P$ < 0.01)*
5	34	6	24	16	6.27 (0.01 < $P$ < 0.05)*
6	25	15	12	28	8.498 (0.001 < $P$ < 0.01)*
7	13	27	10	30	0.549 (N.S.)

N.S. = Not significant.

Table 3 shows a similar analysis based on selection strength, expressed as percentage mortality. The effect of  $C$  is significant ( $P < 0.005$ ,  $df. = 11$ ,  $F = 43.59$ ,  $r^2 = 0.92$ ) but the effect of mortality (i.e. selection strength) is not.

The final example (Table 4) shows that mutation rate and  $C$  both have a significant effect ( $P < 0.0005$ ,  $df. = 11$ ,  $F = 57.22$ ,  $r^2 = 0.94$ ).

The examples given above illustrates that, at least in some circumstances, the SBGA is able to find the global optimum more efficiently than an algorithm based on a single population. To look at a specifically biological problem, we now consider a second example.

### Experiment 2: Daily Routines of Singing and Foraging—the Evolution of a Dawn Chorus

In this section we describe an application of the SBGA to the problem of how a male bird should allocate his time between singing and foraging. The SBGA evolves

TABLE 2

*Performances of a subdivided population (12 demes of six individuals) under different migration frequencies for various lengths of penalty block (C). The migration frequencies were varied from four generations to 32 generations between migrations. Rank-order soft selection was used with the mortality rate set to 16% of the population. The other parameters were  $K = 0.8$ ,  $v = 0.5$ , mutation = 0.005, probability of crossover = 0.5*

Generations between migrations	C	Hit	Miss
4	3	40	0
	4	39	1
	5	34	6
	6	30	10
	7	22	18
8	3	40	0
	4	33	7
	5	29	11
	6	21	19
	7	8	32
16	3	37	3
	4	16	24
	5	14	26
	6	7	33
	7	2	38
32	3	23	17
	4	1	39
	5	1	39
	6	.2	38
	7	0	40

policies which are used during the course of a day to decide whether to sing or feed based upon energy reserves. A high level of singing at dawn is referred to as the dawn chorus (see Kacelnik & Krebs, 1982; Krebs, 1977; Mace, 1986, for discussion). From previous work (McNamara *et al.*, 1987), the optimum policy is known: a dawn chorus should evolve when overnight energy expenditure is variable.

#### THE MODEL AND REPRESENTATION

The model is the same as that developed by McNamara *et al.* (1987). The male is characterized by his level of energy reserves,  $x$ . If  $x$  falls to zero, then the male dies. McNamara *et al.* (1987) considered a male singing over a period of 5 days. Each day is divided into a period of daylight and a period of darkness. There are 24 intervals during the daylight period. At the start of each interval, the male can decide either to sing during the interval or to forage during the interval. If the male sings, then it uses energy (thus decreasing its reserves) but has a probability of attracting a mate. If the male forages, then it also uses energy, but it obtains a

TABLE 3

*Performances of a subdivided population (12 demes of six individuals) under varying levels of mortality. Rank-order soft selection was used with mortality rates set to 33, 50 and 67% of the population. The other parameters were  $K = 0.8$ ,  $v = 0.5$ , mutation = 0.005, probability of crossover = 0.5*

<i>C</i>	Mortality (%)	Hit	Miss
3	33	40	0
	50	40	0
	67	40	0
4	33	33	7
	50	34	6
	67	28	12
5	33	27	13
	50	32	8
	67	30	10
6	33	17	23
	50	18	22
	67	19	21
7	33	11	29
	50	12	28
	67	16	24

TABLE 4

*Performance of a subdivided (12 demes of six individuals) population under different mutation rates. Rank-order soft selection was used with mortality rates set to 16% of the population. The other parameters were  $K = 0.8$ ,  $v = 0.5$  probability of crossover = 0.5*

Mutation rate	<i>C</i>	Hit	Miss
0.01	3	40	0
	4	40	0
	5	38	2
	6	32	8
	7	15	25
0.001	3	28	12
	4	16	24
	5	17	23
	6	18	22
	7	10	30

random amount of food (the amount may be zero). The parameter  $IP$  controls the variability of the food by determining the probability that the bird is interrupted while feeding. During the night the male uses a normally distributed amount of energy with mean  $\mu_N$  and standard deviation  $\sigma_N$  (for further details see McNamara *et al.*, 1987).

To find the policy of singing and foraging that maximizes a male's fitness, McNamara *et al.* (1987) introduced a terminal reward at the end of the fifth day that specifies the male's expected future reproductive success as a function of its state (see McNamara & Houston, 1986 for a detailed discussion of this approach). The reward is zero if the male is dead, one if the male is alive without a mate and two if the male is alive and has obtained a mate. A sensitivity analysis showed that the exact values of this reward function were not crucial. Using dynamic programming (Bellman, 1957) it was found that at time  $t$  on day  $n$  it was optimal to sing if reserves were above  $C^*(t, n)$  and to forage if reserves were below  $C^*(t, n)$ . Apart from the last day ( $n = 5$ ), the policies were virtually independent of  $n$ . We will denote this day-independent critical level by  $C^*(t)$ .

In the investigation of the GA, all animals start at dawn on day 1 with  $x = 8$ . Each bird then follows the policy given by its chromosome until dusk at the end of the second day. We calculate the probability distributions of energy states at the following dawn under this policy. This series of calculations is referred to as the forward procedure. The animal's state at this dawn determines the animal's reward in the following way. Animals which are dead at this time receive no reward. Animals that are alive and have a mate receive a reward of 2. Animals that are alive without a mate receive a reward equal to the mean reward they would have obtained from following the optimal policy for the three remaining days of a 5 day period, as given by McNamara *et al.* (1987). For each chromosome we calculate the expected total reward obtained and equate this to the fitness of the chromosome.

The chromosome determines behaviour as a function of energy reserves and time of day. The optimality analysis carried out by McNamara *et al.* (1987) indicates that the optimal policy always involves a critical level above which the bird should sing and below which it should forage. Thus out of the enormously large range of possible policies, we restrict attention to policies of this form: i.e. a chromosome will specify a critical level  $C(t)$ .

Within this class of policies, there are various ways to code for  $C(t)$ . The results given below are based on the following coding. Each chromosome has six loci, labelled  $i = 0$  to  $i = 5$ . Locus  $i$  contains an integer  $c_i$  between 0 and  $C_{\max}$  ( $C_{\max} = 32$  in all computations reported here). The integers  $c_i$  determine a function  $C(t)$  as follows

$$C(24i/5) = c_i, \quad i = 0, \dots, 5,$$

$C(t)$  at other values is given by linear interpolation.

The male's behaviour is taken to be the same on each day of a 2-day period and is given by the rule forage if  $x(t) < C(t)$  and sing otherwise.

The values at a given locus only influences the policy around that time of day. It is important to note, however, that the fitness consequences of the value at a

given locus depend on the values at all other loci. In other words, there are strong epistatic effects. One of the main points made by McNamara *et al.* (1987) is that the timing of singing cannot necessarily be understood by considering the environmental conditions at the precise time when singing occurs. One may only be able to understand the timing of singing by considering the bird's daily routine as a whole.

We also tried a different form of coding based on a polynomial equation for  $C(t)$ . The equation was of the form  $C(t) = k_0 + k_1t + k_2t^2 + k_3t^3 + k_4t^4$ . There were five loci with locus  $i$  coding for  $k_i$ . This worked less well than the coding described above. A possible reason for this is that the value at each loci influences the policy at all times of the day. Another possibility is that the optimal critical level  $C^*(t)$  is essentially a piece-wise linear function of  $t$  and there are limits to the accuracy with which a fourth-order polynomial can approximate a piece-wise linear function. The polynomial coding may thus have imposed considerable constraints on the set of possible phenotypes.

A population of 36 chromosomes was partitioned into six demes of six chromosomes. The following SBGA parameter settings were used for all runs. The probability of mutation was set to 0.01. When a mutation occurred at a locus, it resulted in the assignment of a random integer between 0 and  $C_{\max}$  to that locus. Migration rate was 0.5 migrants per deme per generation. Selection within demes was by mortality with two deaths per deme per generation. The interlocus crossover probability during reproduction was set to 0.5.

The chromosomes were initialized by setting each locus to a random integer between 0 and  $C_{\max}$ . Preliminary runs indicated that fixation of the chromosomes occurred at  $\approx 100$  generations. After fixation only mutation can alter the chromosomal structure. Since we were primarily interested in selection effects, runs were halted after 100 generations and the evolved policies were saved.

## Results

### NO OVERNIGHT LOSS

We begin by looking at the case in which there is no overnight energy expenditure, i.e.  $\mu_N$  and  $\sigma_N$  are zero. In these circumstances the optimal policy does not depend on time of day, and is given by the rule "sing if reserves are above  $C^*$ , otherwise forage".

The evolved policies resulting from six typical runs of the SBGA are plotted in Fig. 4. It can be seen that the fit between the evolved policies and the optimum is very close. Figure 5 shows the effect of increasing the interruption probability ( $IP$ ). Increasing  $IP$  makes the food supply less reliable, and hence the animal should carry more reserves, i.e.  $C^*$  is increased. It can again be seen that the fit between the evolved policies and the optimum policy is very close.

When there is no overnight loss, the proportion of unmated birds that are singing is constant throughout the day under the optimal policy. The behaviour that results from the SBGA policies evolves to this pattern in both cases that we have considered.



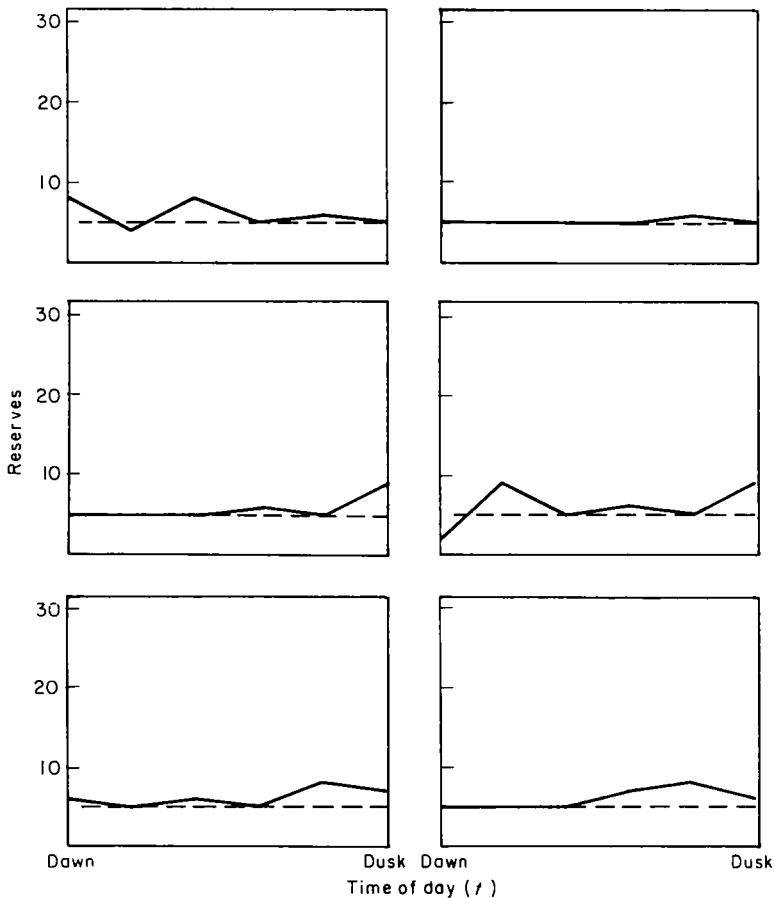


FIG. 4. Evolved policies for singing vs. foraging after 100 generations from six runs of the SBGA. The evolved policy  $C(t)$  is given by a solid line and the optimal policy  $C^*$  (which is the same for each graph) is given by the broken line. The parameter values were:  $\mu_N = 0$ ;  $\sigma_N = 0$ ; and  $IP = 0.0$ . The SBGA parameters were: mutation rate = 0.01; crossover probability = 0.5; migration rate = 0.5 migrants per deme per generation. A population of 36 chromosomes was divided into six demes of six chromosomes. Mortality selection, with two deaths per deme per generation, was used for all runs.

#### OVERNIGHT LOSS

When a variable amount of energy is expended overnight, an animal must build up its reserves before dusk. Consequently, the optimal policy involves a threshold  $C^*(t)$  which increases as dusk approaches. Figure 6 compares six typical evolved policies with the optimal policy. Figure 7 shows the progressive evolution of singing behaviour over 100 generations, as well as the singing behaviour that results from following the optimal policy. It can be seen that the optimal routine involves a dawn chorus, a mid-afternoon dip in singing followed by a rise in singing at dusk (cf. McNamara *et al.*, 1987). After 100 generations, the evolved routine has all of these features.

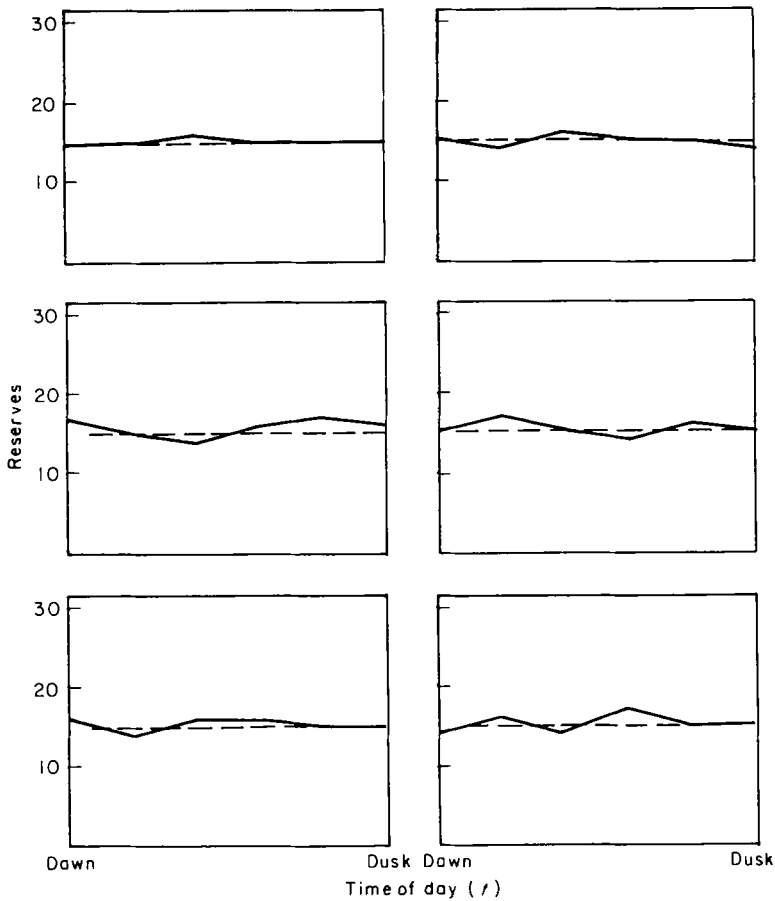


FIG. 5. Evolved policies after 100 generations from six runs of SBGA. The evolved policy  $C(t)$  is given by a solid line and the optimal policy  $C^*$  (which is the same for each graph) is given by the broken line. The parameter values were:  $\mu_N = 0$ ;  $\sigma_N = 0$ ; and  $IP = 0.4$ . The SBGA parameters are as for Fig. 4.

Increasing the interruption probability when there is overnight expenditure means that the animal must not only carry higher reserves but start to build up reserves earlier in the day. Figure 8 shows the optimal policy together with six typical evolved policies and Fig. 9 shows the evolution of singing behaviour over 100 generations together with the singing behaviour under the optimal policy. It can again be seen that the evolved daily routine of singing has the same qualitative form as the optimal singing routine. The policies evolved by the GA achieved close to optimal fitness as can be seen from Table 5.

### Discussion

Viewed as an optimization procedure, GAs have certain advantages over usual methods of optimization. Holland (1975) and De Jong (1980) argue that GAs

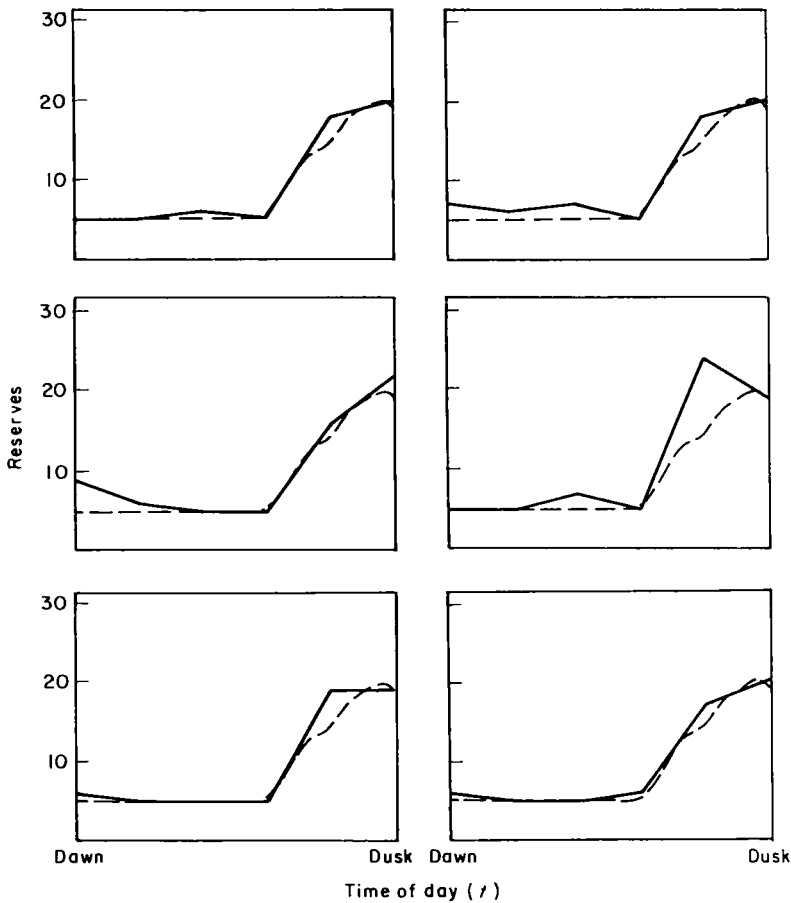


FIG. 6. Evolved policies after 100 generations from six runs of SBGA. The evolved policy  $C^*(t)$  is given by a solid line and the optimal policy  $C^*$  (which is the same for each graph) is given by the broken line. The parameter values were:  $\mu_N = 12$ ;  $\sigma = 4$ ; and  $IP = 0.0$ . The SBGA parameters are as for Fig. 4.

perform better than many standard optimization procedures when the payoff function is discontinuous or has multiple peaks. GAs are flexible in that using the same encoding, they can track changes in the payoff function. Holland (1975) saw this as a necessary feature of an adaptive search procedure. The GAs have their basis in biology, and one can use biological insight to improve their performance. Motivated by Wright's ideas, we have introduced a modified GA in which the population is divided into semi-isolated demes. We have shown that there are circumstances in which this SBGA performs better than a single population GA. Further work needs to be done to investigate the general effect of this modification and the other modifications that we have introduced (e.g. non-overlapping generations, crossover frequency and proportional vs. non-proportional migration) on the search efficiency of GAs.

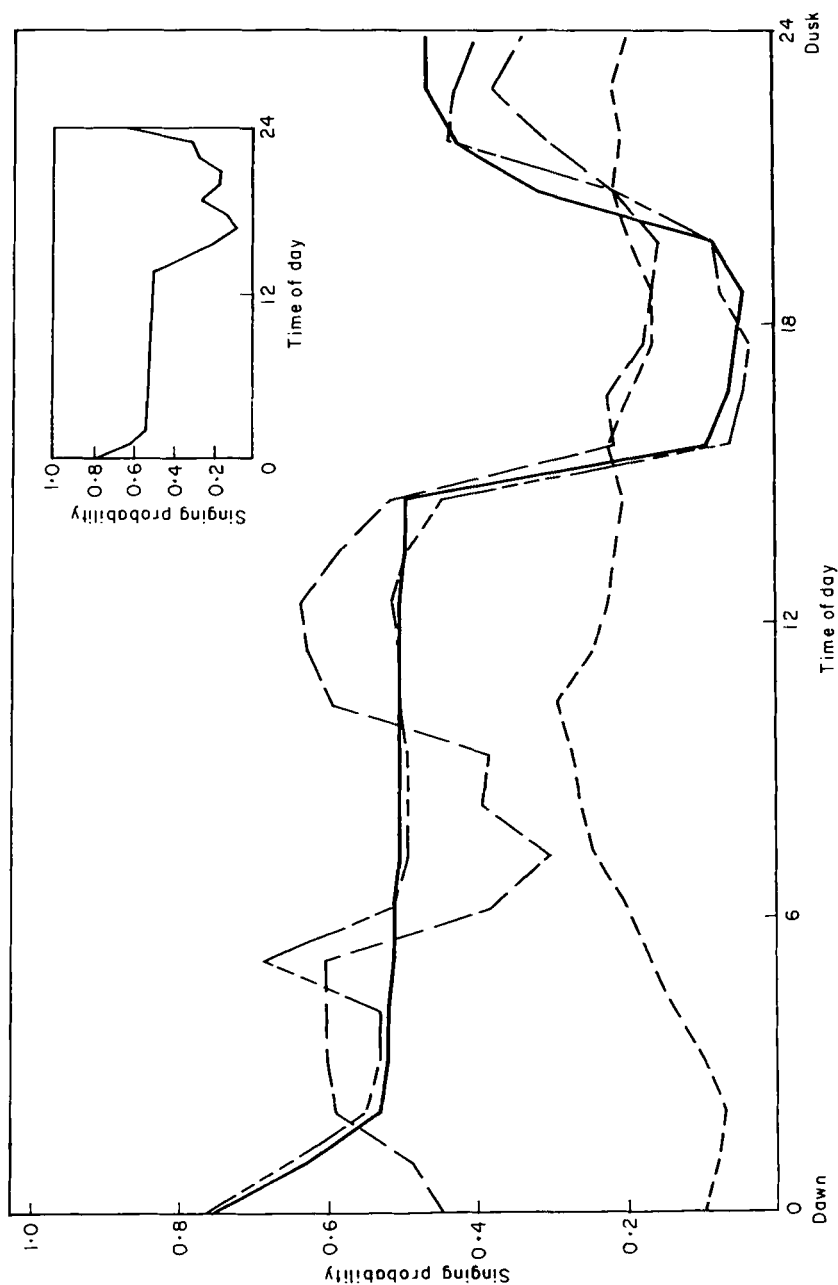


FIG. 7. Dynamics of the evolution of behaviour. The ordinate plots the behaviour, as probability of singing, and the abscissa gives the time of day. Behaviour, calculated using the forward procedure, was plotted after 1 (---); 20 (-.-.); 50 (---); and 100 (—) generations of evolution. Environmental parameters were  $\mu_N = 12$ ,  $\sigma_N = 4$  and  $IP = 0.0$ . The inset shows the routine of singing under the optimal policy.

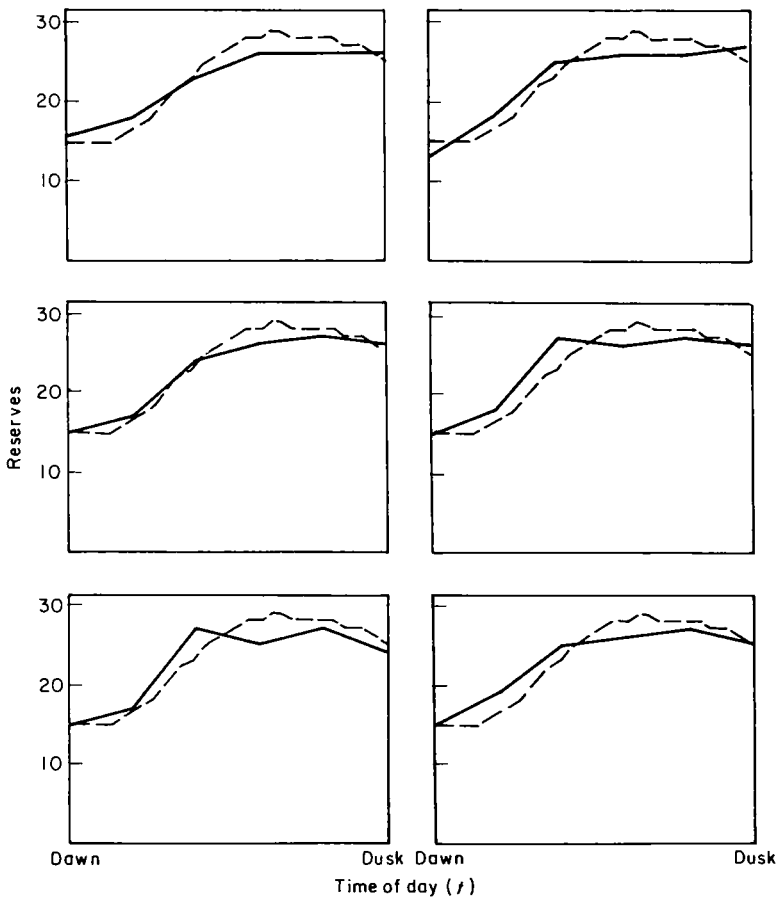


FIG. 8. Evolved policies after 100 generations from six runs of SBGA. The evolved  $C(t)$  is given by a solid line and the optimal policy  $C^*(t)$  (which is the same for each graph) is given by the broken line. The parameter values:  $\mu_N = 12$ ;  $\sigma_N = 4$ ; and  $IP = 0.4$ . The SBGA parameters are as for Fig. 4.

We applied the SBGA to cases in which dynamic programming had already been used (McNamara *et al.*, 1987). The SBGA found near-optimal solutions within a short period of time (within 100 generations). This performance is impressive compared to a procedure based on exhaustively searching the  $32^6 = 1\,073\,741\,824$  possible chromosomes. But in this one-dimensional case, dynamic programming is a more efficient procedure. The computational costs of dynamic programming rise exponentially as the dimensionality of a problem increases ("the curse of dimensionality", Bellman, 1957), whereas for the GA, an increase in the dimensionality of the problem need not substantially increase the computational costs. It follows that GAs are likely to be far more efficient than dynamic programming in high-dimensional problems.

As well as using GAs as efficient search procedures, they can also be used to shed light on evolution in biological systems. Not only do they provide a convenient way

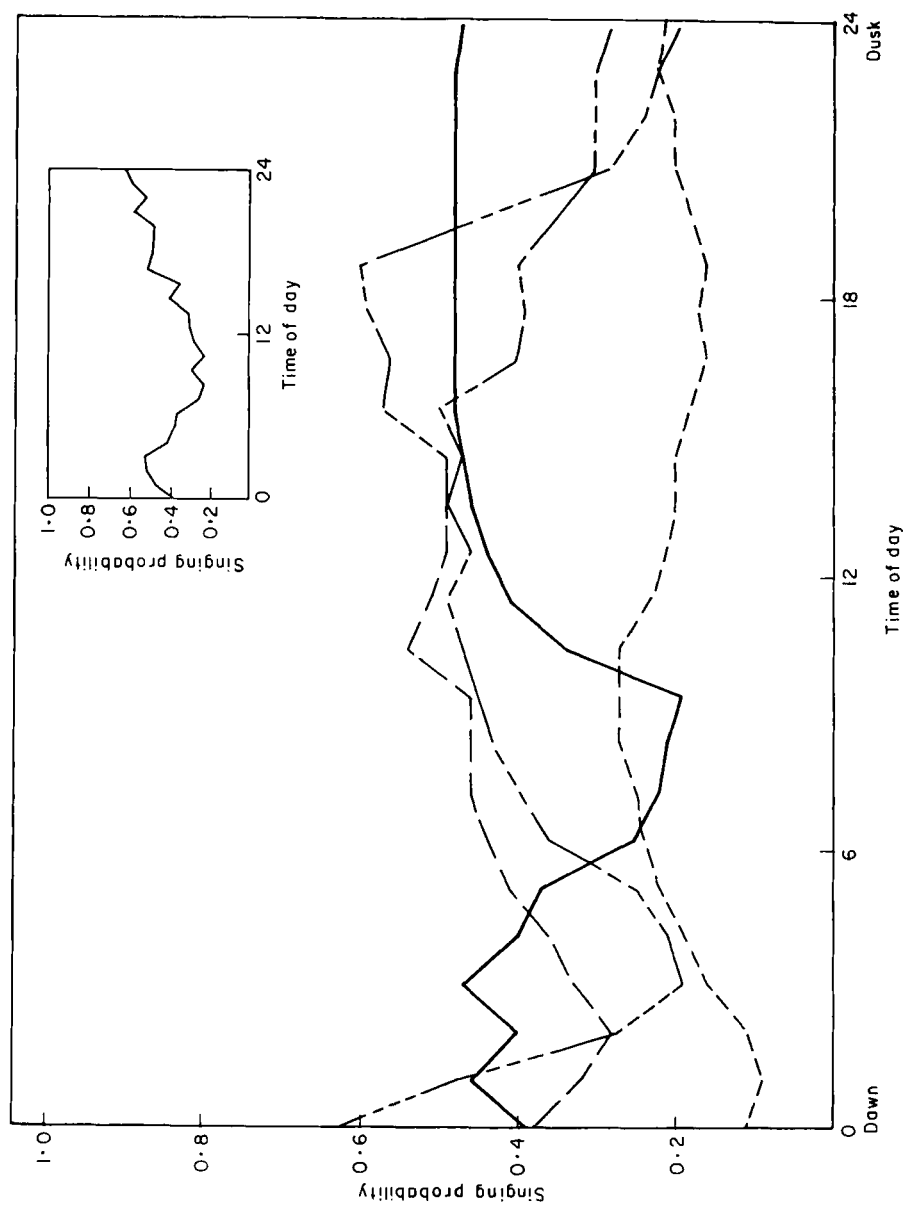


FIG. 9. Dynamics of the evolution of behaviour. The ordinate plots the behaviour, as probability of singing, and the abscissa gives the time of day. Behaviour, calculated using the forward procedure, was plotted after 1 (—), 20 (---), 50 (---), and 100 (— · —) generations of evolution. Environmental parameters were  $\mu_N = 12$ ,  $\sigma_N = 4$  and  $IP = 0.4$ . The inset shows the routine of singing under the optimal policy. Key as for Fig. 7.

TABLE 5

*Evolved policies from replicate runs along with their final fitnesses under four environmental conditions. Genetic algorithm parameters for all runs are: mutation probability = 0.01, mortality = 2, deme number = 6, local population = 6, migrants per generation = 0.5, crossover probability = 0.5 and number of generations = 100*

(a) Condition: $\mu_N = 0$ , $\sigma = 0$ and $IP = 0.0$							
Optimal fitness	Evolved fitness		Evolved policy				
1.5180	1.5178	8	4	8	5	6	5
	1.5180	5	5	5	5	6	5
	1.5179	5	5	5	6	5	9
	1.5178	2	9	5	6	5	9
	1.5178	6	5	6	5	8	7
	1.5178	5	5	5	7	8	6
(b) Condition: $\mu_N = 0$ , $\sigma = 0$ and $IP = 0.4$							
Optimal fitness	Evolved fitness		Evolved policy				
1.5100	1.5100	15	15	16	15	15	15
	1.5100	15	14	16	15	15	14
	1.5100	17	15	14	16	17	16
	1.5100	15	17	15	14	16	15
	1.5100	16	14	16	16	15	15
	1.5100	14	16	14	17	15	15
(c) Condition: $\mu_N = 12$ , $\sigma = 4$ and $IP = 0.0$							
Optimal fitness	Evolved fitness		Evolved policy				
1.4488	1.4486	7	6	7	5	18	20
	1.4487	5	5	6	5	18	20
	1.4486	5	5	7	5	24	19
	1.4485	9	6	5	5	16	22
	1.4487	6	5	5	6	17	20
	1.4487	6	5	5	5	19	19
(d) Condition: $\mu_N = 12$ , $\sigma = 4$ and $IP = 0.4$							
Optimal fitness	Evolved fitness		Evolved policy				
1.4367	1.4364	13	18	25	26	26	27
	1.4364	16	18	23	26	26	26
	1.4364	15	18	27	26	27	26
	1.4364	15	17	24	26	27	26
	1.4364	15	19	25	26	27	25
	1.4360	15	17	27	25	27	24

to find an optimum, but they also make it plausible to assume that evolution could reach this optimum too. It must be remembered, however, that what the GA can evolve is constrained by the scheme of encoding. On the one hand, this can be seen as a disadvantage, in that the optimum is dependent on the genetic system. On the other hand, it can be seen as getting around a criticism of optimality techniques in biology. It has been pointed out that finding optimal solutions ignores genetic constraints; GAs make it possible to investigate the effects of genetic constraints on optimal solutions.

The GAs provide a natural setting for investigating evolution in systems with epistatic interactions between genes. Such interactions are present in both the examples discussed in this paper. In the case of singing and foraging, there are epistatic effects because the fitness consequences of an action at one time of day depend on the actions performed at other times of day.

At a higher level the GA allows one to explore the efficiency of evolution by natural selection as a search mechanism, and in particular to discover which features make the search effective and which features are detrimental (Grefenstette, 1986).

We now give some examples of potential applications of GAs in biology.

(i) Game theoretic interactions between animals. Houston & McNamara (1987) modified the model of McNamara *et al.* (1987) to include interactions between singing males. By considering an infinite population limit, they were able to restrict attention to policies that depended only on a male's level of reserves and time of day. Even so, this dynamic game was computationally much more expensive than the optimization problem faced by a single male. Any reasonable model based on a finite population would have to allow a male's behaviour to depend on the previous behaviour of the other males in the population. Incorporating additional state variables representing this information would render an approach based on dynamic programming completely impractical. One could, however, use GAs to encode the way in which a male responds to the previous behaviour of other males and hence, analyse the action of evolution on such systems.

(ii) Learning rules. There has been considerable discussion of rules which are efficient at exploiting a particular environment (Green, 1980; Iwasa *et al.*, 1981; McNamara, 1982; McNamara & Houston, 1987). Such analyses usually assume that the parameters of the environment are known to the animal. Many animals are constantly exposed to new environments, and so a full account of behaviour needs to consider a higher level of rule which is capable of both learning about, and exploiting a range of *a priori* unknown environments. The GAs provide an obvious and powerful framework for investigating the evolution of such learning rules.

The GAs were originally developed as an artificial intelligence technique. A simple view of evolution by natural selection was used to provide an efficient search procedure. We have now come full circle, where what has been learnt about the GA can be used, in a defined type of strategy-analytic simulation, to improve our understanding of the evolution of natural populations.

B. H. Sumida was supported by S.E.R.C.; A. I. Houston was supported by S.E.R.C., N.E.R.C. and King's College, Cambridge. We thank the following for helpful comments and



criticisms: Andy Barto, Sean Nee and Andrew Pomiankowski. Joy Bergelson provided advice on statistical analysis.

## REFERENCES

- ALEXANDER, R. MCN. (1982). *Optima for Animals*. London: Edward Arnold.
- AXELROD, R. (1987). In: *Genetic Algorithms and Simulated Annealing* (Davis, L., ed.) London: Pitman.
- BELLMAN, R. (1957). *Dynamic Programming*. New Jersey: Princeton University Press.
- BORSTNIK, B., PUMPERNIK, D. & HOFACKER, G. L. (1987). *J. theor. Biol.* **125**, 249–268.
- BOUNDS, D. G. (1987). *Nature, Lond.* **329**, 215–219.
- BRADY, R. M. (1985). *Nature, Lond.* **317**, 804–806.
- BREMNERMAN, H. J. (1962). In: *Self-Organizing Systems* (Yovits, J. & Goldstein, G. D., eds) Washington, D.C.: Spartan Books.
- CONRAD, M. (1979). *Biosystems* **11**, 167–182.
- CROW, J. F. (1969). In: *Proceedings of the XII International Congress on Genetics* **3**, 105–113.
- CROW, J. F. & KIMURA, M. (1970). *An Introduction to Population Genetics*. Minneapolis: Burgess.
- CROW, J. F. & KIMURA, M. (1979). *Proc. natn. Acad. Sci. U.S.A.* **76**, 396–399.
- DAVIS, L. (1987). *Genetic Algorithms and Simulated Annealing*. London: Pitman.
- DE JONG, K. (1980). *IEEE Trans. on Sys. Man and Cyber.* **9**, 566–574.
- EIGEN, M. & SCHUSTER, P. (1979). *The Hypercycle—A Principle of Natural Self-organization*. Berlin: Springer-Verlag.
- FEISENSTEIN, J. (1976). *Ann. Rev. Genet.* **10**, 253–280.
- FISHER, R. A. (1930). *The Genetical Theory of Natural Selection*. Oxford: Clarendon Press, 2nd edn. 1958: Dover Publications.
- FRASER, A. S. (1962). *J. theor. Biol.* **2**, 329–346.
- FRASER, A. S. (1962). *J. theor. Biol.* **2**, 329–346.
- GALAR, R. (1989). *Biol. Cybern.* **60**, 357–364.
- GOEL, N. S. & THOMPSON, R. L. (1986). *Int. Rev. Cytol.* **103**, 1.
- GOLDBERG, D. S. (1989). *Genetic Algorithms*. Massachusetts: Addison-Wesley.
- GOULD, S. J. & LEWONTIN, R. C. (1979). *Proc. R. Soc., Lond. B*, **205**, 581–598.
- GREEN, R. F. (1980). *Theor. pop. Biol.* **18**, 244–256.
- GREFENSTETTE, J. J. (1986). *IEEE Trans. on Sys. Man and Cybern.* **16**, 122–128.
- HOLLAND, J. (1975). *Adaptation in Natural and Artificial Systems*. Ann Harbor: University of Michigan Press.
- HOUSTON, A. I. & McNAMARA, J. M. (1987). *J. theor. Biol.* **129**, 57–68.
- IWASA, Y., HIGASHI, M. & YAMAMURA, N. (1981). *Am. Nat.* **117**, 710–723.
- JACQUARD, A. (1974). *The Genetic Structure of Populations*. Berlin: Springer-Verlag.
- KACELNIK, A. & KREBS, J. R. (1982). *Behaviour* **83**, 287–309.
- KIMURA, M. (1968). *Gen. Res.* **11**, 247–269.
- KIMURA, M. & CROW, J. F. (1978). *Proc. natn. Acad. Sci. U.S.A.* **75**, 6168–6171.
- KING, J. L. & JUKES, T. H. (1969). *Science* **164**, 788–798.
- KONDRASHOV, A. S. (1984). *J. theor. Biol.* **107**, 249–260.
- KREBS, J. R. (1977). In: *Evolutionary Ecology* (Stonehouse, B. & Perrins, C. M., eds) London: McMillan Press.
- McNAMARA, J. M. (1982). *Theor. pop. Biol.* **21**, 269–288.
- McNAMARA, J. M. & HOUSTON, A. I. (1986). *Am. Nat.* **127**, 358–378.
- McNAMARA, J. M. & HOUSTON, A. I. (1987). In: *Quantitative Analyses of Behavior* Vol. 6, *Foraging* (Commons, M. L., Kacelnik, A. K. & Shettleworth, S. J., eds) pp. 23–39, New Jersey: Lawrence Erlbaum.
- McNAMARA, J. M., MACE, R. H. & HOUSTON, A. I. (1987). *Behav. ecol. Socio.* **20**, 399–405.
- MACE, R. (1986). *Anim. Behav.* **34**, 621–622.
- MAYNARD SMITH, J. (1978). *Ann. Rev. ecol. Syst.* **9**, 31–56.
- MORAN, P. A. P. (1962). *The Statistical Process of Evolutionary Theory*. Oxford: Clarendon Press.
- PRICE, T. D., GRANT, P. R., GIBBS, H. L. & BOAG, P. T. (1984). *Nature, Lond.* **309**, 787–791.
- PROVINE, W. B. (1986). *Sewall Wright and Evolutionary Biology*. Chicago: University of Chicago Press.
- OSTER, G. F. & WILSON, E. O. (1978). *Caste and Ecology in the Social Insects*. Princeton, NJ: Princeton University Press.
- SCHAFER, W. M. & ROSENZWEIG, M. L. (1977). *Ecology* **58**, 60–72.
- SCHUSTER, P. (1989). Optimization and complexity in molecular biology and physics. In: *Optimal Structures in Heterogenous Reaction Systems* (Plath, P. J., ed.) Berlin: Springer-Verlag.

- TURNER, J. R. G. (1971). *Am. Nat.* **105**, 267-278.
- WALLACE, B. (1975). *Evolution* **29**, 465-473.
- WANG, Q. (1987). *Biol. Cybern.* **57**, 95-101.
- WATSON, J. D., HOPKINS, N. H., ROBERTS, J. W., STEITZ, J. A. & WEINER, A. M. (1987). *Molecular Biology of the Gene*, 4th edn. Menlo Park: Benjamin/Cummings.
- WRIGHT, S. (1929). *Anat. Rec.* **44**, 287.
- WRIGHT, S. (1945). *Ecology* **26**, 415-419.
- WRIGHT, S. (1969). *Evolution and Genetics of Populations*, Vol. 2. *The Theory of Gene Frequencies*. Chicago: University of Chicago Press.
- WRIGHT, S. (1970). In: *Mathematical Topics in Population Genetics* (Kojima, K., ed.) New York: Springer-Verlag.
- WRIGHT, S. (1977). *Evolution and Genetics of Populations*, Vol. 3. *Experimental Results and Evolutionary Deductions*. Chicago: University of Chicago Press.
- WRIGHT, S. (1988). *Am. Nat.* **131**, 115-123.