9. An Overview

Enough of the theoretical framework has now been erected that we can begin to view it as a whole. To this end, the present chapter will discuss three general aspects of the theory. Section 1 will concentrate on those insights offered by the theory which are useful across the full spectrum of adaptive problems. Section 2 provides a synopsis of several computer studies to give the reader an idea of how the overall theory works in particular contexts. Section 3 will outline several difficult long-range problems which fall within the scope of the theory.

1. INSIGHTS

Within the theoretical framework problems of adaptation have been phrased in terms of generating structures of progressively higher performance. Because the framework itself places no constraints on what objects can be taken as structures, other than that it be possible to rank them according to some measure of performance, the resulting theory has considerable latitude. Once adaptation has been characterized along these lines, it is also relatively easy to describe several pervasive, interrelated obstacles to adaptation—obstacles which occur in some combination in all but the most trivial problems:

1. High cardinality of $\mathcal{A}$. The set of potentially interesting structures is extensive, making searches long and storage of relevant data difficult.
2. Apportionment of credit. Knowledge of properties held in common by structures of above-average performance is incomplete, making it difficult to infer from past tests what untested structures are likely to yield above-average performance.
3. High dimensionality of $\mu_{\mathcal{B}}$. Performance is a function of large numbers of variables, making it difficult to use classical optimization methods employing gradients, etc.

4. Nonlinearity of μ_E. The performance measure is nonlinear, exhibiting "false peaks" and making it difficult to avoid concentration of trials in suboptimal regions.
5. Mutual interference of search and exploitation. Exploitation of what is known (generation of structures observed to give above-average performance) interferes with acquisition of new information (generation of new structures) and vice versa.
6. Relevant non-payoff information. The environment provides much information in addition to performance values (payoff), some of which is relevant to improved performance.

The schema concept suggests a coordinated array of robust procedures for meeting these obstacles. The procedures are all founded on the view that each structure is a "carrier" (or selected sample point) of each of the great number of schemata it instances. Because arbitrary structures are easily represented as strings (by using detectors or more sophisticated techniques such as the broadcast language) the resulting procedures apply to adaptation in all its forms. Once schemata have been defined, there is a natural means (p. 69) of comparing structures and apportioning credit by assigning to each schema the average of payoffs to its *observed* instances (compensating obstacle (2)). A small population of structures, when properly selected (pp. 139–40), can then store the relative performance rankings for very large numbers of schemata (compensating obstacle (1)). It is this broad data base vis-à-vis schemata (p. 87) which enables genetic plans to escape false peaks and other difficulties engendered by nonlinearities (compensating obstacle (4)). Recasting the search problem in terms of the space of schemata sidesteps dimensionality effects (obstacle (3)), at least for intrinsically parallel procedures such as genetic plans (p. 71). Under such plans the succession of structures generated from the data base (the current population) induces a highly parallel, diffusion-like spread of trials in the space of schemata (pp. 104–6). This takes place in such a fashion that there is:

(1) progressive exploitation of the best observed schemata,
(2) increasing confidence in the estimates of the expected payoff to the best observed schemata,
and (3) testing of great numbers of new combinations of schemata (both newly generated schemata and new combinations of already tested schemata of high rank).

In the particular case of the interaction of crossover and inversion with reproduction a net of associations is induced (p. 108). Coadapted attributes (attributes defining schemata of above-average performance) become tightly linked and increase their proportion in the population (p. 127). In fact (p. 137) the expected rate of increase dP_ξ/dt of the proportion of any given schema ξ is closely approximated by

$$dP_\xi/dt = P_\xi(\mu_\xi - \bar{\mu}) = P_\xi \alpha_\xi,$$

where α_ξ is the average excess of the random variable ξ in the population $\mathfrak{A}(t)$. This formula is analogous to Fisher's (1930) classical result for single alleles and reduces to it when ξ is restricted to a single defining position. The resulting intrinsic parallelism greatly ameliorates the conflict between search and exploitation (obstacle (5)). By building up representations and models in terms of a language like the broadcast language (p. 152) the overall advantages of the schema approach can also be brought to bear on the problem of non-payoff information (obstacle (6)). The schemata provide for apportionment of credit to various aspects of the model on the basis of their relevance to realized predictions.

2. COMPUTER STUDIES

At the time of this writing several computer studies of genetic plans have been completed (and more are underway). Four studies closely related to the theoretical framework will be outlined here: R. S. Rosenberg's *Simulation of Genetic Populations with Biochemical Properties* (1967), D. J. Cavicchio's *Adaptive Search Using Simulated Evolution* (1970), R. B. Hollstien's *Artificial Genetic Adaptation in Computer Control Systems* (1971), and D. R. Frantz's *Non-linearities in Genetic Adaptive Search* (1972).

Richard Rosenberg completed his computer study of closed, small populations while formulation of the theoretical framework was still in its early stages. He concentrated on the complicated relationship between genotype and phenotype under dynamic interaction between the population and its environment. The model's central feature is the definition of phenotype by chemical concentrations. These concentrations are controlled by enzymes under genetic control. Epistasis has a critical role because several enzymes (and hence the corresponding genes) can affect any given phenotypic characteristic (chemical concentration). Though the variety of molecules, enzyme-controlled reactions, and genes is kept small to make the study feasible, it still presents a detailed picture of the propagation of

advantageous, linked genes through a small population. Moreover the study suggests general relations between the number of genes, crossover probabilities, and the rate of adaptation under epistasis. Equally important, the study makes it clear that quite complex ("molecular") definitions of phenotype can be simulated without losing relevance, up to and including suggestions for experiments *in vivo* and *in vitro*. (At least two subsequent detailed studies of biological cells were directly encouraged by this experience, R. Weinberg's *Computer Simulation of a Living Cell* [1970] and E. D. Goodman's *Adaptive Behavior of Simulated Bacterial Cells Subjected to Nutritional Shifts* [1972].)

The first study based directly upon the theoretical framework was that of Daniel Cavicchio. (J. D. Bagley's *The Behavior of Adaptive Systems Which Employ Genetic and Correlation Algorithms* [1967] is an earlier study which is a direct precursor of both this study and Frantz's.) The set of structures $\mathfrak{A}$ is taken to be a broad class of pattern classification devices based on those developed by Bledsoe and Browning (1959) and Uhr (1973). Specifically each device uses a set of detectors to process information presented by the sensors in a 25 by 25 array (cf. section 1.3 and Figures 5 through 7). After an initial "training" period, during which the device $A \in \mathfrak{A}$ accumulates information about one or more handwritten alphabets, A is tested and scored on its classification of letters from another handwritten alphabet. This score amounts to A's performance rating, its payoff $\mu(A)$. The adaptive plan, a version of the $\mathfrak{R}_d$ class of reproductive plans (pp. 94–95), generates new detectors (and, in the process, new devices) by using genetic operators which are variations on the operators discussed in sections 6.2 through 6.4.

Because of the sophistication of the problem environment, the first objective is to develop some estimate of the task's difficulty vis-à-vis the devices in $\mathfrak{A}$. Cavicchio does this by testing, in the problem setting, a large number of devices drawn at random from $\mathfrak{A}$. The observed distribution of performances is Gaussian. For a typical environment (Cavicchio calls it the "difficult task"), the mean score is 17 with a standard deviation of 5. (A perfect score would be 100.) This implies that in 1000 random trials of devices drawn from $\mathfrak{A}$ we can expect the best performance to be about 32.

To obtain an idea of the performance of a nonreproductive, but adaptive, plan in the same environment, Cavicchio applied a version of Uhr and Vossler's (1973) "detector evaluation" procedure to the search of $\mathfrak{A}$. This procedure amounts to identifying inferior detectors and replacing them with "mutated" versions. The *best* performance observed over a great many runs of 600 trials each was a score of 52; each of the runs "leveled out" somewhere between the 300th and the 600th trial. This is considerably better than a random search, being 7 standard

deviations above the mean in less than 600 trials (as compared to 3 standard deviations in 1000 trials).

Against this background Cavicchio then developed and tested a series of reproductive plans. The best of these attained a score of 75.5 in 780 trials, a score considerably beyond that attained in any of the "detector evaluation" runs. (In qualitative terms, a score of 52 would correspond to a "poor" human performance, while a score of 75.5 would correspond to a "good" human performance. Because many characters in the "difficult task" are quite similar in form, increments in scoring are difficult to attain after the easily distinguished characters have been handled.) An important general observation of this study is that the sophistication and power of a genetic plan is lost whenever M, the size of the population (data base), is very small. It is an overwhelming handicap to use only the most recent trial ($M = 1$) as the basis for generating new trials (cf. Fogel et al. 1966). On the other hand, the population need not be large to give considerable scope to genetic plans (20 was a population size commonly used by Cavicchio).

Roy Hollstien added considerably to our detailed understanding of genetic plans by making an extensive study of genetic plans as adaptive control procedures. His emphasis is on domains wherein classical "linear" and "quadratic" approaches are unavailing, i.e., domains where the performance function exhibits discontinuities, multiple peaks, plateaus, elongated ridges, etc. To give the problems a uniform setting he transforms them to discrete function optimization problems, encoding points in the domain as strings (see p. 57). An unusual and productive aspect of Hollstien's study is his translation of breeding plans for artificial genetic selection into control policies. A breeding plan which employs inbreeding within related (akin) policies, and recurrent crossbreeding of the best policies from the best families, is found to exhibit very robust performance over a range of 14 carefully selected, difficult test functions. (The test functions include such "standards" as Rosenbrock's ridge, the sum of three Gaussian 2-dimensional density functions, and a highly discontinuous "checkerboard" pattern.) The test functions are represented on a grid of 10,000 points (100 by 100). In each case the region in which the test function exceeds 90 percent of its maximum value is small. For example, test function 7 with *two* false peaks (the sum of three Gaussian 2-dimensional densities) exceeds 90 percent of its maximum value on only 42 points out of the 10,000. The breeding plans are tested over 20 generations of 16 individuals each, special provisions being made to control random effects of small sample size ("genetic drift"). The breeding plan referred to above, when confronted with test function 7, placed *all* of its trials in the "90 percent region" after 12 generations (192 trials). A random search would be expected to take 250 trials

(10,000/42) to place a *single* point in the "90 percent region." The same breeding plan performs as well or better on the 13 other test functions. Given the variety of the test functions, the simplicity of the basic algorithms, and the restricted data base, this is a striking performance.

Daniel Frantz concentrated on the internal workings of genetic plans, observing the effect, upon the population, of dependencies in the performance function. Specifically, he studies situations in which the quantity to be optimized is a function of 25 binary parameters. I.e., $\mathcal{E}$ consists of functions which are 25-dimensional and have a domain of $2^{25} = 3.2 \times 10^7$ points. Dependencies between the parameters (nonlinearities) are introduced to make it impossible to optimize the functions dimension by dimension (unimodality is avoided). Frantz's procedure is to detect the effects of these dependencies upon population structure (gene associations) by using a multidimensional chi-square contingency table. As expected from theoretical considerations (see Lemma 7.2 and the discussion following it) algebraic dependencies (between the parameters) induce statistical dependencies (between alleles). That is, in the population, combinations of alleles associated with dependent parameters occur with probabilities different from the product of the probabilities of the individual alleles. Moreover there is a positional effect on the rate of improvement: For functions with dependencies the rate of improvement is significantly greater when the corresponding alleles are close together in the representation. This effect corresponds to the theoretical result that the ability to pass good combinations on to descendants depends upon the combinations' immunity to disruption by crossover. It is significant that, for the problems studied, the optimum was attained in too short a time for the inversion operator to effectively augment the rate of improvement (by varying positional effects).

3. ADVANCED QUESTIONS

The results presented in this book have a bearing on several problem areas substantially more difficult than those recounted in section 9.1. Each of these problems has a long history and is complex enough to make sudden resolution unlikely. Nevertheless the general framework does help to focus several disparate results, providing suggestions for further progress.

As a first example, let us look at the complex of problems concerned with the dynamics of speciation. These problems have their origin in biology, but a close look shows them to be closely related to problems in the optimal allocation of

limited resources. To see this, consider the following idealized situation. There are two one-armed bandits, bandit ξ_1 paying 1 unit with probability p_1 on each trial, bandit ξ_2 paying 1 unit with probability $p_2 < p_1$. There are also M players. The casino is so organized that the bandits are continuously (and simultaneously) operated, so that at any time t, for a modest fee, a player may elect to receive the payoff (possibly zero) of one of the two bandits. The manager has, however, introduced a gimmick. If M_1 players elect to play bandit ξ_1 at time t, they must *share* the unit of payoff if the outcome is successful. That is, on that particular trial, each of the M_1 players will receive a payoff of $1/M_1$ with probability p_1. Now, let us assume that the M players *must* participate for a period of T consecutive trials. If there is but one player ($M = 1$), clearly he will maximize his income (or minimize his losses) by playing bandit ξ_1 at all times. However, if there are $M > 1$ players the situation changes. There will be stable queues, where no player can improve his payoff by shifting from one bandit to another. These occur when the players distribute themselves in the ratio $M_1/M_2 = p_1/p_2$ (at least as closely as allowed by the requirement that M_1 and M_2 be integers summing to M). For example, if $p_1 = \frac{1}{2}$, $p_2 = \frac{1}{8}$, and $M = 10$, there will be 8 players queued in front of bandit ξ_1 and 2 players in front of bandit ξ_2. We see that with limited resources (in the numerical example, a maximum of 2 units payoff per trial and an expectation of $\frac{8}{8}$ unit) the population of players must divide into two subpopulations in order to optimize individual shares of the resources (the "bandit ξ_1 players" and the "bandit ξ_2 players"). Similar considerations apply when there are $r > 2$ bandits.

We have here a rough analogy to the differentiation of individuals (the subpopulations) to exploit environmental niches (the bandits). The analogy can be made more precise by recasting it in terms of schemata. Let us consider a population of M individuals and the set of 2^{l^0} schemata defined on a given set of l^0 positions. Assume that schema ξ_i, $i = 1, \ldots, 2^{l^0}$, exploits a unique "environmental niche" which produces a total of Q_i units of payoff per time-step. (Q_i corresponds to the renewal rate of a critical, volatile resource exploited by ξ_i.) If the population contains M_i instances of ξ_i, the Q_i units are shared among them so that each instance of ξ_i receives a payoff of Q_i/M_i. Let $Q_{(1)} > Q_{(2)} > \ldots > Q_{(2^{l^0})}$ so that schema $\xi_{(1)}$ is associated with the most productive niche, $\xi_{(2)}$ with the second most productive niche, etc. Clearly when $M_{(1)}$ is large enough that $Q_{(1)}/M_{(1)} < Q_{(2)}$, an instance of $\xi_{(2)}$ will be at a reproductive advantage. Following the same line of argument as in the case of the 2 one-armed bandits, we get as a stable distribution the obvious generalization:

$$M_{(i)} = cQ_{(i)}/Q_{(j)}$$

where j is the smallest index such that

$$\sum_{i=1}^{j+1} Q_{(i)}/Q_{(j+1)} > M$$

and c is chosen so that

$$\sum_{i=1}^{j} cQ_{(i)}/Q_{(j)} = M$$

(modified so that the actual solution is in integers). For example, let $l^0 = 2$ with 2 alleles (attributes) at each locus, yielding schemata ξ_1, ξ_2, ξ_3, ξ_4 with $Q_1 = 1$, $Q_2 = 4$, $Q_3 = 8$, $Q_4 = 1$. Then for $M = 9$ there will be 6 instances of ξ_3, 3 instances of ξ_2, and no instances of ξ_1 or ξ_4 in the stable distribution.

Here we have a simple example of speciation. If the population is restricted to M individuals (by factors other than the niche payoff rates), certain combinations of alleles appear in a stable competition while other combinations are proscribed by the same competition. The example can rapidly be made more realistic by letting the payoff to each schema ξ be a random variable with *expected* payoff

$$\mu_\xi(t) = \min \{\mu_\xi^0, Q_\xi/M_\xi(t), Q/M(t)\}$$

where Q_ξ is the minimum of the renewal rates of resources characterizing the environmental niche associated with ξ, $M_\xi(t)$ is the number of instances of ξ at time t, Q is the minimum of the renewal rates of resources required by all the schemata, and $M(t)$ is the total population at time t. Now the schema ξ will increase its proportion at an intrinsic rate set by μ_ξ^0 *until* it reaches the "carrying capacity" of its niche, determined by Q_ξ, or until the total population has increased to a point that the overall "carrying capacity," determined by Q, limits further expansion. (For the reader familiar with MacArthur and Wilson's [1967] work, the effect of Q_ξ corresponds to a K selection—crowded niche—effect, whereas μ_ξ^0 is the intrinsic rate of increase, possibly wasteful of resources, under classical r selection. Q sets an ultimate limit on the carrying capacity of the environment, no matter what the diversity or organization of the species.) With typical values for the $\{Q_\xi\}$ and Q, the population will once again develop into subpopulations characterized by certain combinations of alleles (schemata), with many combinations being proscribed.

The really interesting form of this theory would characterize niches (and hence the overall payoff function μ) in terms of the varieties of schemata that could exploit them—different schemata exploiting a given niche with differing efficiencies. The dynamics of speciation would then be determined by competition within and across niches. It is interesting that under these circumstances speciation

could take place in the absence of isolation (in contrast to the usual view, cf. Mayr 1963).

Once an adaptive system discovers that given combinations of genes (or their alleles) offer a persistent advantage, new modes of advance become possible. If the given combinations can be handled as units they can serve as components ("super genes") for higher order units. In effect the system can ignore the *details* underlying the advantage conferred by a combination, and operate simply in terms of the advantage conferred. By so doing the system can explore regions of $\mathfrak{A}$, i.e., combinations of the new units, which would otherwise be tried with a much lower probability. (For example, consider two combinations of 10 alleles each under the steady state of section 7.2. If *each* of the alleles involved occurs with a frequency of 0.8, the overall combination of 20 alleles will occur with a frequency $(0.8)^{20} \cong 0.01$. On the other hand, if each of the two 10-allele combinations is maintained at a frequency of only 0.5, then the 20-allele combination will occur with frequency $(0.5)^2 = 0.25$. I.e., the expected time to occurrence will be reduced by a factor of 25.) Since combinations of advantageous units often offer an advantage beyond that of the individual units—as when the units' effects are additive (linear independence) or cooperative—they are good candidates for early testing. (The cooperative case where one unit effects an enrichment which can be exploited by another is particularly common; cf. cooperating cell assemblies or stages of a complex production activity such as illustrated in Figure 3.)

We have already discussed (section 6.3) the way in which inversion can favor association between genes. However, by controlling representation, the adaptive system can bring about changes which go much further, producing a hierarchy of units. The basic mechanism stems from the introduction of arbitrary punctuation marks to control operators (see usage (4) in section 8.3 and the discussion on pages 152–53). The adaptive system introduces a distinct punctuation mark (specific symbol string) to mark off the combinations which are to be treated as units at a given level of the hierarchy. Then the operators for that level are restricted to act only at that punctuation. (E.g., crossover takes place only at the positions marked by the given punctuation.) By introducing another punctuation mark to treat *combinations* of these units, in turn, as *new* units, and so on, the hierarchy can be extended to any number of levels. The resulting structure offers the possibility of quickly pinpointing responsibility for good or bad performance. (E.g., a hierarchy of 5 levels in which each unit is composed of 10 lower level units allows any one of 10^5 components to be selected by a sequence of 5 tests.) In the hierarchy, the units at each level are subject to the same "stability" considerations

as schemata (pp. 100–102), being continually modified by operators at lower levels. Thus certain hierarchies will be favored because of their stability, the corresponding punctuations and operators becoming common features of the overall population. Chapter 4 of Simon's book, *The Sciences of the Artificial* (1969), gives a good qualitative discussion of this and related topics.

It is natural to ask whether these operator-induced hierarchies can account for important features of such observed hierarchies as the organelle, cell, organ, organism, species, . . . hierarchy of biology, or the hierarchical organization of the CNS or a computer program. There would seem to be a strong relation between operator-induced hierarchies and the sequences of developmental biology (embryogenesis and morphogenesis) whereby, for example, a fertilized egg develops into a mature multicellular organism.

As a final problem area we can look to situations wherein payoff to a given structure varies in time and space. For example, in the case of limited resources, the resource renewal rates Q_ξ may be both temporally and spatially inhomogeneous, being described by a function $Q_\xi(x_1, \ldots, x_k, t)$. In such cases we would also expect the population at time t to be distributed spatially, yielding $\mathfrak{B}(x_1, \ldots, x_k, t)$ as the component at coordinate $(x_1, \ldots, x_k)$. After some adaptation any one component of the population, in response to the spatial variations in payoff, will generally exhibit different proportions of schemata than its neighbors.

In ecological situations, as well as in certain control situations, it is appropriate to consider the migration of structures from one component of the population to another (one coordinate to another). That is, under the direction of the adaptive plan, the jth structure $A_j(x_1, \ldots, x_k, t)$ in the population component $\mathfrak{B}(x_1, \ldots, x_k, t)$ may be transferred to a neighboring coordinate $(x_1', \ldots, x_k')$, becoming an element of $\mathfrak{B}(x_1', \ldots, x_k', t+1)$. (Such systems can be usefully described with the help of cellular automata; see R. F. Brender's *A Programming System for Cellular Spaces* 1969 and *Essays on Cellular Automata* edited by A. W. Burks 1970.) Under these conditions we would expect to observe a *spatial* diffusion of schemata. Thus schemata having a large number of instances in $\mathfrak{B}(x_1, \ldots, x_k, t)$ would be expected to appear in fair numbers in neighboring components of the population, even if their performance *there* is poor. At the "boundaries" between different niches the genetic operators will produce unusual "hybrids" of schemata common in each of the niches. That is, where there are sharp changes in the $Q_\xi(x_1, \ldots, x_k, t)$, crossover will yield a wide range of new schemata, which would otherwise occur with low probability. Many of these schemata will be unfit or fit only in the boundary region, but some may exhibit exceptional performance on one or both niches. The relation to Mayr's (1963) description of speciation as the

result of contact between previously isolated, locally adapted populations is manifest. (See, however, the comment on page 166.) There is much to be learned about these processes, particularly with reference to schemata or coadapted sets. (Some of the most interesting work to date has been carried out by A. Brues 1972.) It is clear that the addition of migration rules to reproductive plans affords a sophisticated approach to spatially inhomogeneous environments, but we need to know a great deal more about the efficiency and robustness of such an approach (paralleling the development of chapters 5 and 7 for the homogeneous case).

So far we have been discussing spatial inhomogeneity of payoff, but temporal inhomogeneity or *nonstationarity* is an even more difficult problem. There are four points at which the results of this book have a bearing on such problems. First, and most obvious, the rapid response of reproductive plans, exhibited concretely in the studies of Cavicchio (1970) and Hollstien (1971), permits "tracking" of the changing payoff function. As long as the payoff function changes at a rate comparable to the response rate, overall performance will be good. The proportions of schemata in the population will change rapidly enough to take advantage of current features of the environment. As a second point, it should be noted that the rank bestowed on a schema (its proportion in successive generations) is the geometric mean of the observed averages $\mu_\xi(t)$ (see Lemma 7.2). Thus more rapid fluctuations will favor schemata which exhibit the best (geometric) mean performance when subjected to the fluctuations. Third, if there are repetitive (not necessarily cyclic) features over time, dominance change provides a mechanism for retaining useful schemata when the features are not in force (see pages 115–16). By occasionally (say once every few generations) giving recessive status to *instances* of currently favored schemata, they can be reserved against adverse environmental configurations. In particular, these recessive instances have a much reduced testing rate (see page 115). As a result the recessive versions are relatively unaffected by environmental changes which quickly eliminate the dominant version. By occasionally returning an *instance* of a recessive schema to dominant status it can be tested against the current environmental configuration. If the dominant instance achieves above-average performance it will reproduce rapidly, producing an increasing proportion of dominant instances in the population. (If the performance is below average the newly dominant instance will quickly disappear, at no great cost to the efficiency of the adaptive plan.) Finally, by making the intrachromosomal duplication of a schema ξ subject to the disappearance of an environmental feature currently exploited by ξ, the effective mutation rate of ξ can be increased. For example, let the schema ξ be associated with a *sensor* (see pages 153–54) which detects the environmental feature exploited by ξ. Let intra-

chromosomal duplication be an operator controlled by the sensor; i.e., whenever the sensor is deactivated, intrachromosomal duplication takes place on the sets of genes associated with the sensor. In consequence, disappearance of the environmental feature will result in many copies of the genes, and hence the schemata, associated with the sensor. With a fixed mutation rate for each gene, the number of mutants of a given schema in the population will depend upon the number of copies thereof. Thus by providing many copies within a chromosome, the effective mutation rate is correspondingly increased. As a result, this (hypothetical) mechanism provides many variants relevant to the crisis. At the same time it retains whatever advantage remains to the original schema ξ. In biology there are varying amounts of evidence for the foregoing responses to nonstationarity, and some of the predicted effects have been demonstrated in simulations, but again we are a long way from a theory, or even good experimental confirmation, of their efficiency.

In these nine chapters we have come only a short way in the study of adaptation as a general process. The book's main objective has been to make it plausible that simple mechanisms can generate complex adaptations; however, the book will have fulfilled its role if it has communicated enough of adaptation's inherent fascination to make the reader's effort worthwhile.