Why the Schema Theorem is Correct also in the Presence of Stochastic Effects

Riccardo Poli
School of Computer Science
The University of Birmingham
Birmingham, B15 2TT, UK
R.Poli@cs.bham.ac.uk

Abstract- Holland's schema theorem has been criticised in (Fogel and Ghozeil 1997, Fogel and Ghozeil 1998, Fogel and Ghozeil 1999) for not being able to estimate correctly the expected proportion of a schema in the population when fitness proportionate selection is used in the presence of noise or other stochastic effects. This is incorrect for two reasons. Firstly, the theorem in its original form is not applicable to this case. As clarified in the paper, if the quantities involved in schema theorems are random variables, the theorems must be interpreted as conditional statements. Secondly, the conditional versions of Holland and other researchers' schema theorems are indeed very useful to model the sampling of schemata in the presence of stochasticity. In the paper I show how one can calculate the correct expected proportion of a schema in the presence of stochastic effects when selection only is present, using a conditional interpretation of Holland's schema theorem. In addition, I generalise this result (again using schema theorems) to the case in which crossover, mutation, and selection with replacement are used. This can be considered as an exact schema theorem applicable both in the presence and in the absence of stochastic effects.

1 Introduction

Since John Holland's work in the mid seventies and his well-known schema theorem (Holland 1975), schemata are traditionally used to explain why GAs and more recently GP (Poli and Langdon 1997, Poli and Langdon 1998, Rosca 1997) work. Schemata are similarity templates representing entire groups of points in the search space. A schema theorem is a description of how schemata are expected to propagate generation after generation under the effects of selection and the search operators (typically crossover and mutation). In an alternative interpretation schemata are seen as subsets of the search space, and schema theorems are interpreted as descriptions of the way the expected number of elements (or the proportion) of the population belonging to such subsets changes over time.

The usefulness of Holland's schema theorem has been widely criticised (see for example (Altenberg 1995, Macready and Wolpert 1996, Fogel and Ghozeil 1997, Fogel and Ghozeil 1998, Fogel and Ghozeil 1999)). The theorem certainly has some limitations: it gives only a lower bound for the expected value of the number (or proportion) of instances of a schema at the next generation. The presence of the expectation operator means that it is not easy to use the theorem to predict the behaviour of a genetic algorithm over multiple generations. Also, since Holland's schema theorem provides only a lower bound (it accounts only for schema disruption and survival, not creation), its predictions may be difficult to use in practice. For these reasons, many researchers nowadays believe that Holland's schema theorem is nothing more than a trivial tautology of no use whatsoever (see for example (Vose 1999, preface)). However, as stated in (Radcliffe 1997) the problem with Holland's schema theorem is not the theorem itself, rather its over-interpretations.

One such over-interpretations is present in (Fogel and Ghozeil 1997, Fogel and Ghozeil 1998, Fogel and Ghozeil 1999). In that work, Fogel and Ghozeil identified an important, unknown, bias in the sampling of schemata due to fitness proportionate selection in the presence of stochastic effects, but stated that such a bias was not modelled by Holland's schema theorem. In this paper it will be shown that that was an incorrect interpretation for at least two reasons. Firstly, Holland's schema theorem in its original form is not applicable when the quantities in its r.h.s. are random variables. Holland's schema theorem assumes (and its proof relies on) the fact that the quantities on the r.h.s. are constants. In the paper it will be clarified that if the quantities involved in the r.h.s. of schema theorems are random variables, the theorems must be interpreted as conditional statements. Secondly, when properly interpreted, Holland and other researchers' schema theorems (Stephens and Waelbroeck 1997, Stephens and Waelbroeck 1999, Poli 2000*a*, Poli 2000*b*) are indeed very useful to model the sampling of schemata in the presence of stochasticity. In the paper I show how one can calculate the correct expected proportion of a schema in the presence of stochastic effects when selection only is present, using a conditional interpretation of Holland's schema theorem. In addition, I generalise this result (again using schema theorems) to the case in which crossover, mutation, and selection with replacement are used, both for fixed-size and variable-length GAs and for genetic programming.

The paper is organised as follows. In Section 2 Holland's schema theory, Stephens and Waelbroeck's exact schema theory and Fogel and Ghozeil's results are summarised. Section 3 discusses the conditional interpretation of schema theorems. This is used to derive equations expressing the expected proportion of a schema in the presence of stochastic effects in Section 4. The results presented in the paper are discussed in Section 5 and some conclusions are drawn in Section 6.

2 Background

2.1 Holland's Schema Theory

In the context of GAs operating on binary strings, a schema is a string of symbols taken from the alphabet {0,1,#}. The character # is interpreted as a "don't care" symbol, so that a schema can represent several bit strings. For example the schema #10#1 represents four strings: 01001, 01011, 11001 and 11011. The number of non-# symbols is called the *order* $\mathcal{O}(H)$ of a schema H. The distance between the furthest two non-# symbols is called the *defining length* $\mathcal{L}(H)$ of the schema. Holland obtained a result (often referred to as "the schema theorem") which predicts how the number of strings in a population matching (or belonging to) a schema is expected to vary from one generation to the next (Holland 1975). The theorem can be reformulated as follows:[1]

$$E\left[\frac{m(H,t+1)}{M}\right] \geq p(H,t)\cdot(1-p_m)^{\mathcal{O}(H)}\cdot \left[1-p_{xo}\frac{\mathcal{L}(H)}{N-1}\left(1-p(H,t)\right)\right] \quad (1)$$

where p_m is the probability of mutation per bit, p_{xo} is the probability of crossover, N is the number of bits in a string, M is the number of strings in the population, $m(H,t+1)$ is the number of strings matching the schema H at generation $t+1$, $E[\cdot]$ is the expectation operator, and $p(H,t)$ is the probability of selection of the schema H. In fitness proportionate selection, this is given by $p(H,t)=\frac{m(H,t)f(H,t)}{M\bar{f}(t)}$ where $m(H,t)$ is the number of strings matching the schema H at generation t, $f(H,t)$ is the mean fitness of the strings matching H, and $\bar{f}(t)$ is the mean fitness of the strings in the population.

2.2 Exact Schema Theories

If we define the *total transmission probability* for a schema H, $\alpha(H,t)$, as the probability that, at generation t, every time we create a new individual to be inserted in the next generation such an individual will sample H (Poli *et al.* 1998), then an exact schema theorem takes the simple form:

$$E[m(H,t+1)] = M\alpha(H,t). \quad (2)$$

So, the formulation of an exact schema theorem is possible but it requires the knowledge of the total transmission probability α (the r.h.s. of Equation 1 gives only a lower bound for $\alpha(H,t)$). This requires that not only schema survival and schema disruption events be modelled mathematically but also that schema creation events are.

Modelling schema creation is not an easy task if one wants to do that using only the properties of the schema H (such as the number of instances of H and the fitness of H) and those of the population when expressing the quantity α. Indeed, none of the schema theorems presented to date in the literature have succeeded in doing this. This is the reason why, in general, schema theorems provide lower bounds.

Thanks to the recent work of Stephens and Waelbroeck (Stephens and Waelbroeck 1997, Stephens and Waelbroeck 1999) it is, however, possible to express exactly $\alpha(H,t)$ for GAs operating on fixed-length bit strings by using properties of lower-order schemata which are supersets of the schema under consideration. Assuming $p_m=0$, in a binary GA the total transmission probability is given by the following equation:[2]

$$\alpha(H,t) = (1-p_{xo})p(H,t) + \frac{p_{xo}}{N-1}\sum_{i=1}^{N-1} p(L(H,i),t)p(R(H,i),t) \quad (3)$$

where p_{xo} is the crossover probability, $p(x,t)$ is the selection probability of a schema x at generation t, $L(H,i)$ is the schema obtained by replacing with "don't care" symbols (#) all the elements of H from position $i+1$ to position N, $R(H,i)$ is the schema obtained by replacing with "don't care" symbols all the elements of H from position 1 to position i, and i varies over the valid crossover points.[3] For example, if H =##1111, $L(H,1)$ =######, $R(H,1)$ =##1111, $L(H,3)$ =##1###, and $R(H,3)$ =###111.

Stephens and Waelbroeck's theory has been recently generalised in (Poli 2000*a*, Poli 2000*b*) where an exact expression of $\alpha(H,t)$ for genetic programming with one-point crossover was reported. This is valid for variable-length and non-binary GAs as well as GP and standard GAs.

2.3 Selection-only Case

If selection only is present, i.e. $p_{xo}=0$ and $p_m=0$, then $\alpha(H,t)=p(H,t)$. In this situation Holland's schema theorem and Stephens and Waelbroeck's exact schema theorem coincide. On the assumption that fitness proportionate selection is used, these can be written as:

$$E[m(H,t+1)] = \frac{m(H,t)f(H,t)}{\bar{f}(t)}. \quad (4)$$

2.4 Fogel and Ghozeil's Analysis of Proportional Selection

In (Fogel and Ghozeil 1997, Fogel and Ghozeil 1998, Fogel and Ghozeil 1999) the behaviour of fitness proportionate selection in the presence of stochastic effects was studied. In

[1] Equation 1 is a slightly different version of Holland's original theorem which applies when crossover is performed taking both parents from the mating pool (Goldberg 1989, Whitley 1993) and which is valid for any selection-with-replacement mechanism.

[2] This equation can be obtained either by simplifying the results in (Stephens and Waelbroeck 1997, Stephens and Waelbroeck 1999) or by decomposing the probability that H be created into the sum of the probabilities that this will happen for each possible crossover point.

[3] The symbols L and R stand for "left part of" and "right part of", respectively.

that work Fogel and Ghozeil showed that when schema fitness takes the form of a random variable (e.g. when the fitness function is noisy or when the population is initialised randomly), fitness proportionate selection may not in general allocate trials to competing schemata on the basis of their relative observed fitnesses.

This bias in the sampling of schemata was studied in the case of two competing schemata, H and H', represented by equal proportions of individuals in the population, i.e. $m(H,t) = m(H',t) = M/2$. In this situation Equation 4 may be rewritten as:

$$E\left[\frac{m(H,t+1)}{M}\right] = \frac{f(H,t)}{f(H,t)+f(H',t)}. \quad (5)$$

Fogel and Ghozeil showed that when $f(H,t)$ and $f(H',t)$ are random variables, the correct way to calculate the expected fraction of trials allocated to H is given by:[4]

$$E\Big[\frac{m(H,t+1)}{M}\Big] = E\Big[\frac{f(H,t)}{f(H,t)+f(H',t)}\Big]. \quad (6)$$

This result was proved by calculating directly the expected fraction of times the schema H is selected in n repetitions of a two-step decision process in which, firstly, samples f and f' of $f(H,t)$ and $f'(H,t)$ are drawn and, then, these are used to realise a Bernoulli random variable with success probability $f/(f+f')$. Although valid only in the case of two equally-represented competing schemata under fitness proportionate selection, Equation 6 can be seen as a form of schema theorem applicable in the presence of stochastic schema fitnesses.

So, in the presence of stochastic effects, fitness proportionate selection may easily lead not to sample the schema H according to the ratio $\frac{E[f(H,t)]}{E[f(H,t)]+E[f(H',t)]}$ of the expected fitnesses $E[f(H,t)]$ and $E[f(H',t)]$. This is because fitness proportionate selection samples H according to $E[\frac{f(H,t)}{f(H,t)+f(H',t)}]$ and in general $E[\frac{f(H,t)}{f(H,t)+f(H',t)}] \neq \frac{E[f(H,t)]}{E[f(H,t)]+E[f(H',t)]}$, except for some special cases. In (Fogel and Ghozeil 1997, Fogel and Ghozeil 1998, Fogel and Ghozeil 1999) the difference between $\frac{E[f(H,t)]}{E[f(H,t)]+E[f(H',t)]}$ and $E[\frac{f(H,t)}{f(H,t)+f(H',t)}]$ was calculated and studied for a number of probability different density functions.

If one compares the right-hand sides of Equations 5 and 6 one might think that the (selection-only) schema theorem gives incorrect predictions in the case of stochastic schema fitnesses. If there is no bias in the estimates provided by the fitnesses of the members of the population then $f(H,t)$ and $f(H',t)$ are exactly the same as the mean fitnesses of the points in the hyperspaces H and H', respectively. In this case the r.h.s. of Equation 5 would be $\frac{E[f(H,t)]}{E[f(H,t)]+E[f(H',t)]}$. However, since in general $E[\frac{f(H,t)}{f(H,t)+f(H',t)}] \neq \frac{E[f(H,t)]}{E[f(H,t)]+E[f(H',t)]}$, Holland's schema theorem would appear to fail in the presence of stochastic effects. Indeed Fogel and Ghozeil suggested that:

> (a) "In general, there is no a priori reason to expect the schema theorem to adequately describe the mean sampling of alternative schemata when the fitness evaluation of those schemata is governed by a random variable." (Fogel and Ghozeil 1997)
>
> (b) "Despite the simple case analyzed here, it is shown that the impact of random initialization degrades the schema theorem as a predictive tool for estimating the expected proportion of a particular schema in the population after a single iteration of proportional selection" (Fogel and Ghozeil 1998)
>
> (c) "Even when these factors are not considered, the analysis offered indicates that the allocation of trials to competing schemata will generally be a "misallocation" that deviates from the schema theorem." (Fogel and Ghozeil 1999)

I believe that the criticisms reported above are incorrect for at least two reasons.

Firstly, Holland's schema theorem in its original form is not applicable when the quantities in its r.h.s. are random variables. This was recognised by Fogel and Ghozeil who wrote:

> "The schema theorem in [1] refers only to the specific realized values of competing schemata in a population; it is not intended to handle the case when the fitness of these schemata are described by random variables." (Fogel and Ghozeil 1997)

So, in the light of this, at a superficial level (a) would seem correct, while (b) and (c) are incorrect. However, at a deeper level also (a) is incorrect. Indeed, in the next section, it will be shown that, in general, if the quantities involved in schema theorems are random variables, the theorems are still correct provided they are interpreted as conditional statements.

Secondly, as shown in Section 4, when properly interpreted, Holland and other researchers' schema theorems can be used to obtain exact models of the sampling of schemata not only in the presence of stochasticity and selection but also when other genetic operators are used. So, the allocation of trials to schemata does not really deviate from what predicted using schema theories.

3 Conditional Schema Theorems

The schema theorems described in Sections 2 and in other work are valid on the assumption that the quantities on the

[4] For the sake of uniformity with earlier work, the notation used in this paper is different from the one adopted in (Fogel and Ghozeil 1997, Fogel and Ghozeil 1998, Fogel and Ghozeil 1999). For example, in Fogel and Ghozeil's work the stochastic variables $f(H,t)$, $f(H',t)$ and $\frac{f(H,t)}{f(H,t)+f(H',t)}$ are termed X, Y and Z, respectively.

r.h.s. are constant. This is because the proofs for such theorems have been obtained on this assumption (even if this is rarely stated explicitly). If instead these quantities are random variables, the theorems are still correct but they need to be interpreted as conditional statements (Poli 1999*a*).

For example, exact schema theorems of the form in Equation 2 like the ones in (Stephens and Waelbroeck 1997, Stephens and Waelbroeck 1999, Poli 2000*a*, Poli 2000*b*) should be interpreted as:

$$E[m(H,t+1)|\alpha(H,t)=a] = Ma, \qquad (7)$$

a being an arbitrary constant in [0,1]. That is, these theorems provide information on the conditional expected value of the number of instances of a schema at the next generation, i.e. the expected value of $m(H,t+1)$ on the assumption that $\alpha(H,t)=a$, rather than on $E[m(H,t+1)]$. In other words these theorems should be meant to be saying: "if the total transmission probability of the schema is a, then the expected number of copies of the schema at the next generation is Ma".[5]

Likewise, the form of the schema theorem applicable when fitness proportionate selection only is present (Equation 4) becomes

$$E[m(H,t+1)|m(H,t)=m_H, \\ f(H,t)=f_H, \bar{f}(t)=f_P] = \frac{m_H f_H}{f_P}, \qquad (8)$$

where m_H, f_H and f_P are arbitrary constants and $m(H,t)$, $f(H,t)$ and $\bar{f}(t)$ are random variables. In other words the theorem says: "if the average fitness of the individuals in the population containing the schema H is f_H, there are m_H copies of it in the population and the average fitness of the population is f_P, then the expected number of copies of the schema at the next generation is $\frac{m_H f_H}{f_P}$".

If we consider again a two-competing-schemata situation we can easily see that Equation 8 becomes

$$E\Big[\frac{m(H,t+1)}{M}\Big|m(H,t)=m_H, \\ f(H,t)=f_H, f(H',t)=f_{H'}\Big] \qquad (9) \\ = \frac{m_H f_H}{m_H f_H + (M-m_H)f_{H'}},$$

where I used the fact that $f_P = \frac{m_H f_H + (M-m_H)f_{H'}}{M}$ in this example. If there are equal proportions of the two schemata, like in the example considered by Fogel and Ghozeil and described in Section 2.4, this simplifies to

$$E\Big[\frac{m(H,t+1)}{M}\Big|f(H,t)=f_H, f(H',t)=f_{H'}\Big] \\ = \frac{f_H}{f_H+f_{H'}}. \qquad (10)$$

[5]The proof of Equation 7 is so simple to be almost unnecessary: conditionally to the random variable $\alpha(H,t)$ taking a constant value a, the random variable $m(H,t+1)$ is binomially distributed with success probability a, so, its conditional expected value is Ma.

This is the correct way of writing Equation 5 when schema fitnesses are random variables: as a conditional statement. It should be noted that Equations 10 and 5 have right-hand sides of the same form, but different left-hand sides.

4 Sampling in the Presence of Stochastic Effects

For well-known properties of conditional expected values, it is easy to calculate $E[m(H,t+1)/M]$ in the presence of stochastic effects by using the conditional versions of schema theorems provided in the previous section.

For example, in the case of two competing schemata represented by equal proportions of the population one can write

$$E\Big[\frac{m(H,t+1)}{M}\Big] = \\ \int_{f_H}\int_{f_{H'}} E\Big[\frac{m(H,t+1)}{M}\Big|f(H,t)=f_H, f(H',t)=f_{H'}\Big] \\ \cdot \mathrm{pdf}(f_H, f_{H'})df_H df_{H'}, \qquad (11)$$

where $\mathrm{pdf}(f_H, f_{H'})$ is the joint probability density function for the variables $f(H,t)$ and $f(H',t)$. By using the simplified conditional schema theorem in Equation 10 we obtain:

$$E\Big[\frac{m(H,t+1)}{M}\Big] = \\ \int_{f_H}\int_{f_{H'}} \frac{f_H}{f_H+f_{H'}}\mathrm{pdf}(f_H, f_{H'})df_H df_{H'}. \qquad (12)$$

For a well-know theorem on the expected value of functions of random variables (see for example (Papoulis 1965, page 206)), the r.h.s. of this equation is equal to $E[\frac{f(H,t)}{f(H,t)+f(H',t)}]$. So,

$$E\Big[\frac{m(H,t+1)}{M}\Big] = E\Big[\frac{f(H,t)}{f(H,t)+f(H',t)}\Big], \qquad (13)$$

which is the same result as Fogel and Ghozeil's (Equation 6).

Thanks to the conditional version of the exact schema theorem in Equation 7, this result can easily be generalised to the general case in which selection, crossover, mutation and stochastic effects are all present. We start from

$$E\left[\frac{m(H,t+1)}{M}\right] = \int E\left[\frac{m(H,t+1)}{M}|\alpha(H,t)=a\right] \\ \cdot \mathrm{pdf}(a)da,$$

where $\mathrm{pdf}(a)$ is the probability density function of $\alpha(H,t)$.[6] By following exactly the same steps as for the two-competing-schemata case reported above, this leads to the *general result*:

$$E\left[\frac{m(H,t+1)}{M}\right] = E[\alpha(H,t)], \qquad (14)$$

[6]In this equation, the conditioning event $\alpha(H,t)=a$ can be replaced by a list of conditioning events involving the variables of which α is a function. If this is done, $\mathrm{pdf}(a)$ should be replaced with the joint density function of the same variables. In the case of fitness proportionate selection these variables are: $\bar{f}(t)$, $m(L(H,i),t)$, $f(L(H,i),t)$, $m(R(H,i),t)$, $f(R(H,i),t)$ for $i=1,\cdots,N-1$ (see (Poli 1999*a*) for more details).

where, for binary GAs $\alpha(H,t)$ is given in (Stephens and Waelbroeck 1997, Stephens and Waelbroeck 1999) (and also in Equation 4 when $p_m = 0$), while for genetic programming with one-point crossover and for variable-size and non-binary GAs $\alpha(H,t)$ is given by the exact schema theorems in (Poli 2000*a*, Poli 2000*b*).

Equation 14 can be considered as a general exact schema theorem applicable both in the presence and in the absence of stochastic effects. It is easy to show that this is also valid for discrete random variables.

5 Discussion

The results in the first part of the previous section confirm the important observation made by Fogel and Ghozeil that the correct way of calculating the expected fraction of individuals sampling the schema H (in the two-competing-schemata example) is to use $E[\frac{f(H,t)}{f(H,t)+f(H',t)}]$. This is important because it highlights the fact that, in the presence of stochastic effects, fitness proportionate selection may easily lead not to sample the schema H according to the ratio $\frac{E[f(H,t)]}{E[f(H,t)]+E[f(H',t)]}$ of the expected fitnesses $E[f(H,t)]$ and $E[f(H',t)]$.

However, the proof for Equation 13 presented in this paper is based on the very same thing that (Fogel and Ghozeil 1997, Fogel and Ghozeil 1998, Fogel and Ghozeil 1999) criticised: Holland's schema theorem.

As shown in the second part of the previous section, this proof can be generalised to model the sampling of schemata in the presence of stochasticity, genetic operators, and any type of selection-with-replacement algorithm for any evolutionary algorithm and representation for which an exact expression for $\alpha(H,t)$ exists.

Furthermore, the proofs of the equations in the previous section do not make any assumptions on the independence of the random variables involved (e.g. of $f(H,t)$ and $f(H',t)$), which was instead assumed in some of the results reported in (Fogel and Ghozeil 1997, Fogel and Ghozeil 1998, Fogel and Ghozeil 1999).

6 Conclusions

The work in this paper is not motivated by the desire to defend Holland's schema theorem at all costs. As indicated in Section 1, Holland's schema theorem does have some limitations. However, one of the motivations for this work is certainly to rectify common incorrect beliefs regarding schema theories.

One such beliefs is that the limitations of Holland's theorem, mainly deriving from the presence of the expectation operator and the inequality in Equation 4, are shared by schema theories in general (see for example (Vose 1999, page 211)). While this might be true for some schema theories proposed in the past, as shown in more recent work (Stephens and Waelbroeck 1997, Stephens and Waelbroeck 1999, Poli 2000*a*, Poli 2000*b*) it is now possible, for example, to make the effects and the mechanisms of schema creation explicit obtaining exact schema theorems (rather than lower bounds). In addition, it is now also possible to get rid of the expected value and obtain schema-theory-based GA convergence results (Poli 1999*a*, Poli 1999*c*, Poli 1999*b*).

Another such belief is that Holland's schema theorem does not adequately model the sampling behaviour of a GA in the presence of stochastic effects (Fogel and Ghozeil 1997, Fogel and Ghozeil 1998, Fogel and Ghozeil 1999). This paper clarifies that Holland and other researchers' schema theorems must be interpreted as conditional statements when the quantities involved in their right-hand sides are not constants, but random variables. Thanks to this interpretation, it was shown how Holland's schema theorem can be used to calculate correctly the expected proportion of a schema in the population produced by fitness proportionate selection in the presence of stochastic effects. This shows that the criticisms to the schema theorem expressed in (Fogel and Ghozeil 1997, Fogel and Ghozeil 1998, Fogel and Ghozeil 1999) were largely unjustified.

The paper has also shown how the exact schema theorems recently proposed in the literature can be used to model the sampling of schemata in the presence of stochasticity, one-point crossover, mutation, and selection in standard binary GAs, in variable-length and non-binary GAs, and in genetic programming.

Together with other recent work, these results corroborate the author's belief that, when correctly interpreted, properly developed and used, schema theorems can be very useful tools to understand evolutionary algorithms, make predictions on their behaviour, and help design competent algorithms.

Acknowledgements

The author wishes to thank the members of the Evolutionary and Emergent Behaviour Intelligence and Computation (EEBIC) group at Birmingham for useful comments and discussion. D. B. Fogel is also warmly thanked for the useful email discussion and clarifications, for providing a copy of (Fogel and Ghozeil 1999) and for his very helpful comments on a earlier version of this paper.

Bibliography

Altenberg, Lee (1995). The Schema Theorem and Price's Theorem. In: *Foundations of Genetic Algorithms 3* (L. Darrell Whitley and Michael D. Vose, Eds.). Morgan Kaufmann. Estes Park, Colorado, USA. pp. 23–49.

Fogel, D. B. and A. Ghozeil (1997). Schema processing under proportional selection in the presence of random effects. *IEEE Transactions on Evolutionary Computation* **1**(4), 290–293.

Fogel, D. B. and A. Ghozeil (1998). The schema theorem and the misallocation of trials in the presence of stochastic effects. In: *Evolutionary Programming VII: Proc. of the 7th Ann. Conf. on Evolutionary Programming* (V.W. Porto, N. Saravanan, D. Waagen and A.E. Eiben, Eds.). Springer. Berlin. pp. 313–321.

Fogel, D. B. and A. Ghozeil (1999). Schema processing, proportional selection, and the misallocation of trials in genetic algorithms. *Information Sciences*.

Goldberg, David E. (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley. Reading, Massachusetts.

Holland, J. (1975). *Adaptation in Natural and Artificial Systems*. University of Michigan Press. Ann Arbor, USA.

Macready, William G. and David H. Wolpert (1996). On 2-armed gaussian bandits and optimization. Sante Fe Institute Working Paper 96-05-009.

Papoulis, A. (1965). *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill Kogakusha, Ltd.. Tokyo.

Poli, Riccardo (1999*a*). Probabilistic schema theorems without expectation, recursive conditional schema theorem, convergence and population sizing in genetic algorithms. Technical Report CSRP-99-3. University of Birmingham, School of Computer Science.

Poli, Riccardo (1999*b*). Schema theorems without expectations. In: *Proceedings of the Genetic and Evolutionary Computation Conference* (Wolfgang Banzhaf, Jason Daida, Agoston E. Eiben, Max H. Garzon, Vasant Honavar, Mark Jakiela and Robert E. Smith, Eds.). Vol. 1. Morgan Kaufmann. Orlando, Florida, USA. p. 806.

Poli, Riccardo (1999*c*). Schema theory without expectations for GP and GAs with one-point crossover in the presence of schema creation. In: *Foundations of Genetic Programming* (Thomas Haynes, William B. Langdon, Una-May O'Reilly, Riccardo Poli and Justinian Rosca, Eds.). Orlando, Florida, USA.

Poli, Riccardo (2000*a*). Hyperschema theory for GP with one-point crossover, building blocks, and some new results in GA theory. In: *Genetic Programming, Proceedings of EuroGP'2000* (Riccardo Poli, Wolfgang Banzhaf, William B. Langdon, Julian F. Miller, Peter Nordin and Terence C. Fogarty, Eds.). Vol. 1802 of *LNCS*. Springer-Verlag. Edinburgh. pp. 163–180.

Poli, Riccardo (2000*b*). Exact schema theorem and effective fitness for GP with one-point crossover. In: *Proceedings of the Genetic and Evolutionary Computation Conference* (D. Whitley *et al.*, Ed.). Morgan Kaufmann. Las Vegas.

Poli, Riccardo and W. B. Langdon (1997). A new schema theory for genetic programming with one-point crossover and point mutation. In: *Genetic Programming 1997: Proceedings of the Second Annual Conference* (John R. Koza, Kalyanmoy Deb, Marco Dorigo, David B. Fogel, Max Garzon, Hitoshi Iba and Rick L. Riolo, Eds.). Morgan Kaufmann. Stanford University, CA, USA. pp. 278–285.

Poli, Riccardo and William B. Langdon (1998). Schema theory for genetic programming with one-point crossover and point mutation. *Evolutionary Computation* **6**(3), 231–252.

Poli, Riccardo, William B. Langdon and Una-May O'Reilly (1998). Analysis of schema variance and short term extinction likelihoods. In: *Genetic Programming 1998: Proceedings of the Third Annual Conference* (John R. Koza, Wolfgang Banzhaf, Kumar Chellapilla, Kalyanmoy Deb, Marco Dorigo, David B. Fogel, Max H. Garzon, David E. Goldberg, Hitoshi Iba and Rick Riolo, Eds.). Morgan Kaufmann. University of Wisconsin, Madison, Wisconsin, USA. pp. 284–292.

Radcliffe, Nicholas J. (1997). Schema processing. In: *Handbook of Evolutionary Computation* (T. Baeck, D. B. Fogel and Z. Michalewicz, Eds.). pp. B2.5–1–10. Oxford University Press.

Rosca, Justinian P. (1997). Analysis of complexity drift in genetic programming. In: *Genetic Programming 1997: Proceedings of the Second Annual Conference* (John R. Koza, Kalyanmoy Deb, Marco Dorigo, David B. Fogel, Max Garzon, Hitoshi Iba and Rick L. Riolo, Eds.). Morgan Kaufmann. Stanford University, CA, USA. pp. 286–294.

Stephens, C. R. and H. Waelbroeck (1997). Effective degrees of freedom in genetic algorithms and the block hypothesis. In: *Proceedings of the Seventh International Conference on Genetic Algorithms (ICGA97)* (Thomas Bäck, Ed.). Morgan Kaufmann. East Lansing.

Stephens, C. R. and H. Waelbroeck (1999). Schemata evolution and building blocks. *Evolutionary Computation* **7**(2), 109–124.

Vose, Michael D. (1999). *The simple genetic algorithm: Foundations and theory*. MIT Press. Cambridge, MA.

Whitley, Darrell (1993). A genetic algorithm tutorial. Technical Report CS-93-103. Department of Computer Science, Colorado State University.