Lecture 11: Critiques of Schema Theory

Recap

- The Schema Theory tracks changes in frequency of representation of subsets of the search space (schemata), independent of the other schemata in the population
  - Its formulation reveals the conflicting pressures of positive selection, and destructive genetic operators
- It has been used to make some proposals about how GAs (should) work
  - The Building Block Hypothesis
  - The Two-Armed Bandit Analogy
  - The Principle of Minimum Alphabets
- But is it a good explanation of GA behaviour?

“Schema Theory tells us almost nothing about GA behaviour”

Michael Vose, 1999

The Building Block Hypothesis

- From the Schema Theorem we know that the representation of a schema S in the population should increase when

$$r(S,t) \geq 1 + \chi \frac{l_S}{\ell - 1} + \mu k_S$$

- So, a GA should work well when it can combine short, low-order schemata (*building blocks*) to form better building blocks
- *But...* as 'building blocks' are combined together the order and length of the new building block is increased
  - The new building block becomes more vulnerable to disruption by crossover and selection
  - So, the BBH suggests how GAs *shouldn't* work!
  - On the 'Royal Road' fitness function, a GA is outperformed by hill-climbing (more later...)

Implicit Parallelism

- Any chromosome is a member of 2^{ℓ} schemata
- By evaluating a chromosome's fitness we are simultaneously evaluating many schemata
  - I.e. we are parallelising the search of the solution space
- *But...* this assumes that the population is uniformly distributed
  - As the population converges, the number of schemata represented will decrease
  - Hence the parallelisation of the search decreases

The Two-Armed Bandit Analogy

- Holland drew an analogy between competition amongst schemata, and the n-armed bandit from statistical decision theory
- Holland argued that by allocating exponentially increasing numbers of trials to the superior schemata, the GA approximates the optimal strategy
- *But...*
  - For the GA there are actually several bandits
  - As we are solving a non-linear problem, the order in which the bandits are solved is likely to be important
  - Identifying the schemata involved in the competitions is non-trivial, due to *confounding*
    - E.g. are we comparing (1 * *) with (0 * *), or (* 1 *) with (* 0 *)
  - It is not clear in any case that exponentially increasing trials is the optimal strategy
  - Furthermore, the GA is unlikely to achieve exponentially increasing trials in the long run

The Principle of Minimal Alphabets

- Holland argued for the optimality of binary encoding, using schema theory
- The number of possible schemata for an alphabet $\mathcal{A}$ is $|\mathcal{A}+1|^{\ell}$
- This will be maximised when chromosome length ℓ is maximised, which is minimised when $|\mathcal{A}|$ is minimised
  - So the maximum number of possible schemata is achieved by binary encoding
- *But...* what is the benefit of maximising the number of schemata if they can't be processed effectively?
  - However, binary encoding is still optimal in terms of minimum population size during random initialisation (see lecture 3)

The Schema Theorem

- The Schema Theorem predicts that above average fitness schemata proliferate in the population
- *But...* the Schema Theorem only tells us what happens to a schema *independent* of other schemata in the population
  - Other schemata will also be proliferating
  - Over time the population will converge and its fitness variance will decrease
  - Hence the fitness of a schema will tend to the population average, it’s fitness ratio will approach 1...
  - ...and ultimately its expected number of copies will decrease due to destruction by crossover and mutation

The Schema Theorem

- The Schema Theorem predicts that above average fitness schemata proliferate in the population
- *But...* the Schema Theorem only considers crossover and mutation to be *destructive*
  - The *constructive* effects of crossover in particular are ignored
  - This means the lower bounds on transmission probability from the Schema Theorem are not crisp
  - More importantly, the Schema Theorem doesn't actually tell us anything about the positive role that the defining feature of a GA plays, namely crossover
    - Hence another reason why the Building Block Hypothesis doesn't really follow from the Schema Theorem

Price's Theorem

- A more general formulation of selection is available from population genetics

$$\Delta Q = \frac{\mathrm{Cov}(z, q)}{\bar{z}} + \frac{\sum z_i \Delta q_i}{\bar{z}}$$

- where Q is the frequency of some trait in the population, z is individual fitness (i.e. number of offspring produced), q is individual trait value, and Δq_i is change in trait value from parent i to offspring *not due to selection*
- Trait value Q can be any linear combination of alleles in an individual
  - Schema membership is a linear function of all alleles in an individual
  - Hence the Schema Theorem emerges as a special case of Price's Theorem

Price's Theorem

- Price's Theorem requires that for some trait to spread in the population it should be positively correlated with fitness
- So the question is, for a particular trait, would a linear regression of trait value on fitness have a positive slope?
  - This is not a general property for schemata, it will depend heavily on the fitness function

Fisher’s Fundamental Theorem

- Price’s Theorem can also be used to derive Fisher’s Fundamental Theorem
  - *Theorem (Fisher’s Fundamental Theorem):* The rate of increase in mean fitness of a population is proportional to the variance in its fitness due to genetic effects
- So Price’s Theorem and the Schema Theorem can be used to formalise a familiar idea in EC (and artificial and natural selection in general)
  - Genetic variability is required for evolution through natural selection
  - As the population converges, evolution through natural selection slows

Fitness Transmission

- As well as formalising the importance of diversity in EC, Price's Theorem and the Schema Theorem highlight the importance of *transmission* from parents to offspring
  - Schemata are transmitted from parent to offspring, but may be destroyed by crossover and mutation (Schema Theorem)
  - Traits in general are transmitted from parent to offspring, but their value may change during transmission (Price's Theorem)
  - *C.f.* the Breeder's equation $R = sh^2$
- Clearly heritability of a fitness-enhancing trait is important
- One obvious and simple way to measure such heritability is to measure transmission of fitness
  - Perform a linear regression of offspring fitness on parent fitness, and check the slope of the line

Formae and Respect

- Tracking transmission of genotype may not be the most relevant thing to do for EC
- Tracking transmission of phenotype can be much more relevant
- This is the idea behind the principles of *formae* and *respect* encountered in the ‘Grouping Problems’ lectures