

# Using real data sets to simulate evolution within complex environments

Bruce Edmonds, Centre for Policy Modelling, Manchester Metropolitan University

**Target Question:** Does the **complexity** of the environment significantly affect evolutionary processes?

- (here "complexity" means that there are exploitable patterns in the environment but these are difficult to discover)
- Adding randomness to the environment (or the fitness function) is not satisfactory and the NK model of fitness adjusts the difficulty in a uniform manner (second order uniformity)

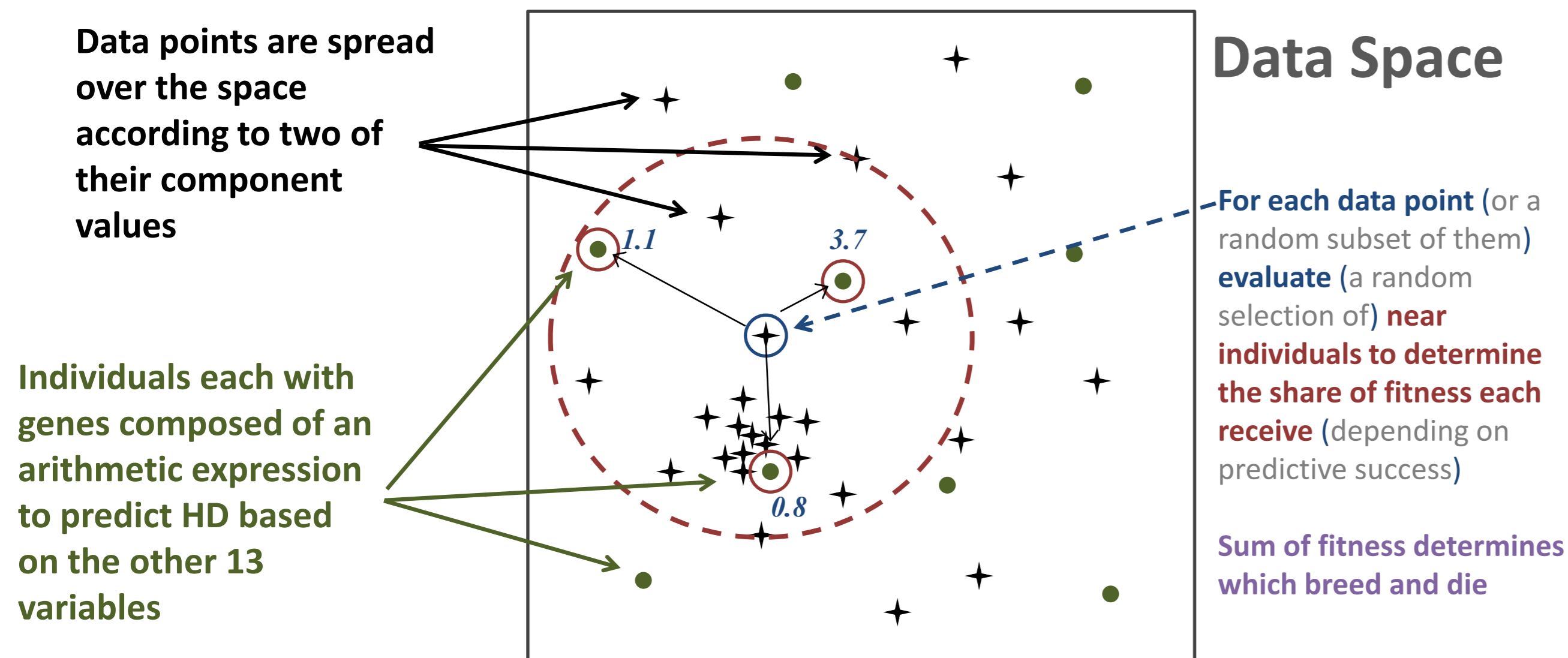
## Main Ideas:

- Evolutionary data-mining is where ideas from biological evolution are applied to data-mining – *finding patterns in data*
- Data sets exist for the purpose of testing different ML algorithms that have patterns in them, albeit difficult to discover
- *Reversing this...* I am suggesting the use of complex data sets as a **test bed** to investigate how the complexity of the environment might affect evolution

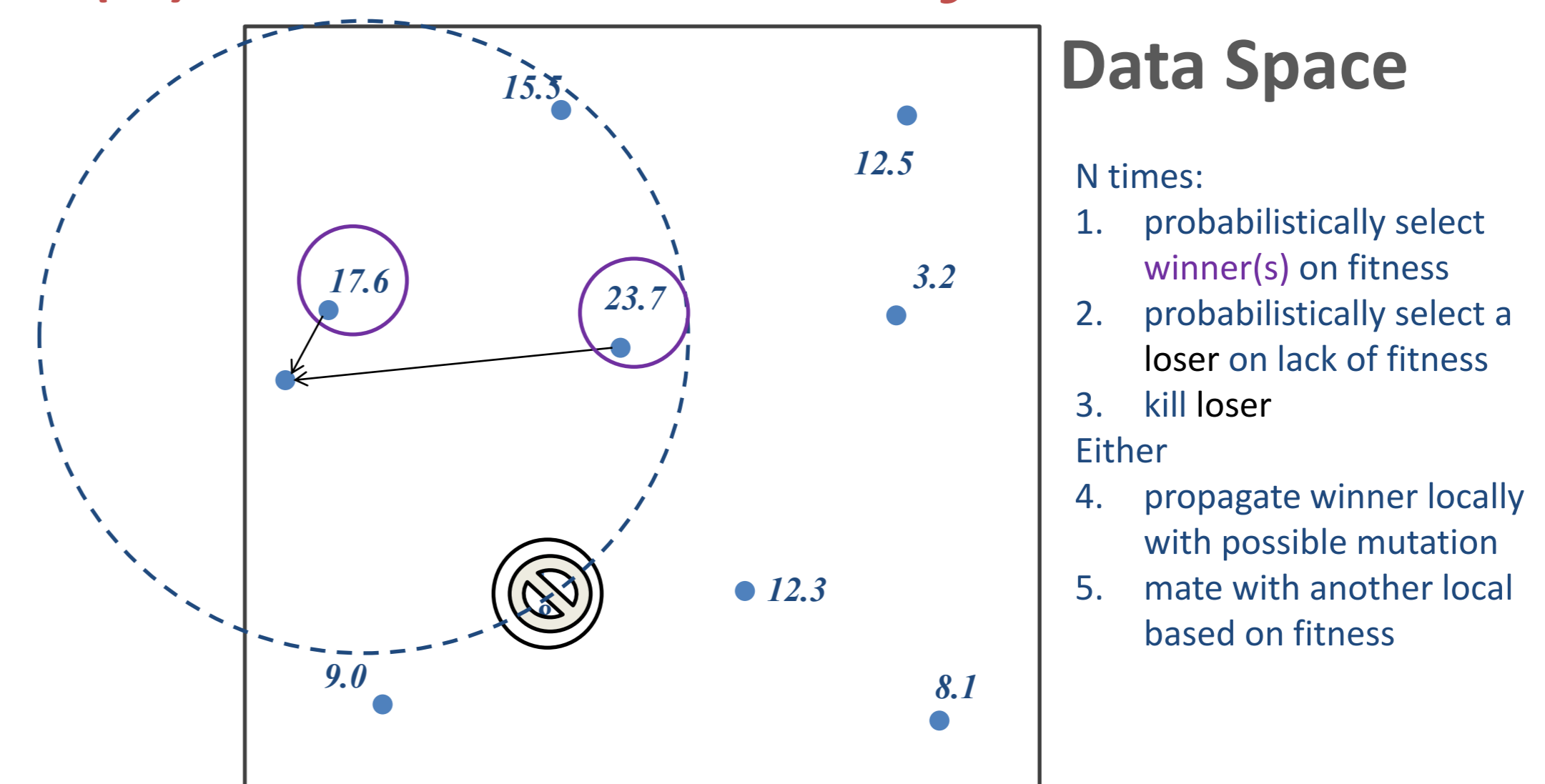
## The Data Set Environment:

- Find a rich data set (preferably one derived from a naturally complex system) with many independent variables
- The gene of an individual is an arbitrary arithmetic expression stored as a tree (or similar technique)
- Resource in the model is modelled by distributing to individuals predicting the outcome variable of *local data* better than its competitors
- The gene are mutated and crossed as the simulation progresses
- Individuals are selected for/against depending on their total success in predicting

## (1) Evaluation Phase of Model:



## (2) Selection Phase of Model:



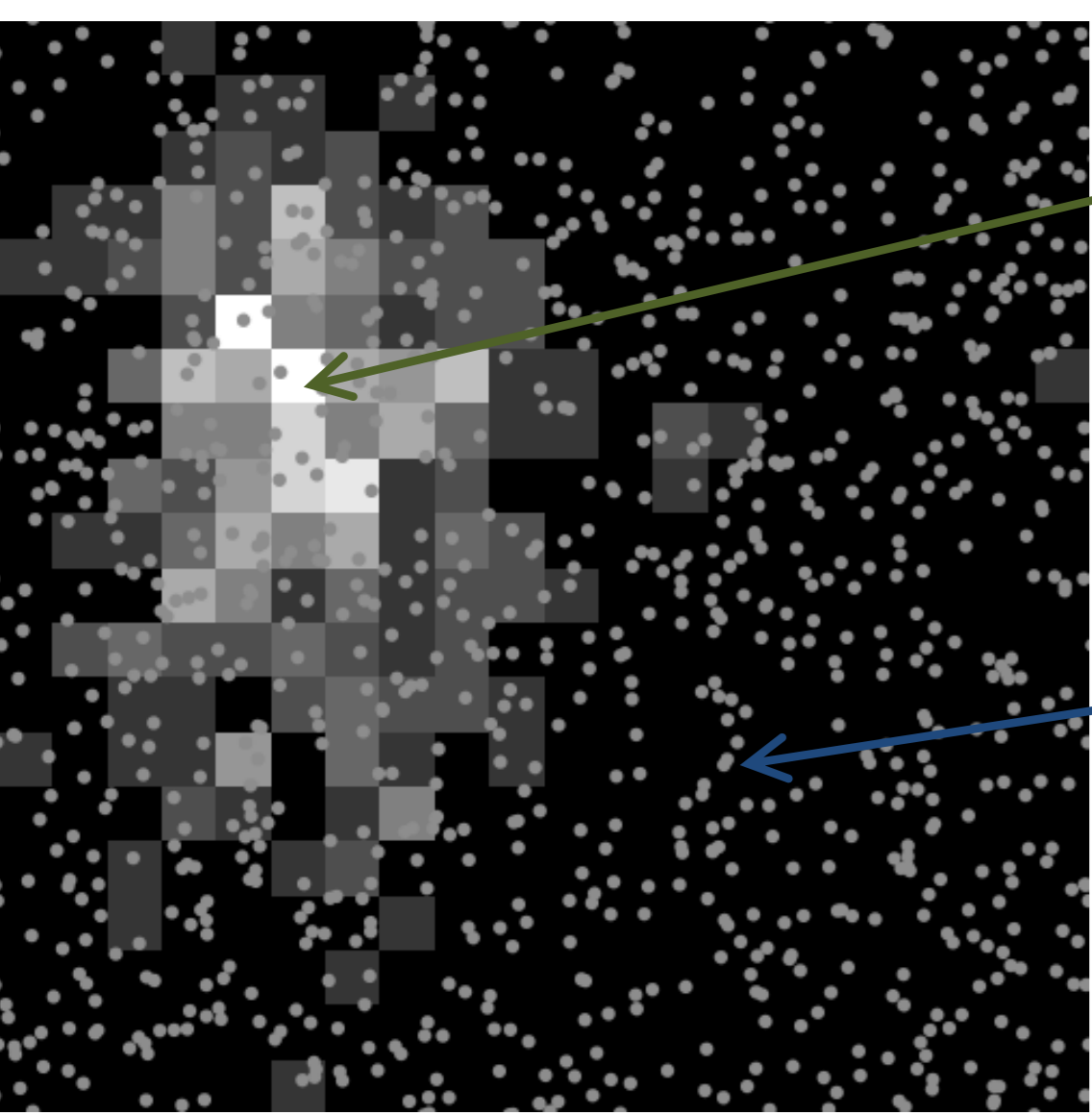
## An example simulation run using the HD data set:

Start of Simulation (HD Data)

Simulation – 25 ticks

Simulation – 50 ticks

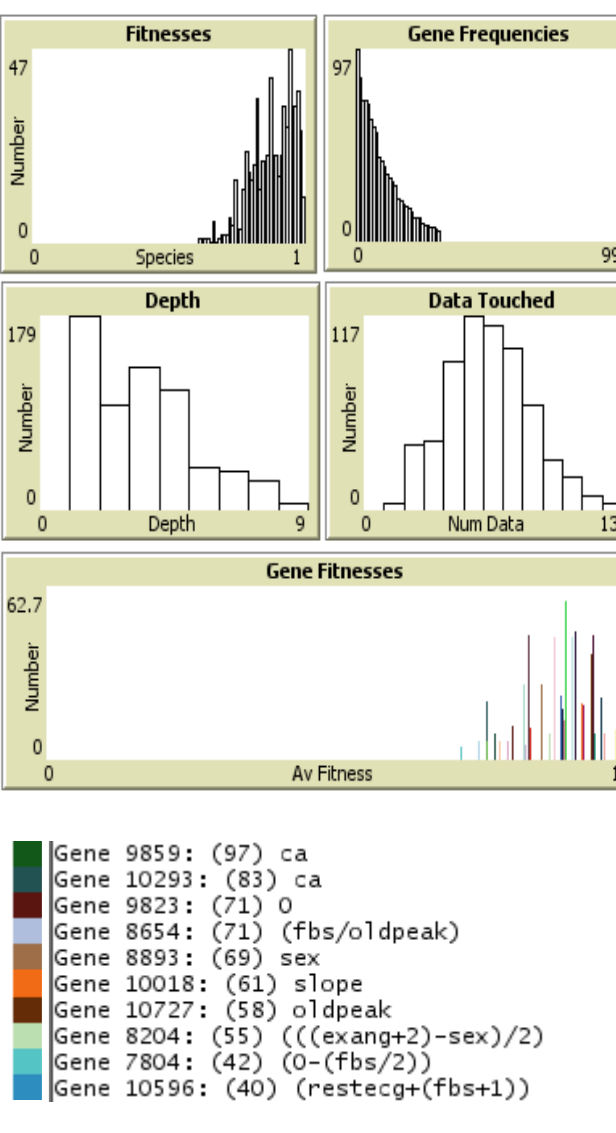
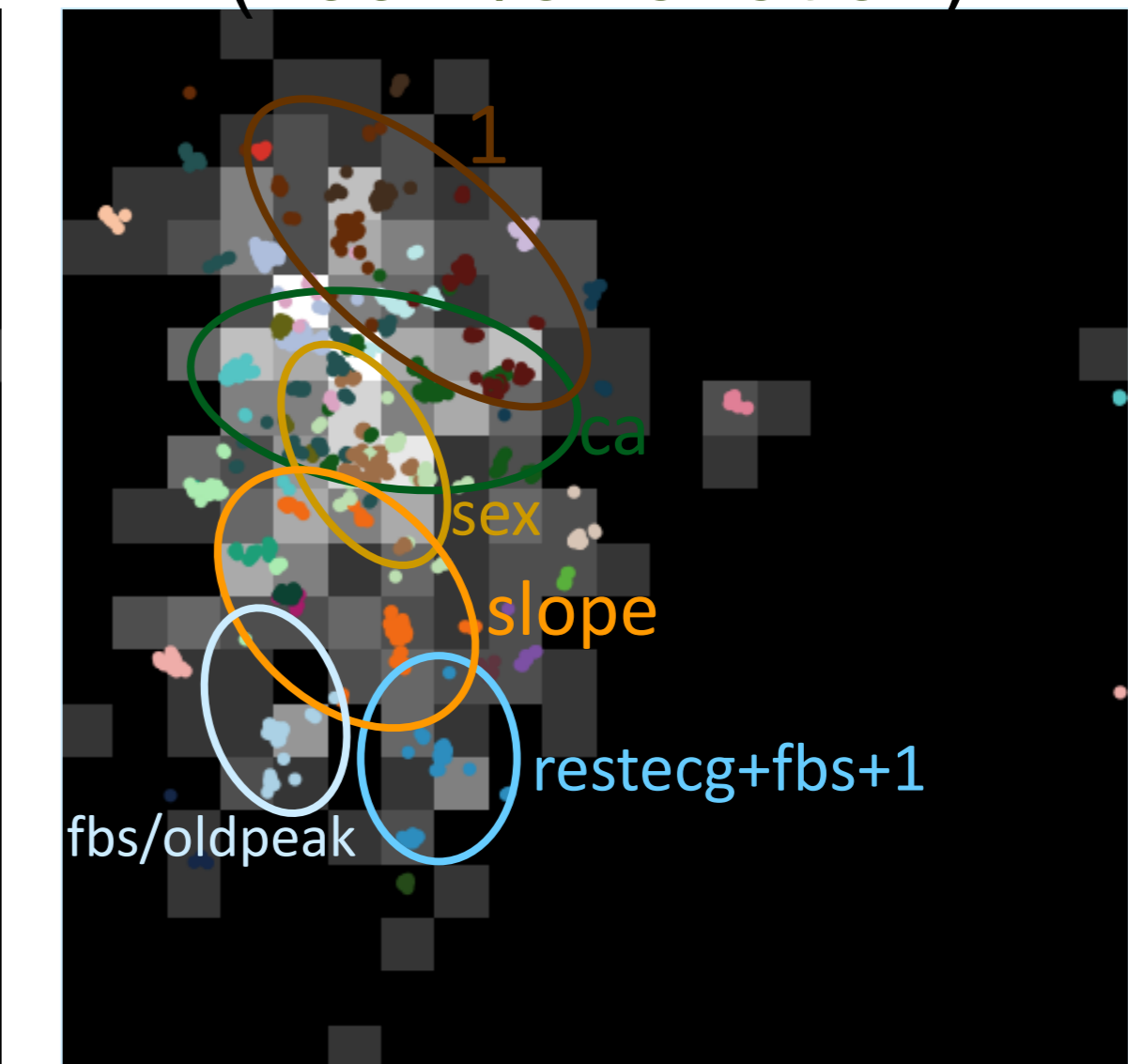
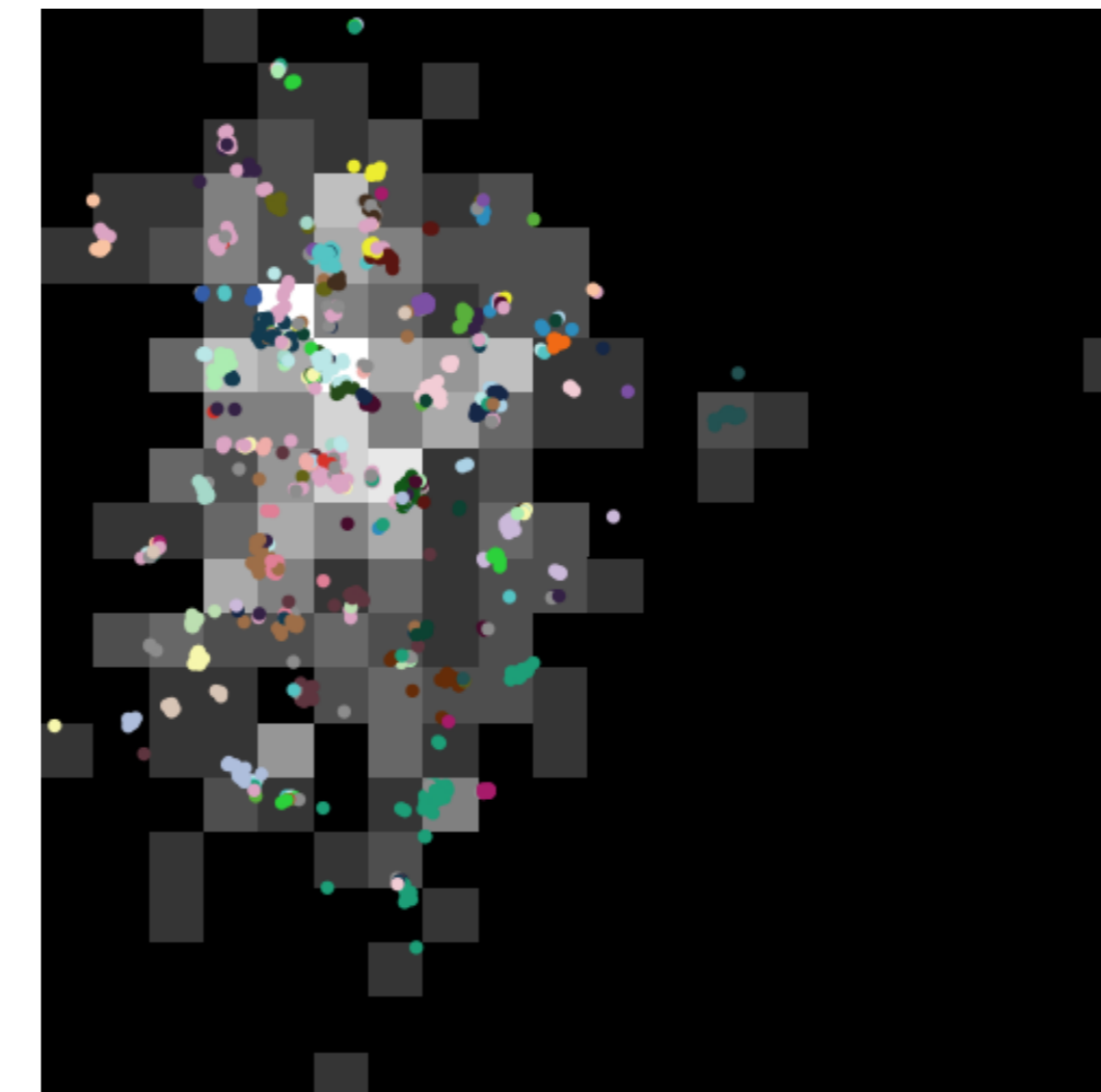
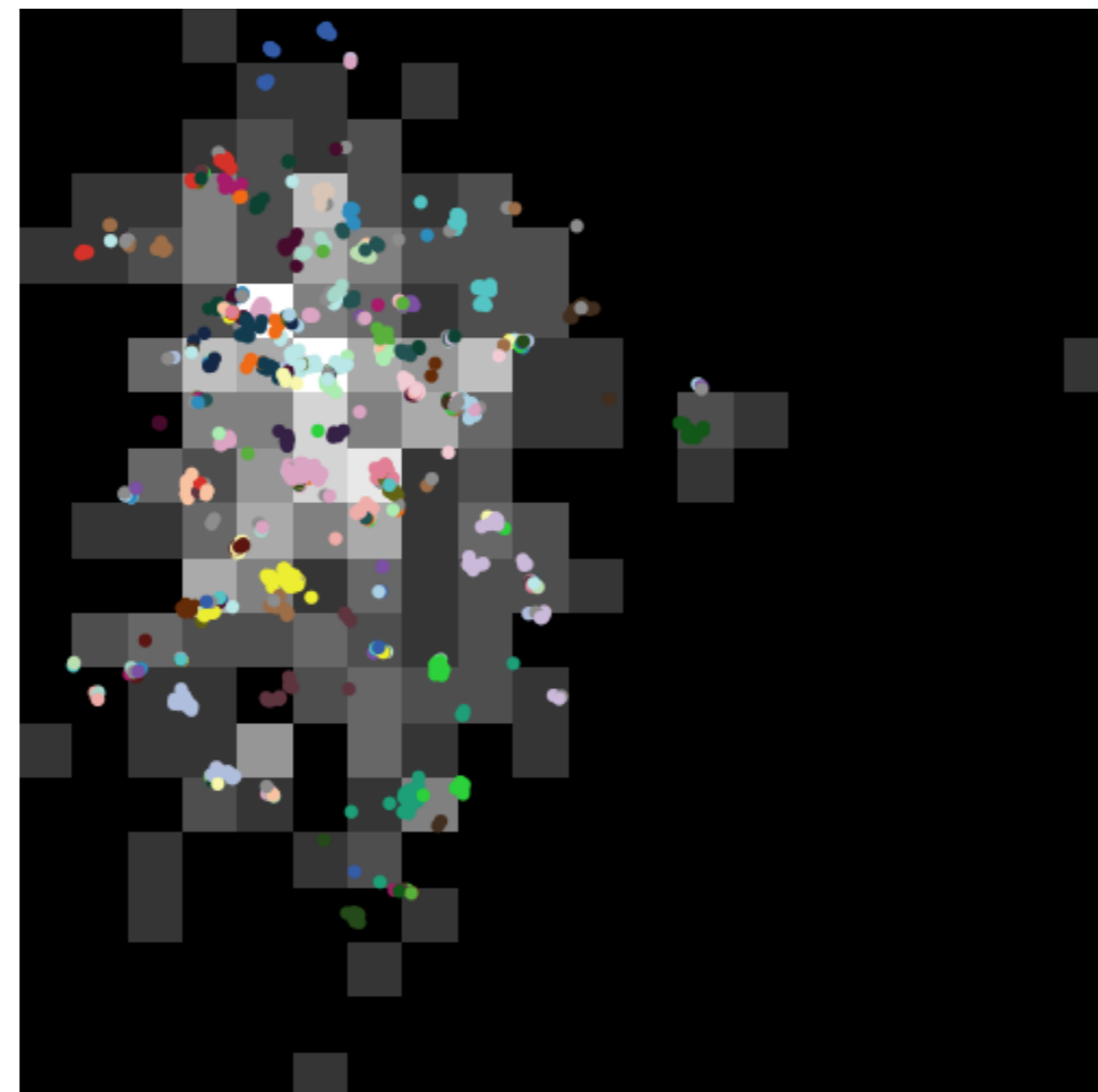
Simulation – 300 ticks (100 w/o variation)



Data points from set distributed over space dependent on 2 variables chol (x) & thalach (y)

Individuals each with gene which is an arithmetic expression, e.g.:

```
Gene 516: (1) ((restbps/9)/exang)
Gene 864: (1) ((ca*restbps)/0.5)
Gene 1518: (1) ((0*ca)+thalach)
Gene 1950: (1) ((cp/3)*(1-age))
Gene 1274: (1) ((0-exang)/5)
Gene 1222: (1) ((20+slope)*11)
Gene 1378: (1) ((restcpg-100)-exang)
Gene 1460: (1) ((tal+age)
Gene 1986: (1) ((0+cp)*(100/fbs))
Gene 1596: (1) ((0/01slope)*(tal/01slope))
```

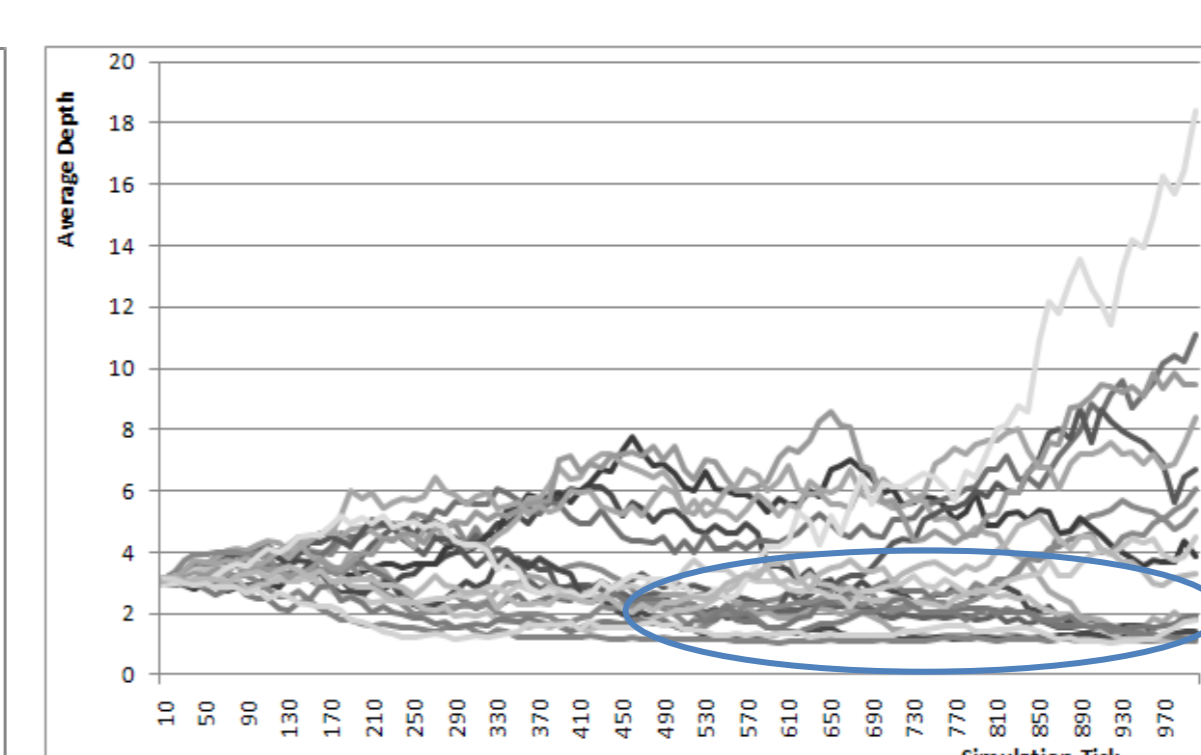
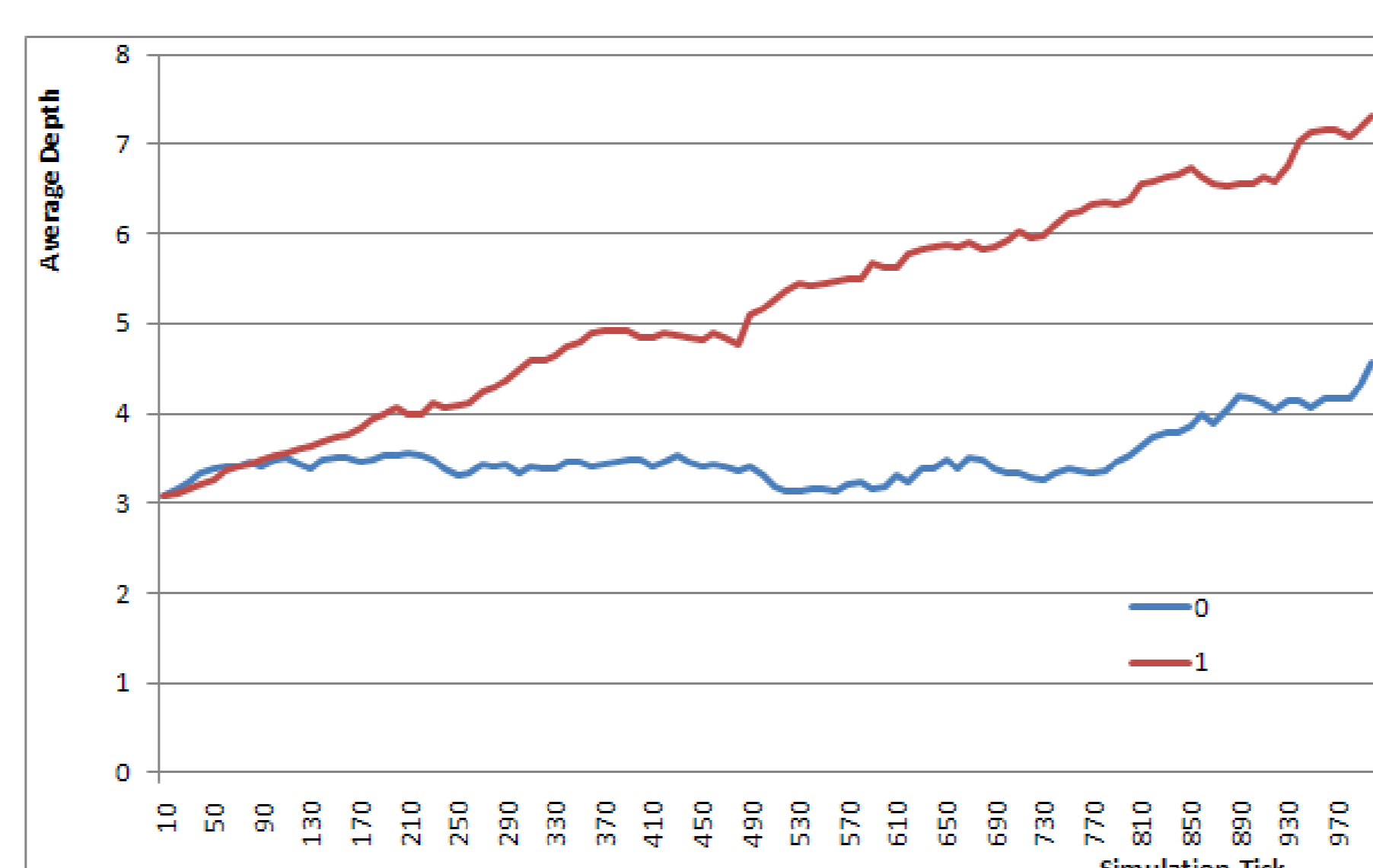
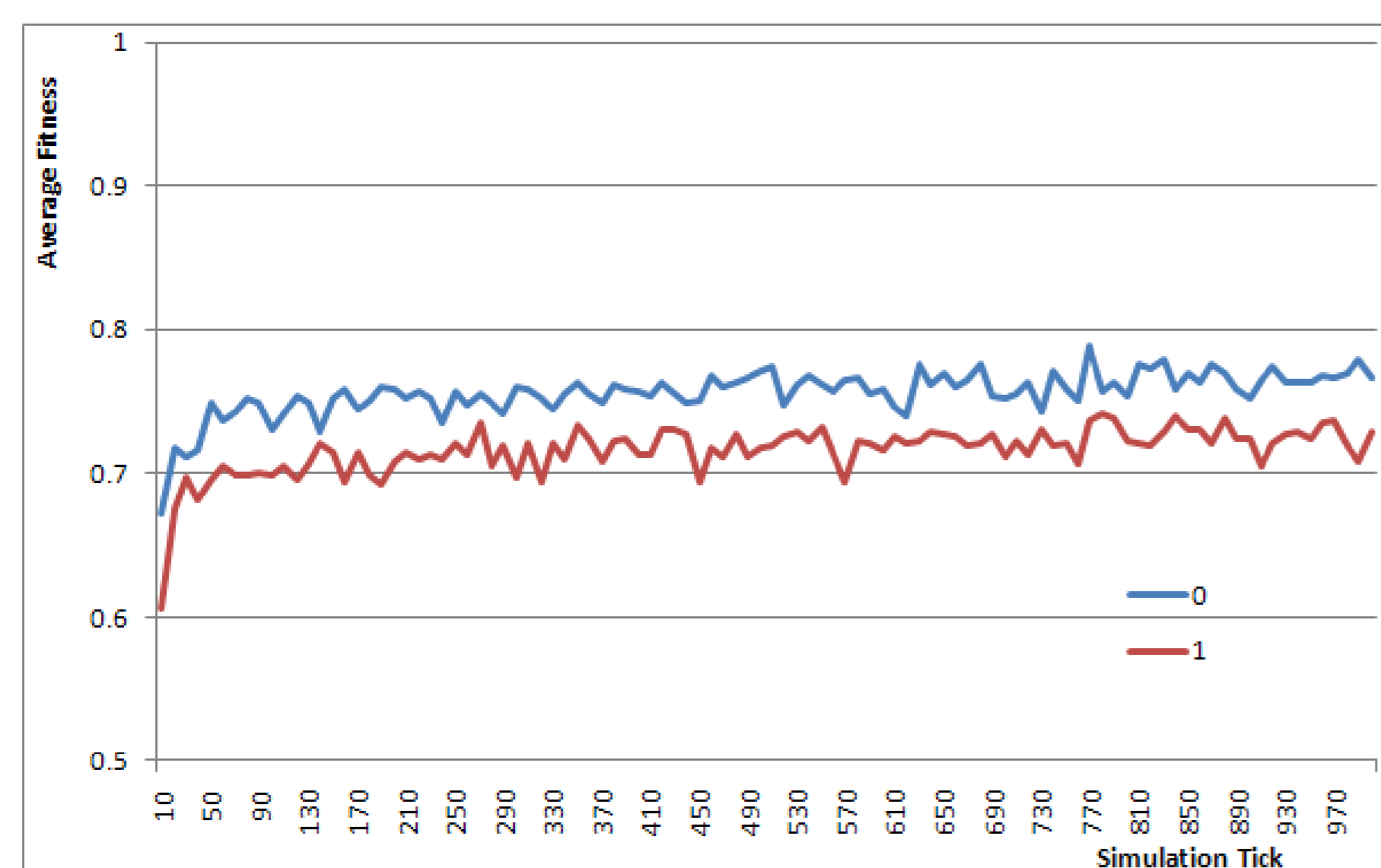


## Illustrative Results:

Using Reduced Cleveland Heart Disease Data Set, 20 runs with each setting, 1000 individuals, 1000 iterations, Locality parameter 0.1 (radius), Comparison of **Original** vs **Ersatz** Data Sets (same distribution in each dimension but random), Fixed normal noise (0, 0.1) added to both data sets

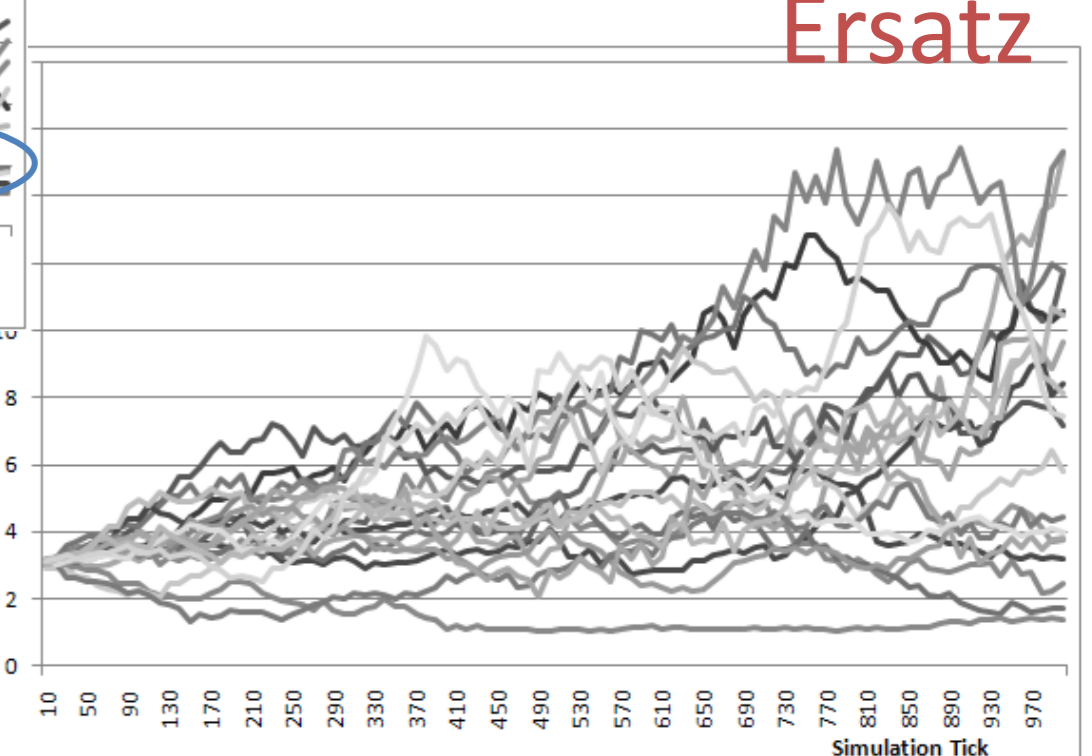
Fitness

Gene Complexity/Depth



Original

Ersatz



## Questions:

That might be explored using this approach

When might the complexity of the environment effect evolutionary processes?  
 How might the complexity of the environment effect evolutionary processes?  
 Will models with a simple environment tell us about evolution in the wild?  
 • When and about what aspects will models with simple environments be sufficient?  
 • In what ways might evolution differ when in complex environments?  
 What kind of complexity might we need?  
 How might one measure this complexity in the wild (if this is even possible)?