Taxonomy and Coarse Graining in Evolutionary Computation

A Tale of Elephants, Blind Men and Soup!

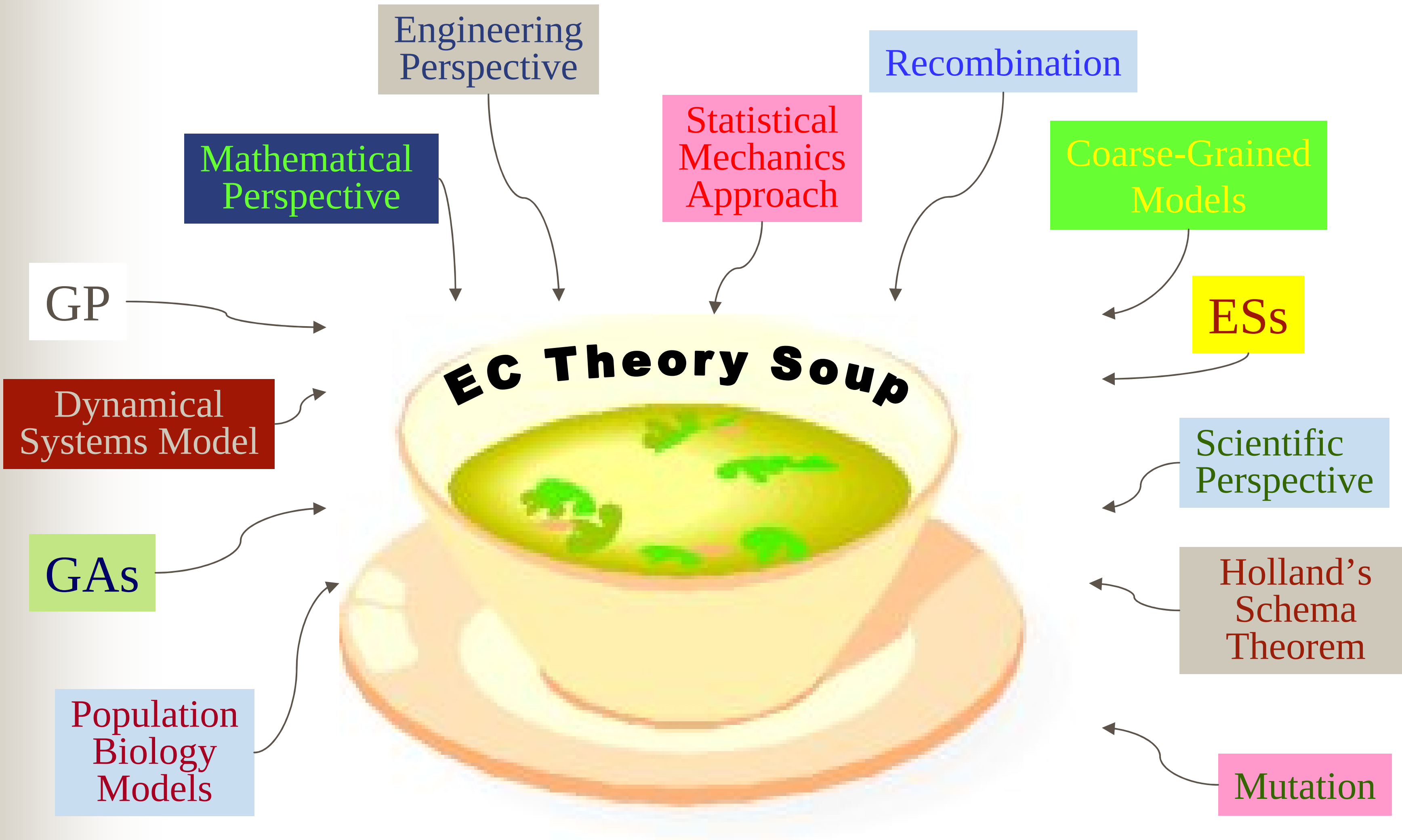
Chris Stephens,
Instituto de Ciencias Nucleares, UNAM
Tutorial GECCO2004, 26/06/2004
stephens@nuclecu.unam.mx

WARNING!

EQUATIONS

G Slides suitable for a general audience
A Slides suitable only for those accompanied by someone with knowledge of arithmetic and elementary linear algebra
AA Slides only for those not shocked by a coordinate transformations and giving new names to things that previously had other names
X No X rated slides – no mention of the words “Theorem”, “Lemma” or “Proof”

ALL CLEAR!

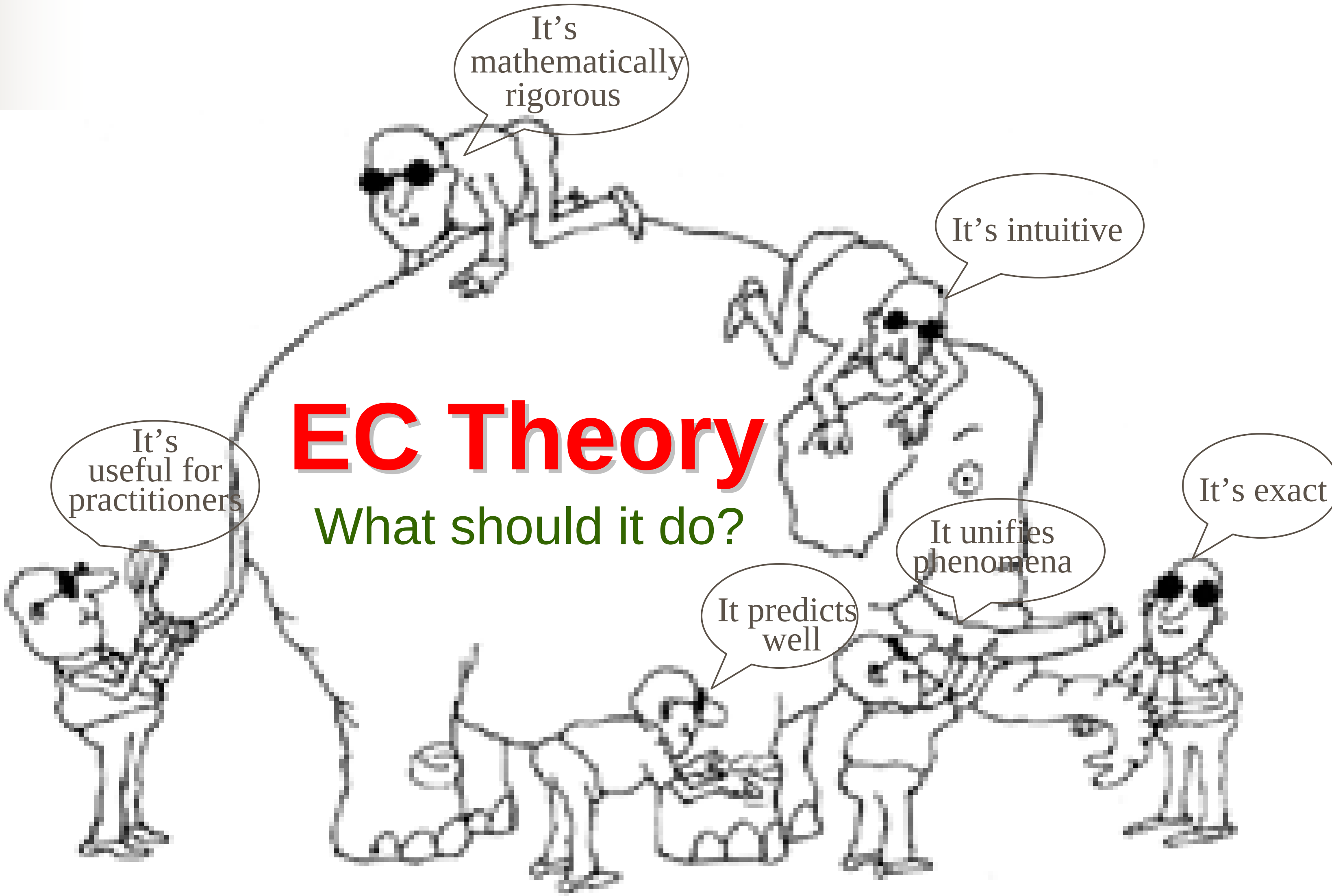

EC Theory Soup
Engineering Perspective
Recombination
Statistical Mechanics Approach
Mathematical Perspective
Coarse-Grained Models
GP
ESs
Dynamical Systems Model
Scientific Perspective
GAs
Holland's Schema Theorem
Population Biology Models
Mutation

GA with no selection, 1-pt crossover and mutation
GA with 1-pt crossover and unitation landscape
GA with 2-pt crossover and unitation landscape
GA with 2-pt crossover and unitation landscape and mutation
.
.
GA with uniform crossover and weak epistasis
GA with uniform crossover and “quasilinear” fitness
GA with uniform crossover and NK-landscape
.
.
.
GA with inversion and n-city TSP landscape
.
.
GA for a 555-job job-shop scheduling problem with three point crossover with probability 0.9 and mutation probability 0.015
.
.
GA for the Multi-Resource Traveling Gravedigger Problem with Variable Coffin Size

GAs

Taxonomy

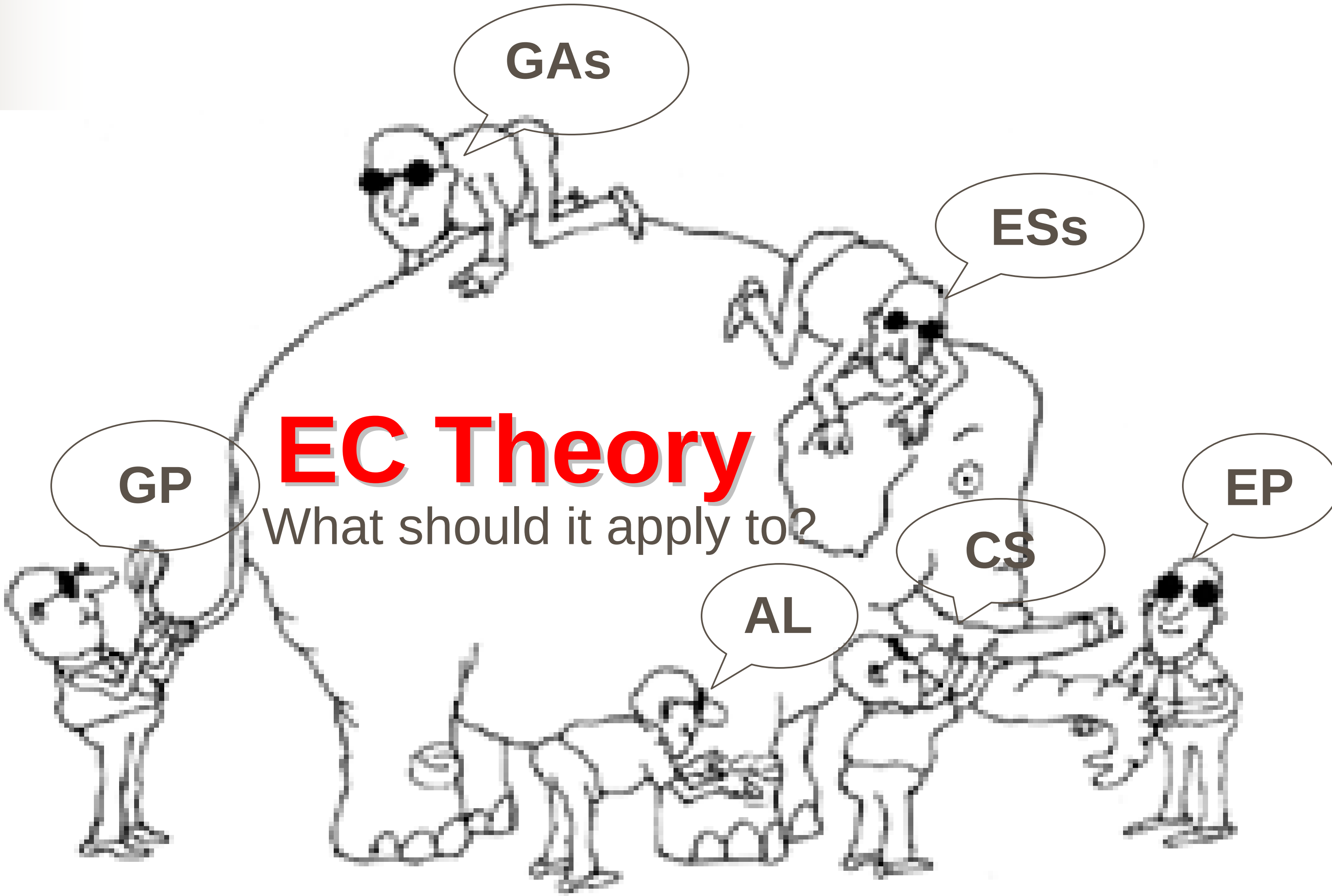
The Problem of Taxonomy…



EC Theory

What should it do?

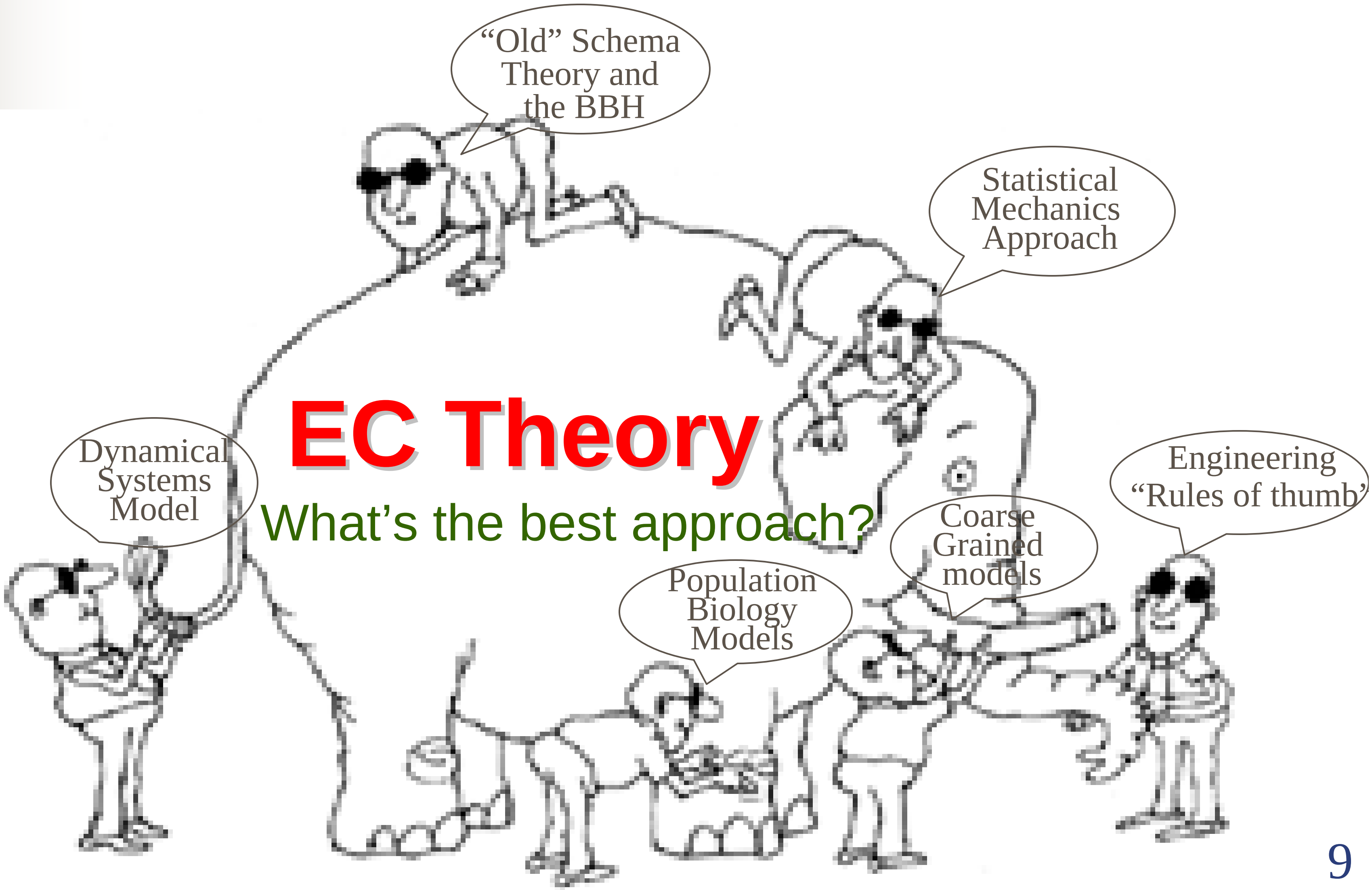
The Problem of Taxonomy…



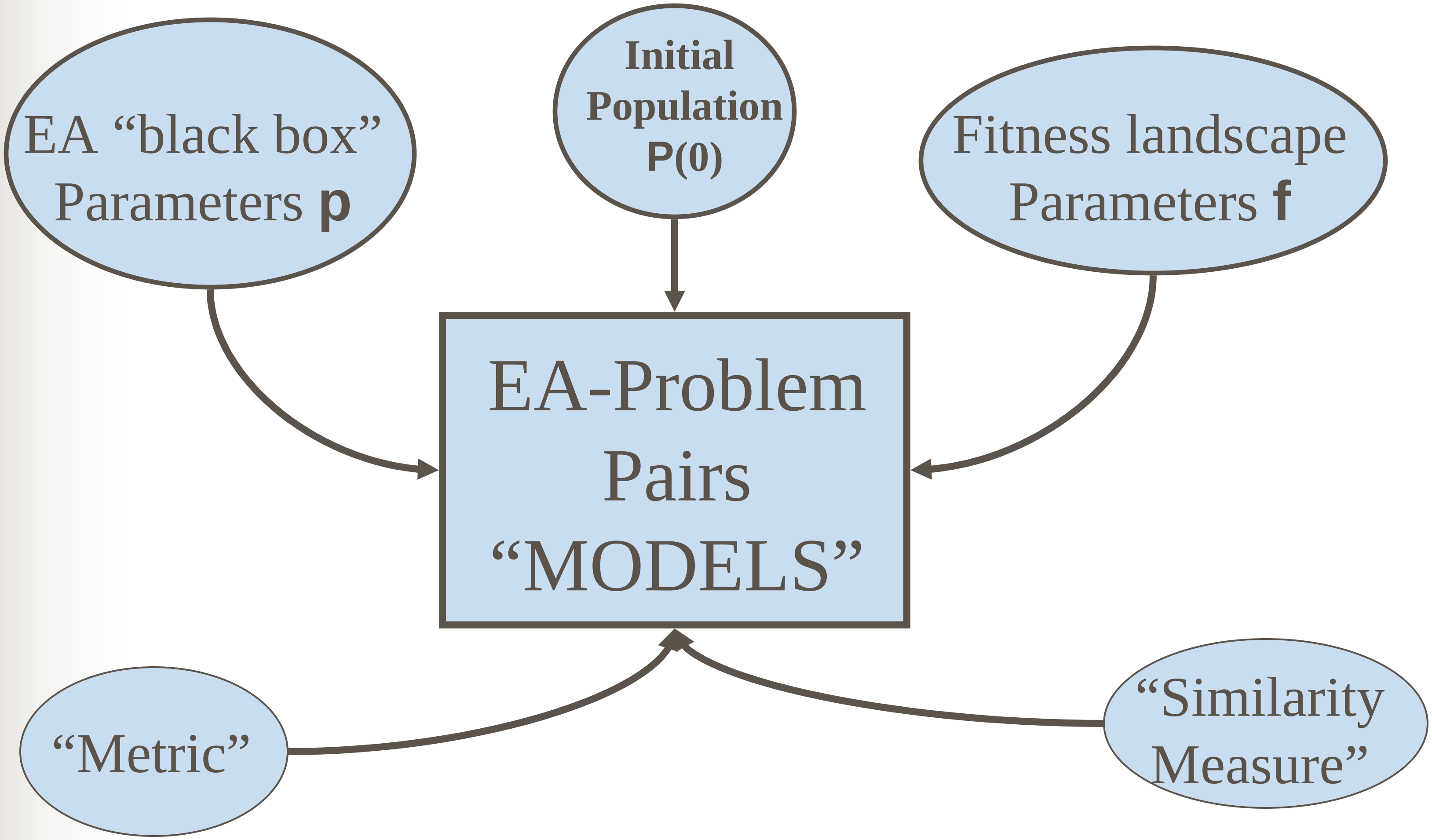
EC Theory

What should it apply to?

The Problem of Taxonomy…



The Problem of Taxonomy…



The Space of EAs

How “far apart” are a GA with one-point crossover, pc=0.8, mutation, p=0.08, and a NK fitness landscape, N=27, K=3 and GP for K-SAT, K=4, sub-tree crossover, pc=0.5, mutation, p=0.05?

How “far apart” are a giraffe and a grasshopper?

How “far apart” are hydrogen and uranium?

How “far apart” are a GA with one-point crossover, pc=0.8, mutation, p=0.08, and a NK fitness landscape, N=27, K=3 and a GA with one-point crossover, pc=1, mutation, p=0.1, and a NK fitness landscape, N=35, K=3?

How “far apart” are a giraffe and a horse?

How “far apart” are sodium and potassium?

Taxonomy is easier with “distance” measures

Taxonomy

Theory – what can it tell us? E.g. “electronic structure”

Phenomenology – we want more of that! e.g. “periodic table”

History – contingency, that’s what we’ve had

Universality

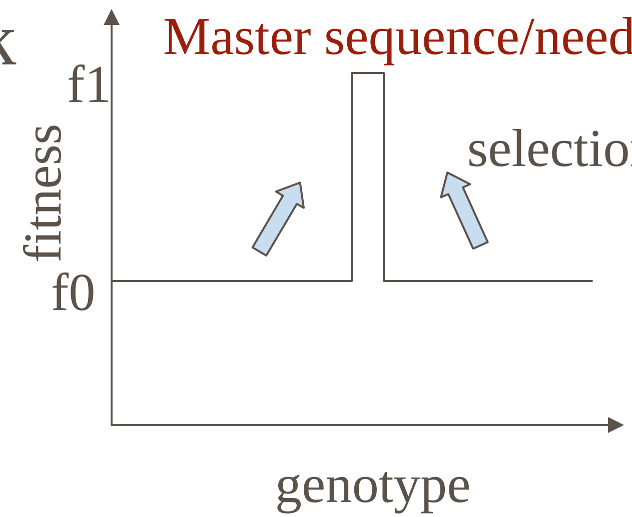
Phenomenology

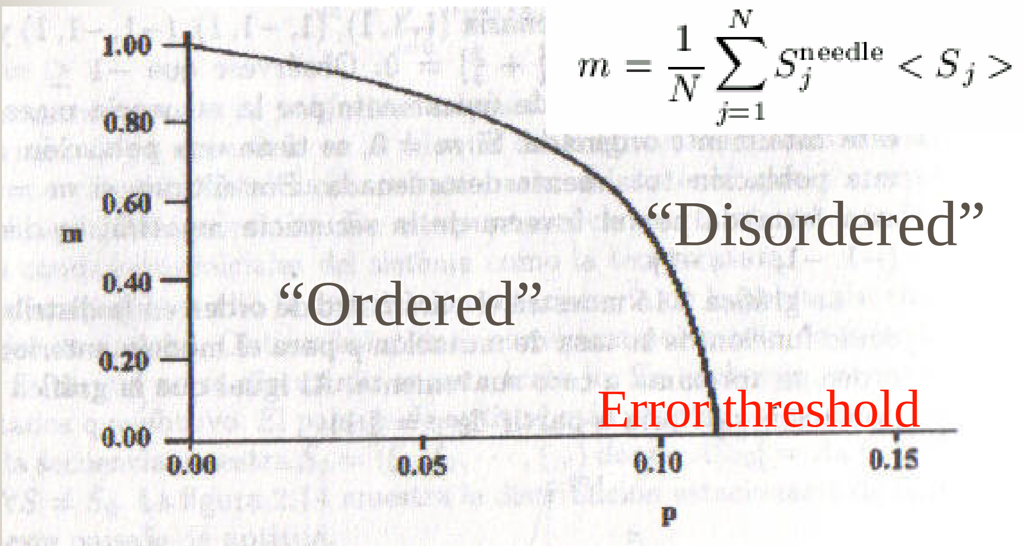
Universality – when is the devil in the details?

Eigen model/Needle-in-a-haystack

Characteristic of viruses and real world BRITTLE problems (it works or it doesn't't!)

Qualitative behavior dominated by existence of error threshold – doesn't depend on "details" - **UNIVERSAL**





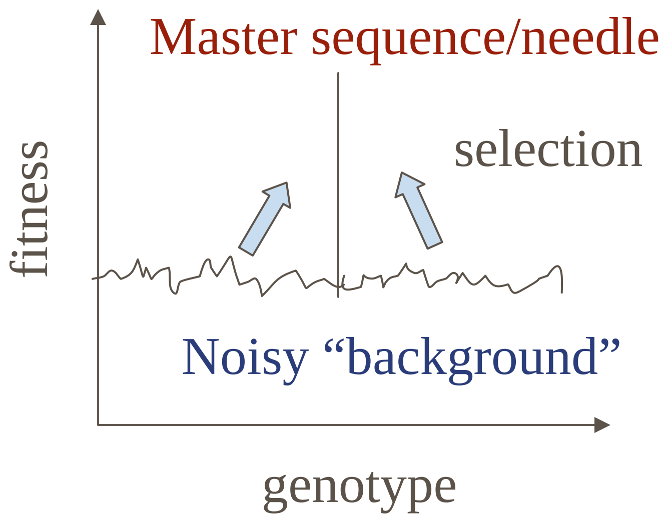
$$m = \frac{1}{N}\sum_{j=1}^{N} S_j^{\text{needle}} < S_j >$$

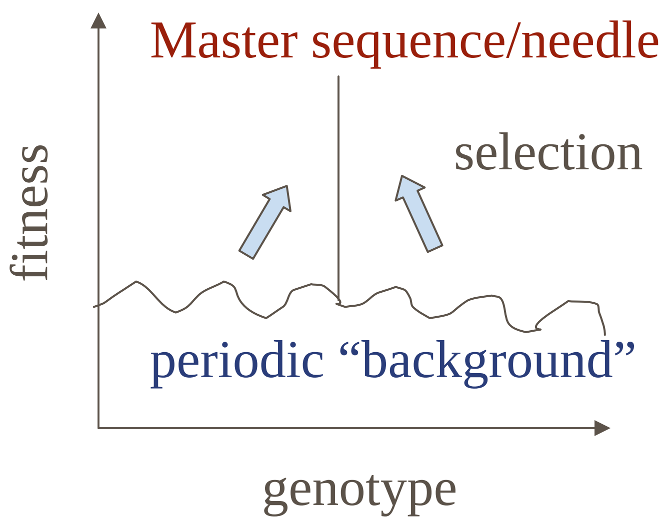
Value of critical mutation rate does depend on details (N, f1, f0 …)

– **NON-UNIVERSAL**

Universality – when is the devil in the details?

Same **UNIVERSALITY CLASS** as NIAH

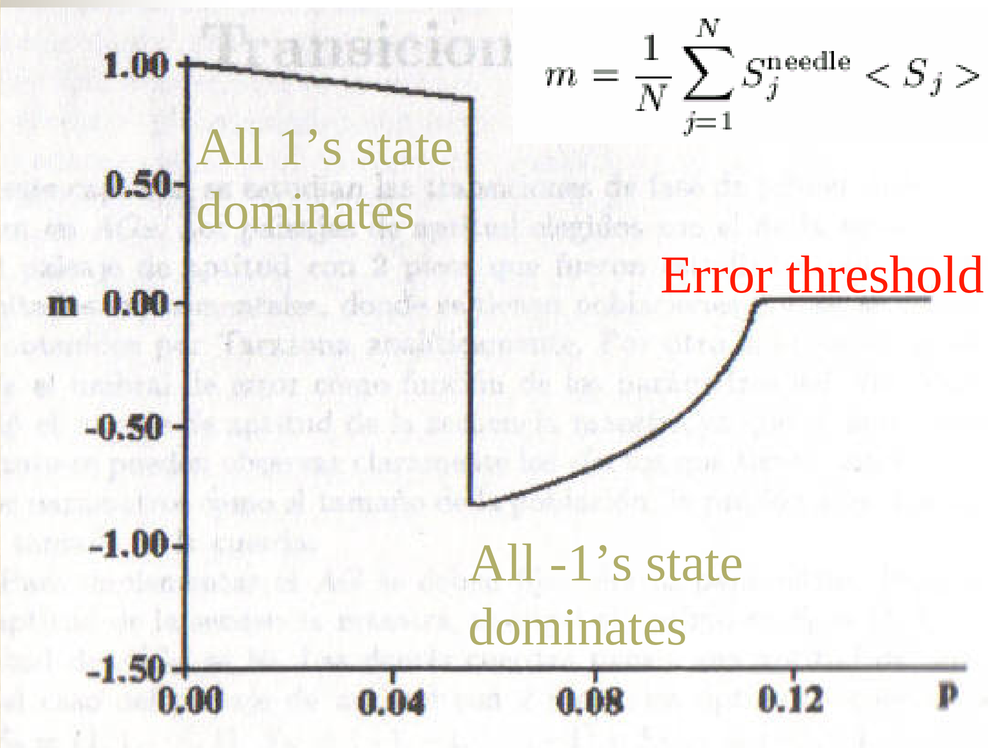


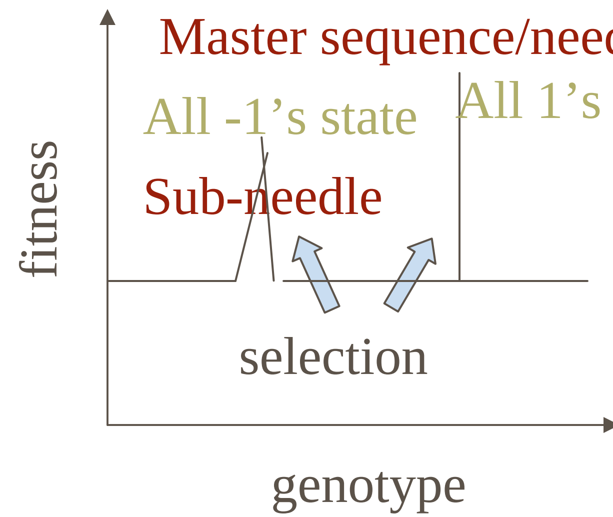


Universality – when is the devil in the details?

NOT in the same **UNIVERSALITY CLASS** as NIAH!

$$m = \frac{1}{N}\sum_{j=1}^{N} S_j^{\text{needle}} < S_j >$$



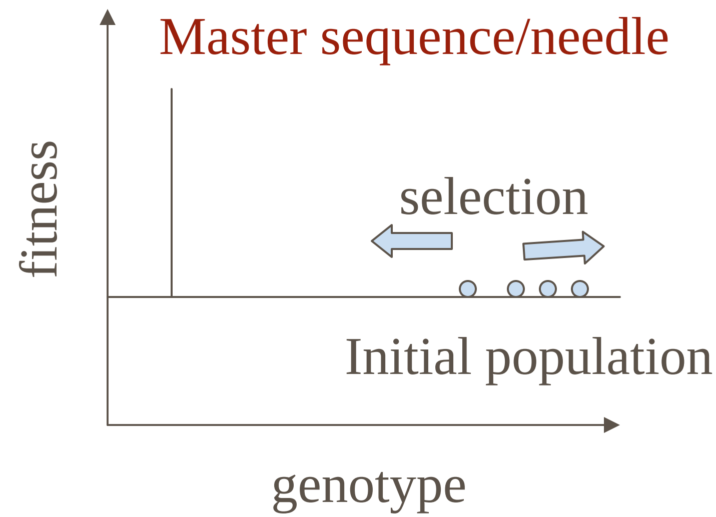


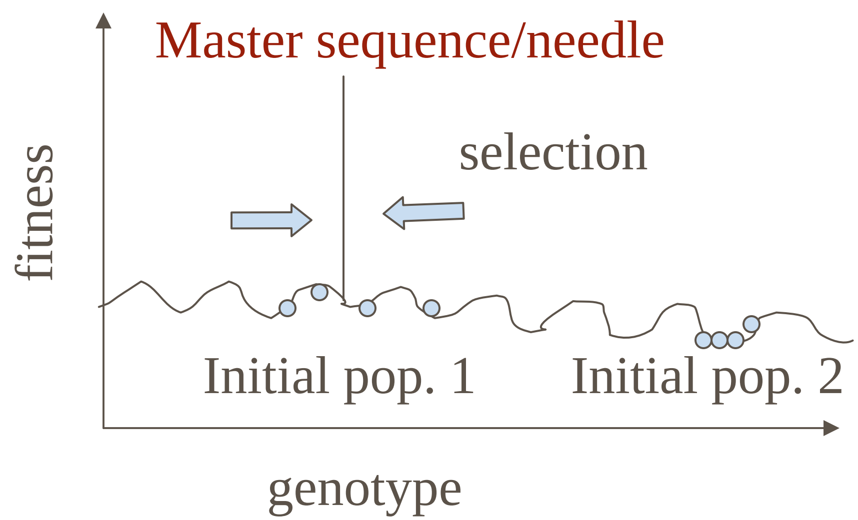
Corresponds to a system where there's several "it kind of works" states as well as a "it definitely works" state

Universality – when is the devil in the details?

Same **UNIVERSALITY CLASS** as NIAH? YES

What typically happens?





Universality – when is the devil in the details?

Phase transition for K-SAT

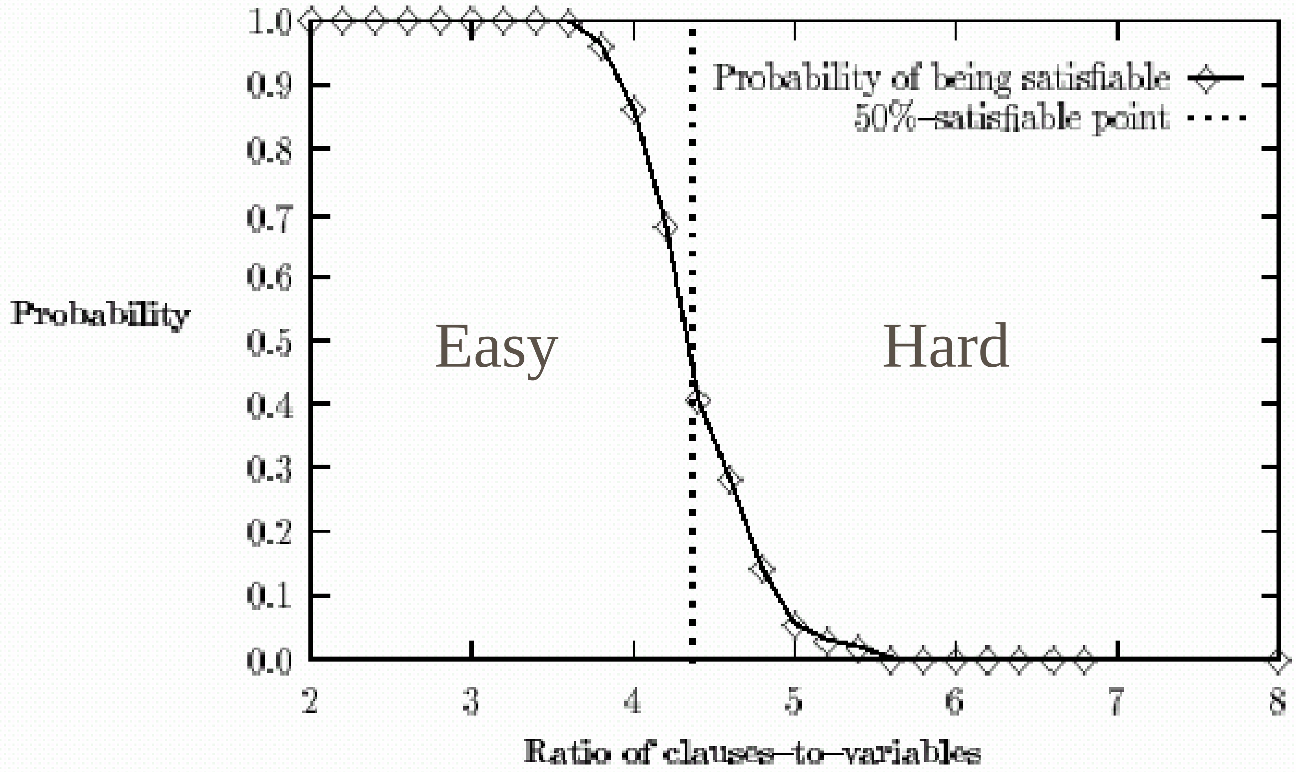


Figure 4: Probability of satisfiability of 50-variable formulas, as a function of the ratio of clauses-to-variables.

From: Mitchell et al.(1992) Hard and easy distributions of SAT problems

Similarity measures

Need objective criteria by which to judge the degree of affinity between different models. There are many possibilities…e.g.

- Average population fitness vs. time
- Best in population versus time
- Diversity versus time
- "Order parameter" (e.g. % of population that is optimal as function of EA parameters)
- Time to find optimum
- "Hardness"
- "Robustness"
- Fixed points (asymptotic behaviour)

Example of a similarity measure

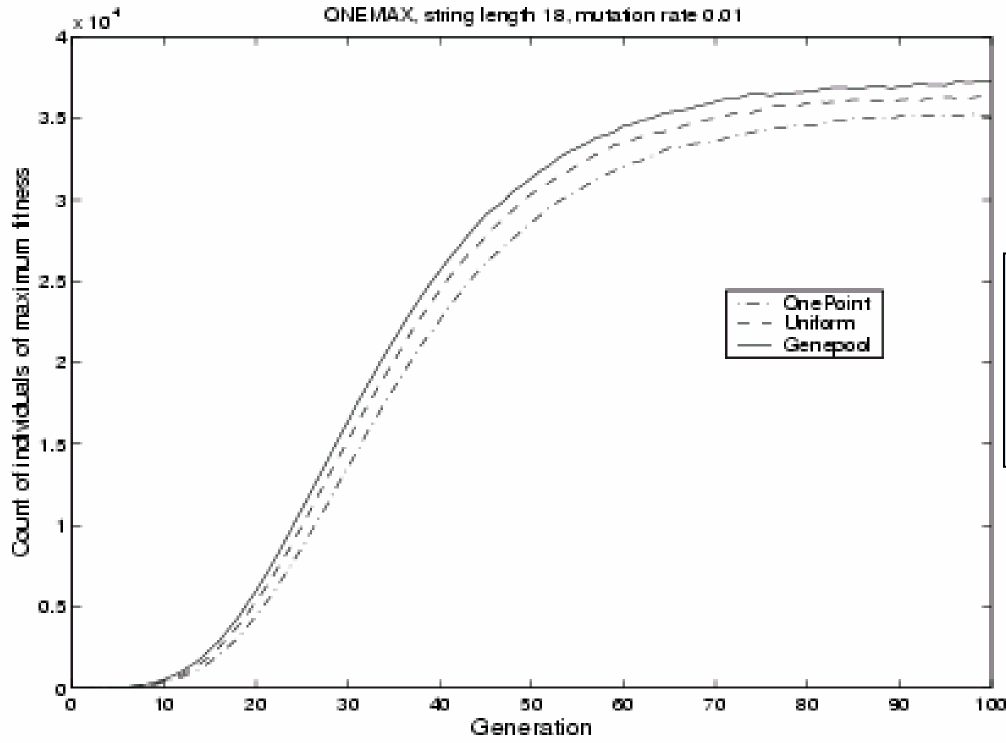


Figure 2: The number of optimal individuals for different types of recombination

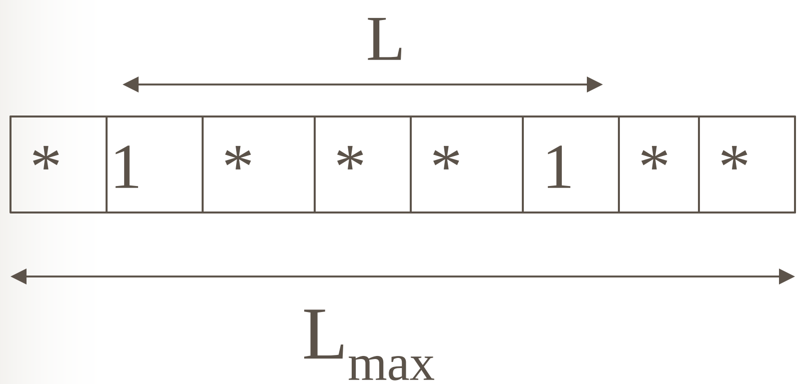
< 10% difference between One-point and genepool

"Model"/"Toy" problems tell us much more than you think!

Example of a similarity measure

Stephens, Waelbroeck and Aguirre – FOGA 7

$$M(L) = (n_{opt}(L) - n_{opt}(L_{max}))/ n_{opt}(L_{max})$$



$n_{opt}(L)$ = no. of optimal 2-schemata/total no. possible per string

Interested in whether short or long building blocks are preferred. $M(L) > 0$ ⇒ preference for short blocks, $M(L) < 0$ ⇒ preference for long blocks

Example of a similarity measure

Experiment: popsize = 5,000; L_{max} = 8;   30 runs

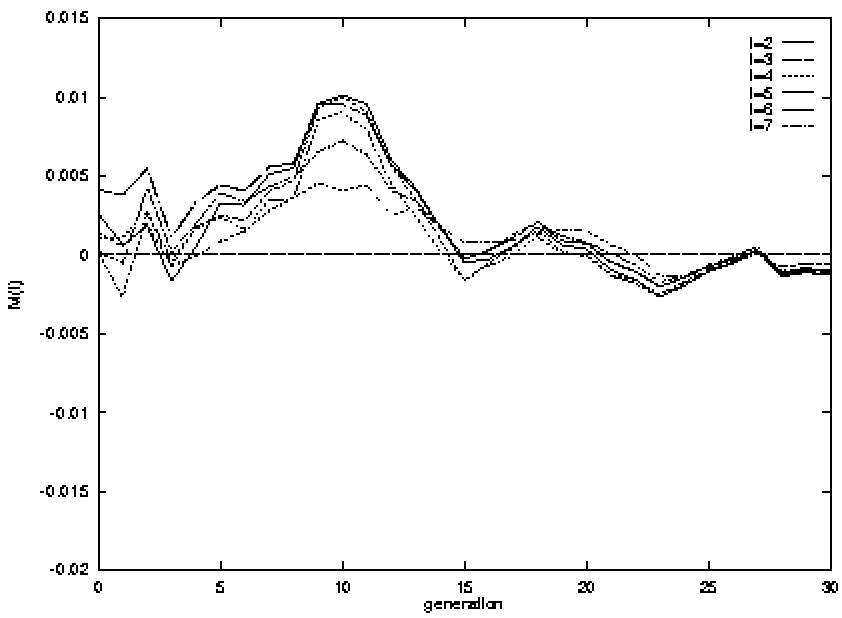


Figure 1: Graph of $M(l)$ versus t in the unitation model with $p_c = 0$.

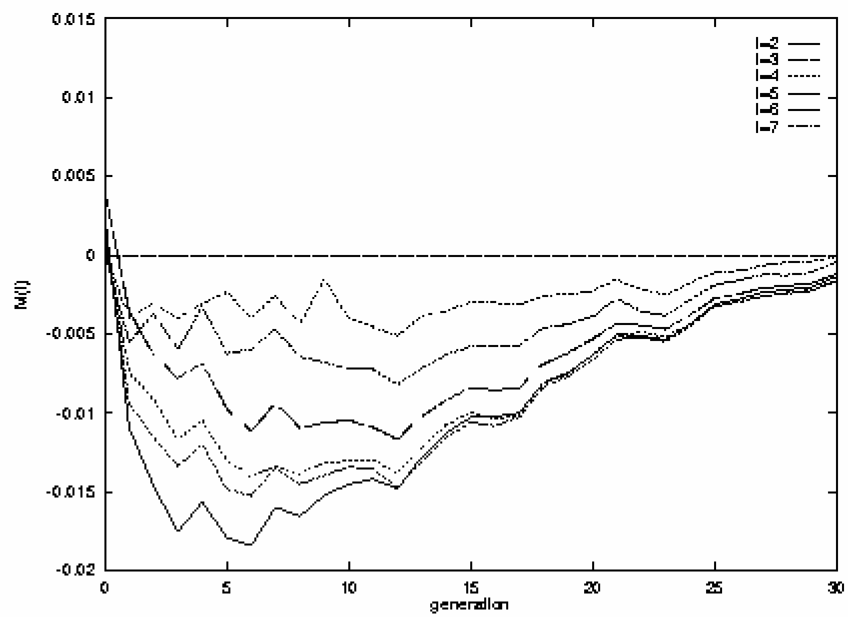


Figure 2: Graph of $M(l)$ versus t in unitation model with $p_c = 1$.

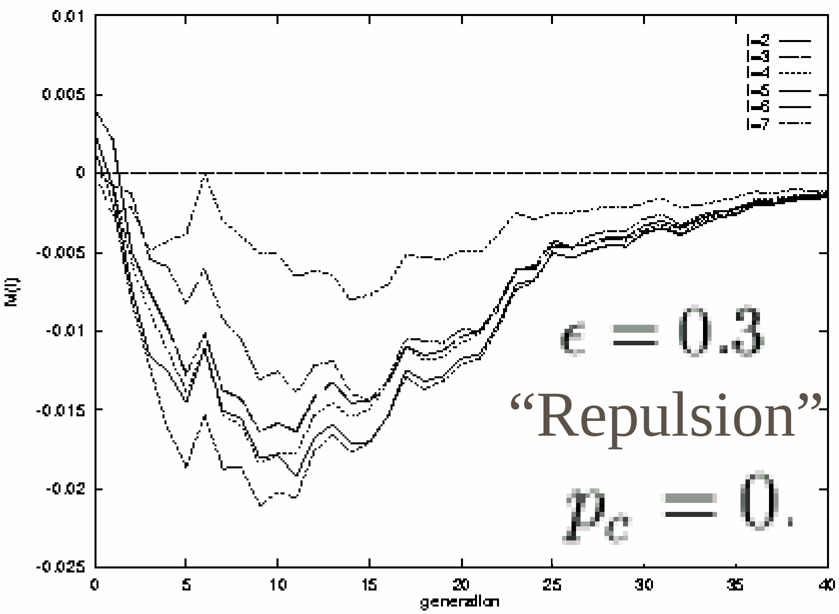
Without crossover – no preference for one size versus another

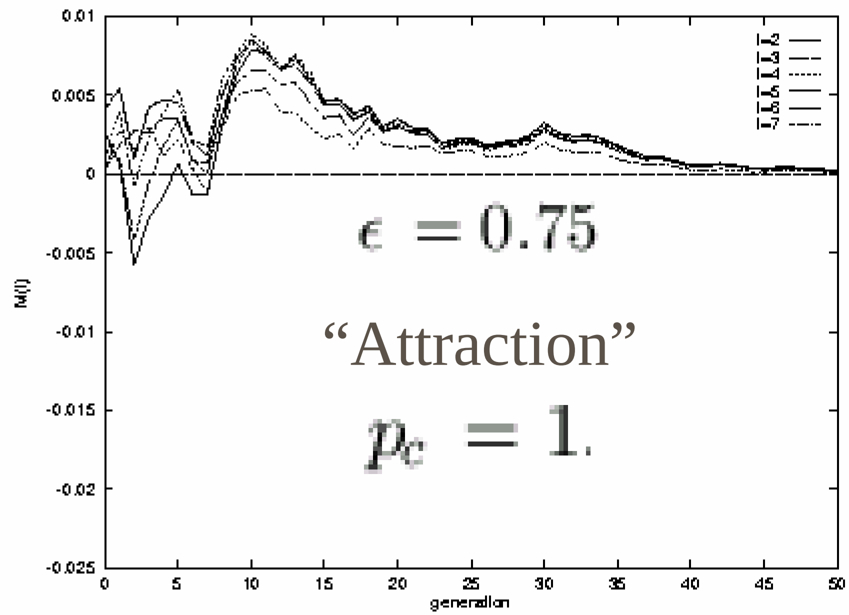
With crossover – large schemata grow, short schemata diminish – opposite of Building Block Hypothesis

Example of a similarity measure

$$f(c_i) = \sum_j 1_j + \frac{\epsilon}{N^{\pm}} \sum_{jk \in C_i} l_{jk}^{\pm 1}$$

Add pair epistasis: + →repulsion;
- →attraction





Results for $p_c = 1.$ with no epistasis are similar to those with $p_c = 0.$ and an epistatic repulsion between bits

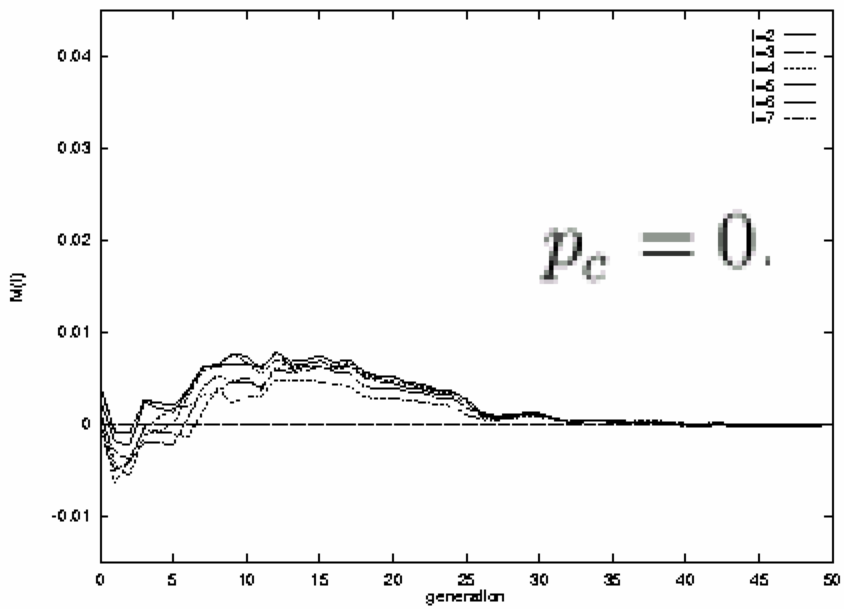
Results for $p_c = 0.$ with no epistasis are similar to those with $p_c = 1.$ and an epistatic attraction between bits

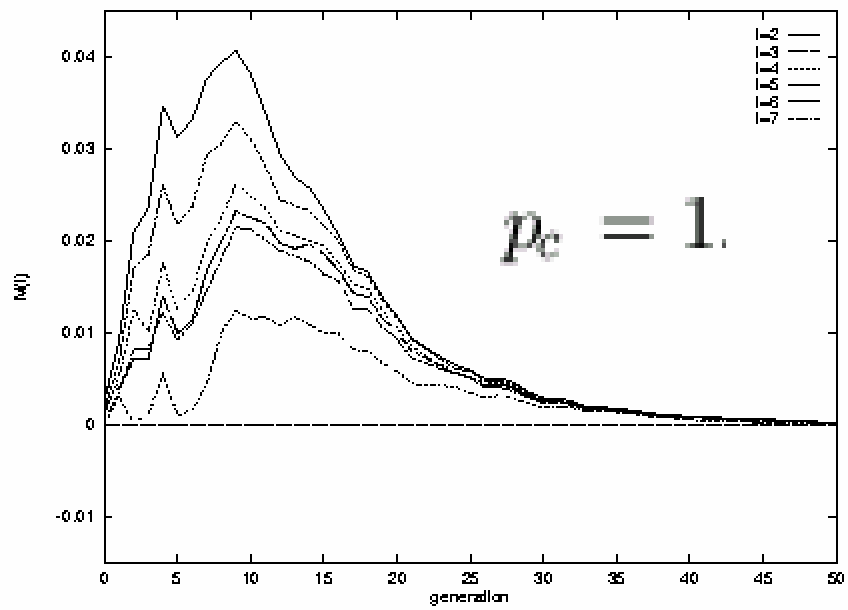
UNIVERSALITY

Example of a similarity measure

"Deceptive" landscape

$f(11) = 3,\ f(01) = f(10) = 1,\ f(00) = 2;$ for every pair



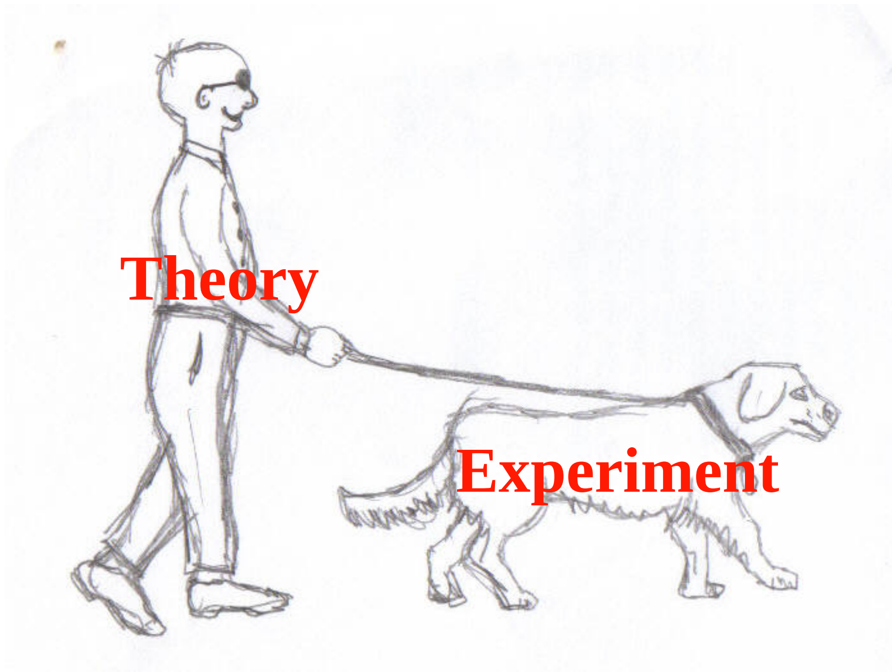


Without crossover - no preference for one size versus another

With crossover – short blocks preferred

Theory

The Problem of Theory…

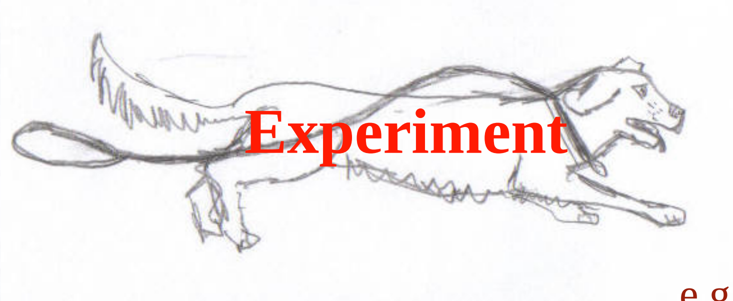


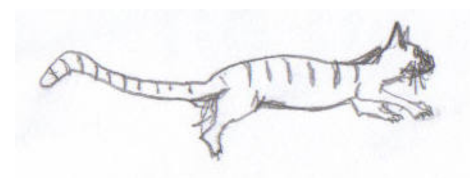
The "ideal"

The Problem of Theory…

In EC …

New Applications
New Algorithms





e.g. Multi-Resource Traveling Gravedigger Problem with Variable Coffin Size

"Most algorithms are NEVER used (except by the people who created them)" - Darrell Whitley, GECCO 2003 tutorial

The Problem of Theory…

The EC Expectation Gap

What theoreticians think practitioners are and what practitioners think theoreticians should be

What practitioners think theoreticians are and what theoreticians think practitioners should be

The Problem of Theory…

“Professors in every branch of the sciences prefer their own theories to truth; the reason is that their theories are private property, but truth is common stock” – Charles Caleb Colton, Lacon (1825).

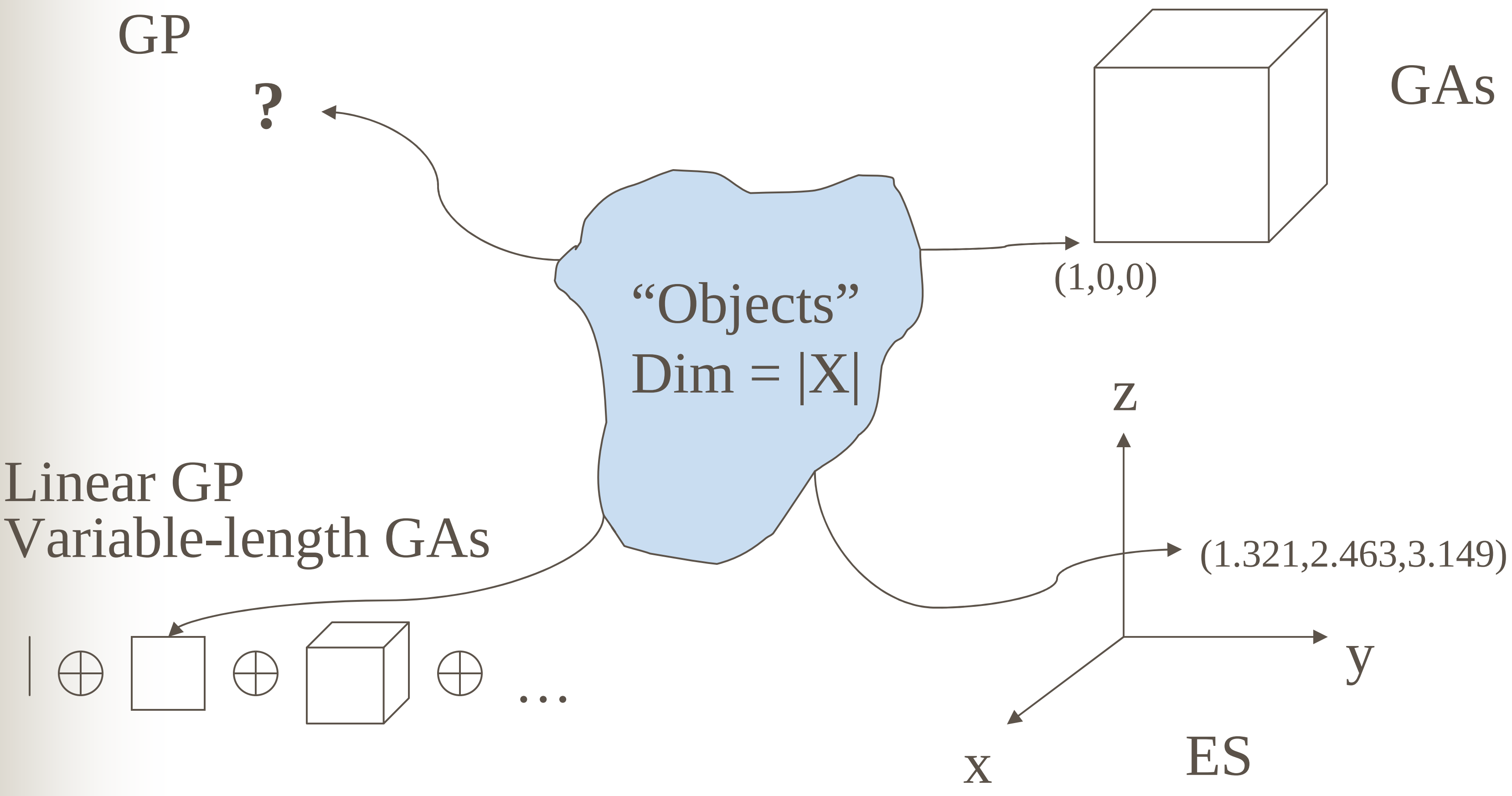
“It is the nature of an hypothesis, when once a man has conceived it, that is assimilates everything to itself as proper nourishment, and, from the first moment of your begetting it, it generally grows the stronger by everything you see, hear, read, or understand” – Laurence Sterne, Tristram Shandy (1767).

“EC theory is hard!” - Chris Stephens (most weeks).

The EC “Expectation Gap” ⇕ The Population Biology “Lesson”

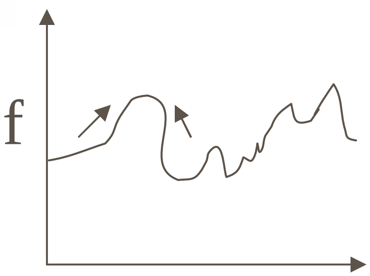
“How does this help practitioners…?” – most referees

EC Theory – the “Bare Necessities”



EC Theory – the "Bare Necessities"

Objects have fitness: $f_X : X \rightarrow R^+$

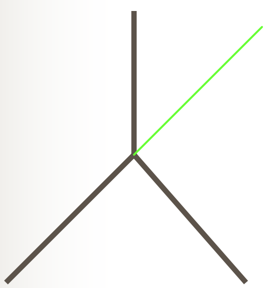


Objects have interactions:

$Object_i \rightarrow Object_i$ $\quad P'_i$ Selection

$Object_i \longleftrightarrow Object_j$ $\quad \mathcal{P}_{ij}$ Mutation

$Object_i + Object_j \longleftrightarrow Object_k + Object_l$ Recombination



m – recombination "mode"

$$\mathbf{P}(t+1) = \mathcal{H}(\mathbf{p}, \mathbf{f}, \mathbf{P}(t))$$

Dynamics

WARNING!

EQUATIONS

A

EC Theory – the "Bare Necessities"

$$P_I(t+1) = M_I{}^J \left((1-p_c)P'_J + p_c \sum_M \sum_{K,L} \frac{1}{2}(p(M) + p(\bar{M}))\lambda_J{}^{KL}(M)P'_K P'_L \right)$$

$P_I(t+1)$ - Probability to find "object" I

P'_J - Probability to select "object" J

$M_I{}^J$ - Probability to mutate "object" J to "object" I

$p(M)$ - Probability for recombination mask/mode M

p_c - Probability to implement recombination

$\lambda_J{}^{KL}(M)$ - Probability that given "objects" K and L and mode M "object" J is created (= 0, 1).

Sums are over all possible recombination modes and all objects J and K. e.g. for GA and homologous crossover 2^{3N} terms

EC Theory – the "Bare Necessities"

$$P_I(t+1) = M_I{}^J \left((1-p_c)P'_J + p_c \sum_M \sum_{K,L} \frac{1}{2}\left(p(M) + p(\bar{M})\right) \lambda_J{}^{KL}(M) P'_K P'_L \right)$$

- $M_I{}^J$: Probability that "object" J is mutated to "object" I
- $(1-p_c)P'_J$: Probability that "object" J is cloned
- $\sum_M \sum_{K,L} \frac{1}{2}\left(p(M) + p(\bar{M})\right) \lambda_J{}^{KL}(M) P'_K P'_L$: Probability that parent "objects" K and L are selected and "mixed" to form child "object" J via mode M

Example: 2-bit GA with $p(M) = 1/4$ for all M, $I = (11)$

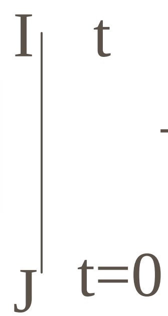
$$\lambda_{(11)}((00)) = \begin{pmatrix} 0&0&0&0 \\ 0&0&0&0 \\ 0&0&0&0 \\ 1&1&1&1 \end{pmatrix} \qquad \lambda_{(11)}((01)) = \begin{pmatrix} 0&0&0&0 \\ 0&0&0&0 \\ 0&1&0&1 \\ 0&1&0&1 \end{pmatrix}$$

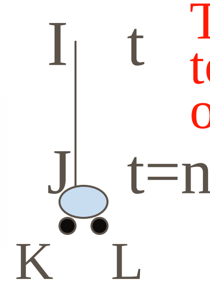
Ugly!

$$\lambda_{(11)}((11)) = \lambda_{(11)}((00))^T \qquad \lambda_{(11)}((10)) = \lambda_{(11)}((01))^T$$

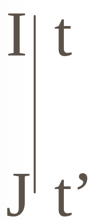
EC Theory – the “Bare Necessities”

Can integrate the equation and represent the solution graphically -

$P_I(t) = \sum_J$ 

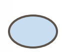
$+ \sum_{JKL} \sum_M \sum_{n=0}^{t}$ 

Term exclusively due to constructive effect of recombination



$= G_{IJ}(t, t')$

Probability that object J propagates from t to t’ and converts to I on the way


$= \frac{1}{2}\left(p(M) + p(\bar{M})\right)\lambda_J{}^{KL}(M)\,\frac{f_K}{\bar{f}(t)}\,\frac{f_L}{\bar{f}(t)}$

Measures strength of interaction between objects J, K and L

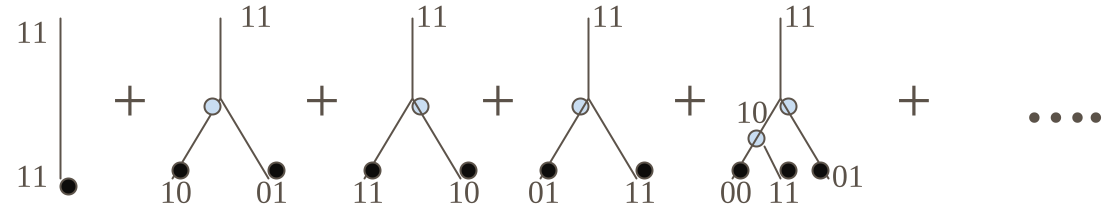
K • $= P_I(t)$

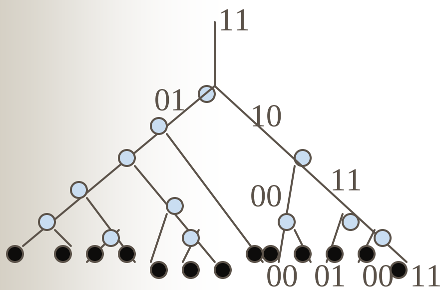
EC Theory – the "Bare Necessities"

Iterate … by recursively substituting for • until get to t = 0

Example – 2-bits 1-point crossover

$P_{11}(t)$ =





\+ ….

Each tree tells us the probability of forming 11 by a given process. In principle can see which processes are most important. But …

Tree depth bounded only by t!

COMPLICATED!

EC Theory – the "Bare Necessities"

General (Feynman) rules:

1) Draw all possible tree diagrams that contribute to creation of "object"
2) For each internal line ———— attach a propagator

$$G_{IJ}(t,t') = (1-p_c)^{t-t'} \frac{(\mathbf{FM})_{IJ}^{t-t'}}{\sum_I (\mathbf{FM})_{IJ}^{t-t'} P_J(t')}$$

3) To each vertex ⬭ attach a weight

$$\frac{1}{2}\left(p(M)+p(\bar{M})\right)\lambda_J{}^{KL}(M)\,\frac{f_K}{\bar{f}(t)}\,\frac{f_L}{\bar{f}(t)}$$

4) To each root • attach a factor $P_I(t')$
5) Carry out integration over time for all vertices

ALL CLEAR!

EC Theory – the "Bare Necessities"

So what do we have so far from the "microscopic" theory?

Exact	Yes
Mathematically rigorous	Yes ?
Unifies Phenomena	Yes/No
Intuitive	No
Predicts well	No
Useful for Practitioners	No

EC Theory – the “old stuff”

Let’s compare with the old Schema theorem and Building Block Hypothesis approach

Exact	No
Mathematically rigorous	Yes ??
Unifies Phenomena	Yes/No
Intuitive	Yes/No
Predicts well	No ?
Useful for Practitioners	Yes/No

Coarse Graining

Coarse Graining

Why?

What?

How?

Coarse Graining

Why?

1. **Emergence of "Effective Degrees of Freedom" (EDOF)/Collectivity/Universality**

2. **Curse of dimensionality/intractable dynamics**

   Coarse-grained degrees of freedom are combinations of the underlying "microscopic" degrees of freedom. EDOF are those coarse-grained degrees of freedom that are important for the dynamics

Coarse Graining

What?

1. “Direct” dimensional reduction
2. Phenotypes
3. Schemata
4. Hyperschemata
5. Building Blocks
6. Lowest cumulants of fitness distribution
7. “Normal (e.g. Walsh) modes”
8. Others

What is the most natural coarse graining depends on the operators and their corresponding parameters, the fitness landscape and the population.

Coarse Graining

How?

1. Phenotype Dynamics
2. Schemata Dynamics
3. Hyperschemata Dynamics
4. Building Block Dynamics
5. Aggregation of Markov chain
6. Truncation of cumulants
7. Walsh analysis
8. Others

Is it exact?

Coarse Graining By Coordinate Transformations

Identifying "Effective Degrees of Freedom"

WARNING!

EQUATIONS

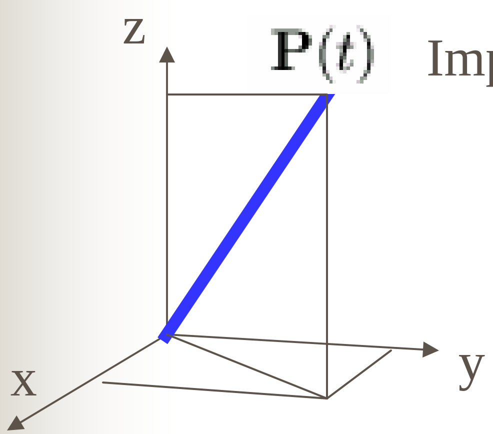
AA

Coarse graining via Coordinate Transformations

$$P_I(t+1) = M_I{}^J \left((1-p_c)P'_J + p_c \sum_M \sum_{K,L} \frac{1}{2}\left(p(M) + p(\bar{M})\right) \lambda_J{}^{KL}(M) P'_K P'_L \right)$$

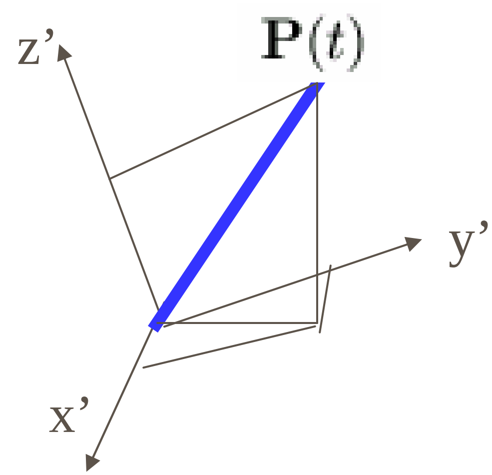
is COVARIANT, i.e. has the same content in ANY coordinate system

$\mathbf{P}(t)$ is INVARIANT in any coordinate system, its components $P_I(t)$ however, do change



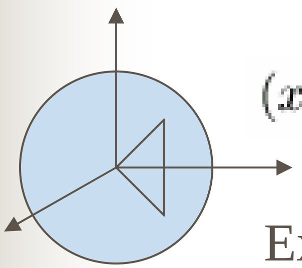
Implemented by rotation matrix

$$P_I = \sum_J \Lambda_I{}^J P_J$$



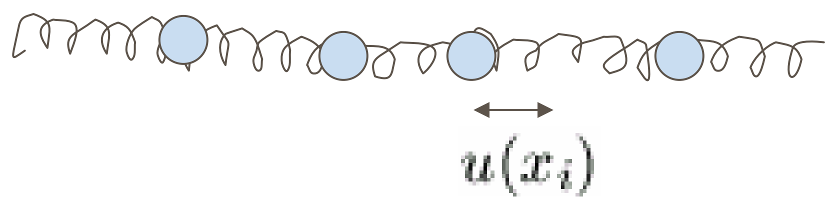
Coarse graining via Coordinate Transformations

- Appropriate choice of coordinate system can make manifest the **Effective Degrees of Freedom** and greatly facilitate calculations



$(x, y, z) \rightarrow (R, \theta, \phi)$

Exploit spherical symmetry



$$u(k) = \frac{1}{\sqrt{2}} \sum_{x_i} \exp(ikx_i) u(x_i)$$

Normal modes - waves

Coordinate system used up to now is the “object” system – e.g. strings, trees etc. ⇒ OK when EDOF are strings, trees etc.

Appropriate in “strong” selection regime

Coarse graining via Coordinate Transformations

Mutation …

Walsh basis (for fixed length binary strings)

$$W_{i,j} = 2^{-\ell/2}(-1)^{\#(i \otimes j)} \qquad W = \frac{1}{2}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}$$

Coordinate transformation matrix is orthonormal

$$\mathbf{M} = \begin{pmatrix} (1-p)^2 & p(1-p) & p(1-p) & p^2 \\ p(1-p) & (1-p)^2 & p^2 & p(1-p) \\ p(1-p) & p^2 & (1-p)^2 & p(1-p) \\ p^2 & p(1-p) & p(1-p) & (1-p)^2 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & (1-2p) & 0 & 0 \\ 0 & 0 & (1-2p) & 0 \\ 0 & 0 & 0 & (1-2p)^2 \end{pmatrix}$$

"Frequencies" of normal modes

EDOF are discrete versions of normal modes

Coarse graining via Coordinate Transformations

In Walsh basis …

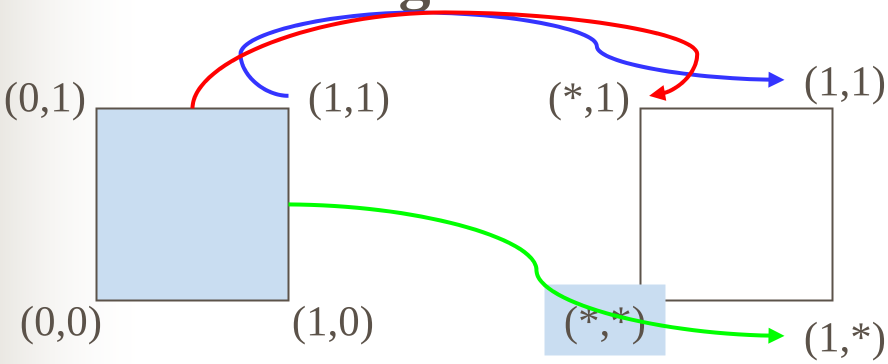
- Mutation matrix is diagonal
- Selection matrix is non-diagonal
- Crossover – O(n) Walsh coefficients made up from crossing O(m) and O(n-m) coefficients
- "Normal modes" not simply interpretable
- Useful for landscape analysis
- **Gives exact solution for mutation only**

Appropriate in "strong" mutation regime

Coarse graining via Coordinate Transformations

Crossover…

Building Block basis



$\Lambda =$

	111	110	101	011	100	010	001	000
111	1	0	0	0	0	0	0	0
11*	1	1	0	0	0	0	0	0
1*1	1	0	1	0	0	0	0	0
*11	1	0	0	1	0	0	0	0
1**	1	1	1	0	1	0	0	0
1	1	1	0	1	0	1	0	0
**1	1	0	1	1	0	0	1	0
***	1	1	1	1	1	1	1	1

$\tilde{\lambda}_{\alpha\beta\gamma}(m) = \mathbf{\Lambda}_{\alpha i}\lambda_{ijk}\mathbf{\Lambda}^{-1}_{\beta j}\mathbf{\Lambda}^{-1}_{\gamma k} = 0$ unless γ is the complement of β with respect to α and β is equivalent to m

Example: 2-bit GA with $p(M) = 1/4$ for all M, $I = (11)$

$$\lambda^{m}_{(11)} = \frac{1}{4}\begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

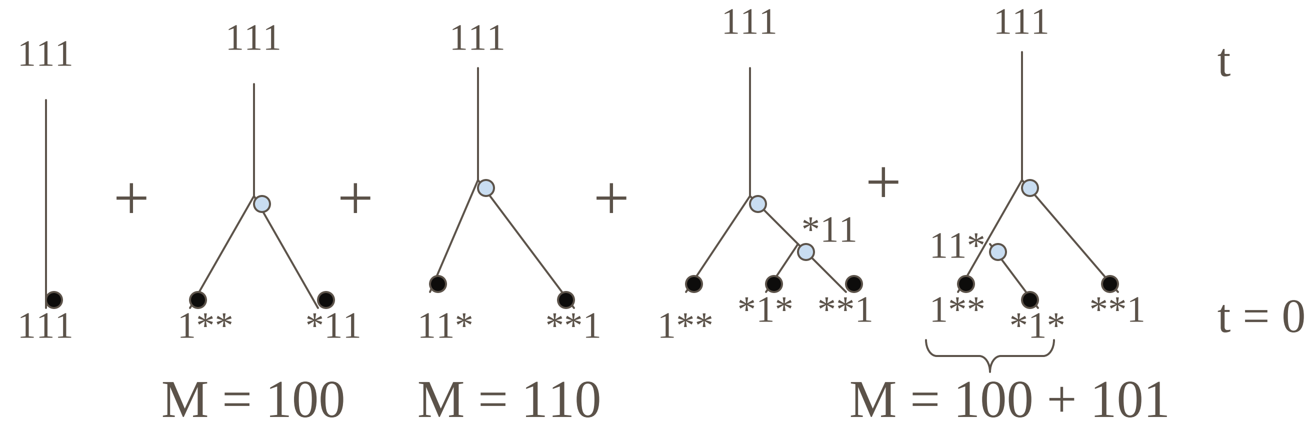
- In Building Block basis interaction matrix is skew diagonal
- Mask simply tells you which skew diagonal elements interact, e.g. mask 101011 points to building block 1*1*11 which interacts with *1*1** to give 111111

Coarse graining via Coordinate Transformations

Iterate … by recursively substituting for • until get to t = 0

Example – 3-bits 1-point crossover

$P_{111}(t) =$



I | t … I | t' $= G_{II}(t, t')$ Probability that "Building Block" I propagates from t to t'

⬭ $= \frac{1}{2}\left(p(M) + p(\bar{M})\right)\lambda_J{}^{KL}(M)\,\frac{f_K}{\bar{f}(t)}\,\frac{f_L}{\bar{f}(t)}$ Measures strength of interaction between "Building Blocks" J, K and L

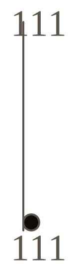
K • $= P_I(t)$

Skew-diagonal – only conjugate "Building Blocks" interact!

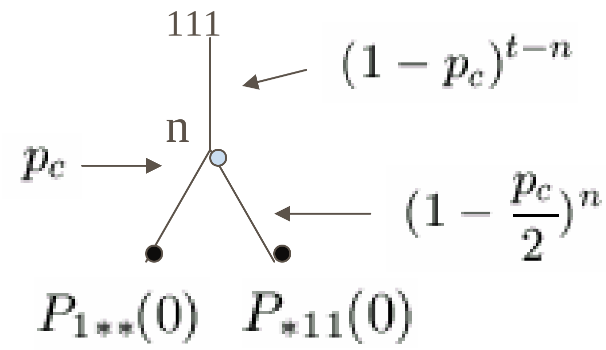
Coarse graining via Coordinate Transformations

Each tree tells us the probability of forming 111 by a given process. In principle can see which processes are most important. Tree depth bounded by N. MUCH SIMPLER THAN STRING (OBJECT) BASIS!

Examples:

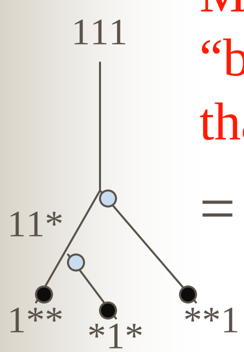


$$= (1-p_c)^t P_{111}(0)$$



$$= 2(1-p_c)\left((1-\frac{p_c}{2})^t - (1-p_c)^t\right)$$

Moral: No point putting in "building blocks" of higher order than one!



$$= 2(1-p_c)\left((1-p_c)^t - 2(1-\frac{p_c}{2})^t + 1\right) P_{1**}(0)P_{*1*}(0)P_{**1}(0)$$

Dominates in long time limit – Geiringer's theorem

Coarse graining via Coordinate Transformations

For a particular recombination "channel" (mode) whether recombination contributes positively or negatively to the effective fitness is determined by

$$\Delta(m) = P'_i(t) - P'_{i_m}(t)P'_{i_{\bar{m}}}(t)$$

BBs of i

SWLD (Selection Weighted Linkage Disequilibrium) Coefficient

If $\Delta(m) < 0$ "channel" is non-deceptive ⇨ long schemata preferred (see page 21)

If $\Delta(m) > 0$ "channel" is deceptive; deception – just like BBs - is dynamic

Standard Two-bit deception: f(0*) > f(1*) ⇨ $\Delta(m) > 0$

i.e. $P'_{11}(t) - P'_{1*}(t)P'_{*1}(t) > 0$

ALL CLEAR!

Coarse graining via Coordinate Transformations

In Building Block basis …

- Building Blocks schemata are the natural EDOF for recombination
- They are dynamical and not necessarily “short” or “fit”
- They are the ONLY way in which higher order “objects” can be built up by recombination
- Generically, the “construction” term dominates
- BBB is complete but not orthonormal
- There are |X| equivalent BBB (related by simple permutations)
- Only “dual” objects (i.e. conjugate BBs) interact, e.g. line and plane intersect at a point
- **Gives exact solution for recombination only**

Appropriate in “strong” crossover regime

Coarse Graining By Projections

Making intractable dynamics more tractable

WARNING!

EQUATIONS

AA

Coarse graining via Projections

Introduce a general coarse-graining operator $\mathcal{R}(\eta, \eta')$

Which coarse grains from the variables $\eta \in X_\eta$ to the variables $\eta' \in X_{\eta'} \subset X_\eta$

Given two such coarse grainings:

$$\mathcal{R}(\eta, \eta')P_\eta(t) = P_{\eta'}(t) \qquad \mathcal{R}(\eta, \eta'')P_\eta(t) = P_{\eta''}(t)$$

but $$\mathcal{R}(\eta', \eta'')P_{\eta'}(t) = P_{\eta''}(t)$$

hence
$$\mathcal{R}(\eta, \eta'') = \mathcal{R}(\eta, \eta')\mathcal{R}(\eta', \eta'')$$

i.e. coarse grainings form a semi-group – **“Renormalization Group”**

Coarse graining via Projections

Dynamics coarse grains via

$$\mathcal{R}(\eta,\eta')\mathcal{H}(\mathbf{p},\mathbf{f},\mathbf{P}_{\eta}(t))$$

If this can be written in the form

$$\mathcal{H}(\mathbf{p}',\mathbf{f}',\mathbf{P}_{\eta'}(t))$$

with suitable "renormalizations"

$\mathsf{f} \longrightarrow \mathsf{f}'$ and $\mathsf{p} \longrightarrow \mathsf{p}'$

then the dynamics is form covariant or invariant under the coarse graining. If $\mathsf{f} = \mathsf{f}'$ and $\mathsf{p} = \mathsf{p}'$ dynamics is "compatible"

Coarse graining via Projections

Examples: Compatible Coarse grainings

1. Selection and Phenotypes
   - Unitation, e.g. 2^N genotypes → (N+1) phenotypes
   - Eigen model (NIAH), e.g. 2^N genotypes → 2 phenotypes

2. Mutation and Crossover and Schemata
   - 2^N genotypes → 2^{N_2} coarse grained genotypes

Incompatible Coarse grainings

1. Selection, Mutation and Crossover and Schemata
   - 2^N genotypes → 2^{N_2} coarse grained genotypes
   - $f_\alpha = \mathcal{R}(x,\alpha) f_x = \sum_{x\in\alpha} f_x P_x(t) / \sum_{x\in\alpha} P_x(t)$. time-dependent

Coarse graining via Projections

In BBB for 1-point crossover…

$$P_{111}(t+1) = (1-p_c)P_{111}(t) + \frac{p_c}{2}(P_{1**}(t)P_{*11}(t) + P_{11*}(t)P_{**1}(t))$$

"Zap" (projection) $111 \rightarrow 11*$ ⇨

$$P_{11*}(t+1) = (1-p_c)P_{11*}(t) + \frac{p_c}{2}(P_{1**}(t)P_{*1*}(t) + P_{11*}(t)P_{***}(t))$$

⇨ $$P_{11*}(t+1) = (1-\frac{p_c}{2})P_{11*}(t) + \frac{p_c}{2}P_{1**}(t)P_{*1*}(t)$$

Note – coarse grained (projected) 3-bit equation same as "microscopic" 2-bit equation with "renormalization" $p_c \rightarrow \frac{p_c}{2}$

$$P_{11}(t+1) = (1-p_c)P_{11}(t) + p_c P_{1*}(t)P_{*1}(t)$$

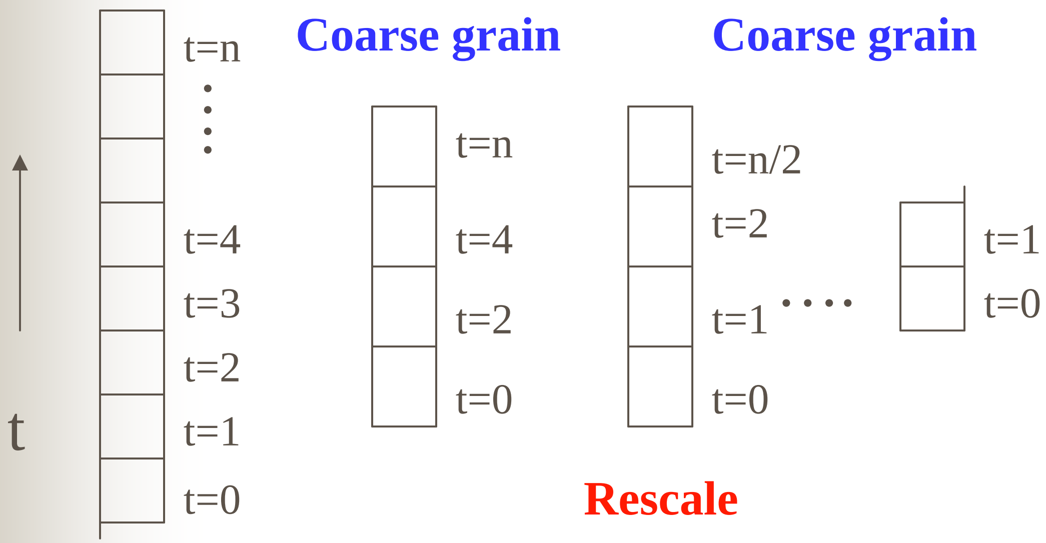
<u>**FORM INVARIANCE**</u>

Coarse graining via Projections

- Generalizes to the case of variable-length GAs and GP; Building Block Schemata →Building Block Hyperschemata; "form invariance" of equations over different types of EA and form invariant upon coarse graining to schemata;
- Gives exact form of the Schema Theorem and generalizes it to EAs other than GAs
- Neglecting the "construction" terms leads to standard Holland Schema Theorem as an approximation

Coarse Graining by Projection - "Divide and Conquer"

Example: 1-bit



Can we coarse grain an n generation problem to a one generation problem? Much easier to solve the dynamics over only one generation!

$X_1(t)$ – unnormalized incidence vector

p – mutation rate

$$\begin{pmatrix} X_1(t+2) \\ X_0(t+2) \end{pmatrix} = \underbrace{\begin{pmatrix} (1-p)f_1 & pf_0 \\ pf_1 & (1-p)f_0 \end{pmatrix}^2}\begin{pmatrix} X_1(t) \\ X_0(t) \end{pmatrix}$$

Evolves bit two time steps in landscape f(1), f(0) with mutation p

Coarse Graining by Projection - "Divide and Conquer"

$$\begin{pmatrix} X_1(t'+1) \\ X_0(t'+1) \end{pmatrix} = \underbrace{\begin{pmatrix} (1-p_1')f_1' & p_0' f_0' \\ p_1' f_1' & (1-p_0')f_0' \end{pmatrix}} \begin{pmatrix} X_1(t') \\ X_0(t') \end{pmatrix}$$

Evolves bit one time step in "renormalized" landscape f'(1), f'(0) with asymmetric mutation rates p'(1) and p'(0)

Equivalent dynamics (all we did was "change names"!, i.e. "renormalize")

$$f_1' = (1-p_1)f_1^2 + p_1 f_0 f_1$$
$$f_0' = (1-p_0)f_0^2 + p_0 f_0 f_1$$
$$p_1' = p_1 \left(\frac{(1-p_1)f_1 + (1-p_0)f_0}{(1-p_1)f_1 + p_1 f_0} \right)$$
$$p_0' = p_0 \left(\frac{(1-p_0)f_0 + (1-p_1)f_1}{(1-p_0)f_0 + p_0 f_1} \right)$$

ALL CLEAR!

Coarse Graining by Projection - “Divide and Conquer”

Evolution of mutation/selection GA over n time steps with fitness landscape f(1), f(0) and mutation rates p(2) and p(1) is same as that of a GA with “renormalized” landscape and mutation rates, f’(1), f’(0), p’(2), p’(1) over n/2 time steps!

⇓

UNIVERSALITY

Fixed points of Renormalization Group transformation:

$|\ln(f(1)/f(0))| = 0$, $p(1) = p(0) = 0$; no selection/mutation – “FERROMAGNETIC”

$|\ln(f(1)/f(0))| = \text{infinity}$, $p(1) = p(0) = 0$; strong selection – “FROZEN”

$|\ln(f(1)/f(0))| = \text{constant}$, $p(1) + p(0) = 1$; neutral evolution – “PARAMAGNETIC”

Coarse Graining by Projection - “Divide and Conquer”

- Iterated map takes you to a problem with fewer degrees of freedom – NOT associated with “trivial” symmetries.
- Linearization around the fixed points of the equations give the late time asymptotics
- Can understand “universality” of behaviour
- Can coarse grain in both “space” and “time”
- Coarse graining can almost never be done exactly
- Have to decide what coarse graining is most appropriate for a given model

EC Theory – Coarse-grained

So what do we have so far?

Exact	Yes
Mathematically rigorous	Yes ?
Unifies Phenomena	Yes
Intuitive	Yes
Predicts well	No
Useful for Practitioners	Yes/No

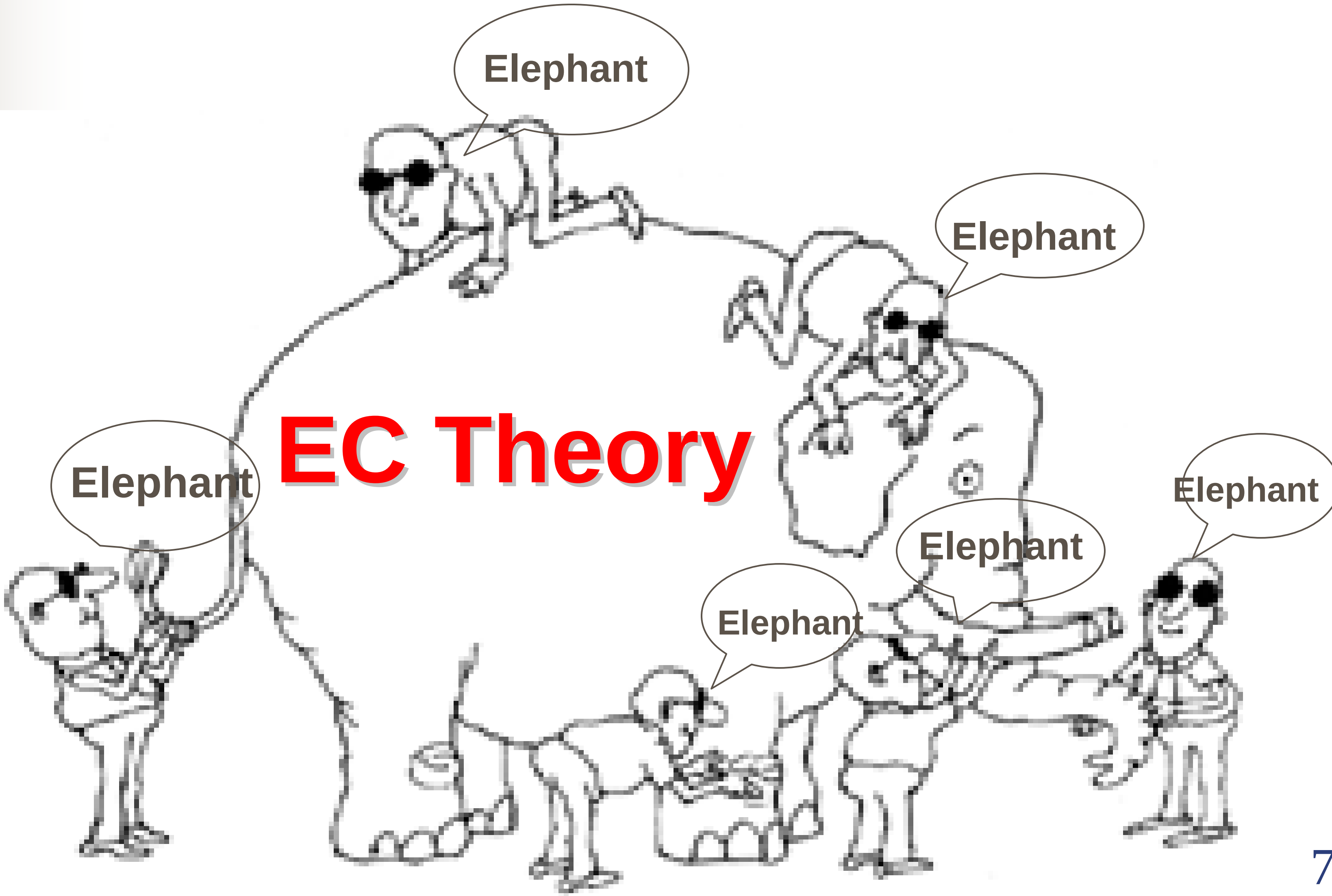
The Bottom Line …

- Present taxonomy in EC theory is inadequate
- Taxonomy can be greatly improved by using "distance" measures
- Taxonomy and universality can also be much better understood using an appropriate coarse graining
- Coarse grained genetic dynamics unifies and makes compatible different areas of EC and different previous theoretical formulations
- GAs and GP – different sides of the same coin
- Old Schema theory/BB hypothesis and Vose type models – different sides of the same coin
- Coarse graining and the Renormalization Group offer a generic methodology for approximating genetic dynamics

The Bottom Line …

- Theory in EC is **NOT** particularly well developed. If it was there wouldn't be such a huge expectation gap between theoreticians and practitioners. No systematic approximation techniques for attacking problems from first principles
- Practitioners have to realize what is and isn't theoretically feasible (theoretical population biologists have spent nearly a century achieving things that "practitioners" would scorn).
- Practitioners could really help by stress testing theory (too much testing of theory in the hands of people who make up the theory and too much testing of "never to be used" algorithms by practitioners)

The Bottom Line …



Acknowledgements

Collaborators:

Henri Waelbroeck, Riccardo Poli, Alden Wright, Peter Stadler, Chryssomalis Chryssomalakos, Jon Rowe, Bill Langdon

Students and postdocs: Adolfo Zamora, Rosalia Aguirre, Jaime Mora, Odon Palacios, Andres Aguilar, Ian Garcia

Valuable Critics:

Michael Vose, Dave Goldberg, Anil Menon, Ken De Jong, Nic McPhee

Funding: CONACyT; DGAPA, UNAM

References

CRS and H. Waelbroeck, ICGA97, 34-41 (1997)

CRS and H. Waelbroeck, Evol. Comp. 7, 109-124 (1999)

CRS, H. Waelbroeck and R. Aguirre, FOGA5, 117-135 (1999)

CRS, GECCO2001, 631-638 (2001)

CRS, Acta Phys. Slov. 52, 515-524 (2002)

R. Poli, GECCO2000, 469-476 (2000)

R. Poli, Genetic Prog. and Evol. Mach. 2, 123-163 (2001)

W.B. Langdon and R. Poli, Foundations of Genetic Programming, Springer-Verlag (2002)

R. Poli et al, EuroGP2002, 211-227, Springer-Verlag (2002)

P.F. Stadler and CRS, Comments in Theor. Biol. 8, 389 (2003)

R. Poli et al, GECCO2002, 211-227 (2002)

CRS and A. Zamora, GECCO2003

CRS and C. Chryssomalakos, GECCO2004

CRS and R. Poli, in Frontiers of Evolutionary Computation, Kluwer (2004)

References

Vose, M.D. (1999) *The Simple Genetic Algorithm: Foundations and Theory*, (MIT Press, Cambridge MA).

Bürger, R. (2000) "The Mathematical Theory of Selection, Recombination and Mutation", Wiley Series in Mathematical and Computational Biology.

Alden H. Wright. The exact schema theorem. http://www.cs.umt.edu/CS/FAC-/WRIGHT/papers/schema.pdf, January 2000.

Altenberg, L. (1995). The Schema Theorem and Price's Theorem. In Whitley, L. D. and Vose, M. D., editors, *Foundations of Genetic Algorithms 3*, pages 23–49, Estes Park, Colorado, USA. Morgan Kaufmann.

Goldberg, D. E. (1989a). Genetic algorithms and Walsh functions: I. A gentle introduction. *Complex Systems*, 3(2):129–152.

Goldberg, D. E. (1989b). Genetic algorithms and Walsh functions: II. Deception and its analysis. *Complex Systems*, 3(2):153–171.

Prügel-Bennett, A. and Shapiro, J. L. (1994). An analysis of genetic algorithms using statistical mechanics. *Physical Review Letters*, 72:1305–1309.

Whitley, D. (1992). An executable model of a simple genetic algorithm. In Whitley, D., editor, *Foundations of Genetic Algorithms Workshop (FOGA-92)*, Vail, Colorado.

J. Rowe and C. Reeves, Genetic Algorithms – Principles and Perspectives A Guide to GA Theory, Kluwer 2002.