Genetic Algorithms

M. Anthony Kapolka III

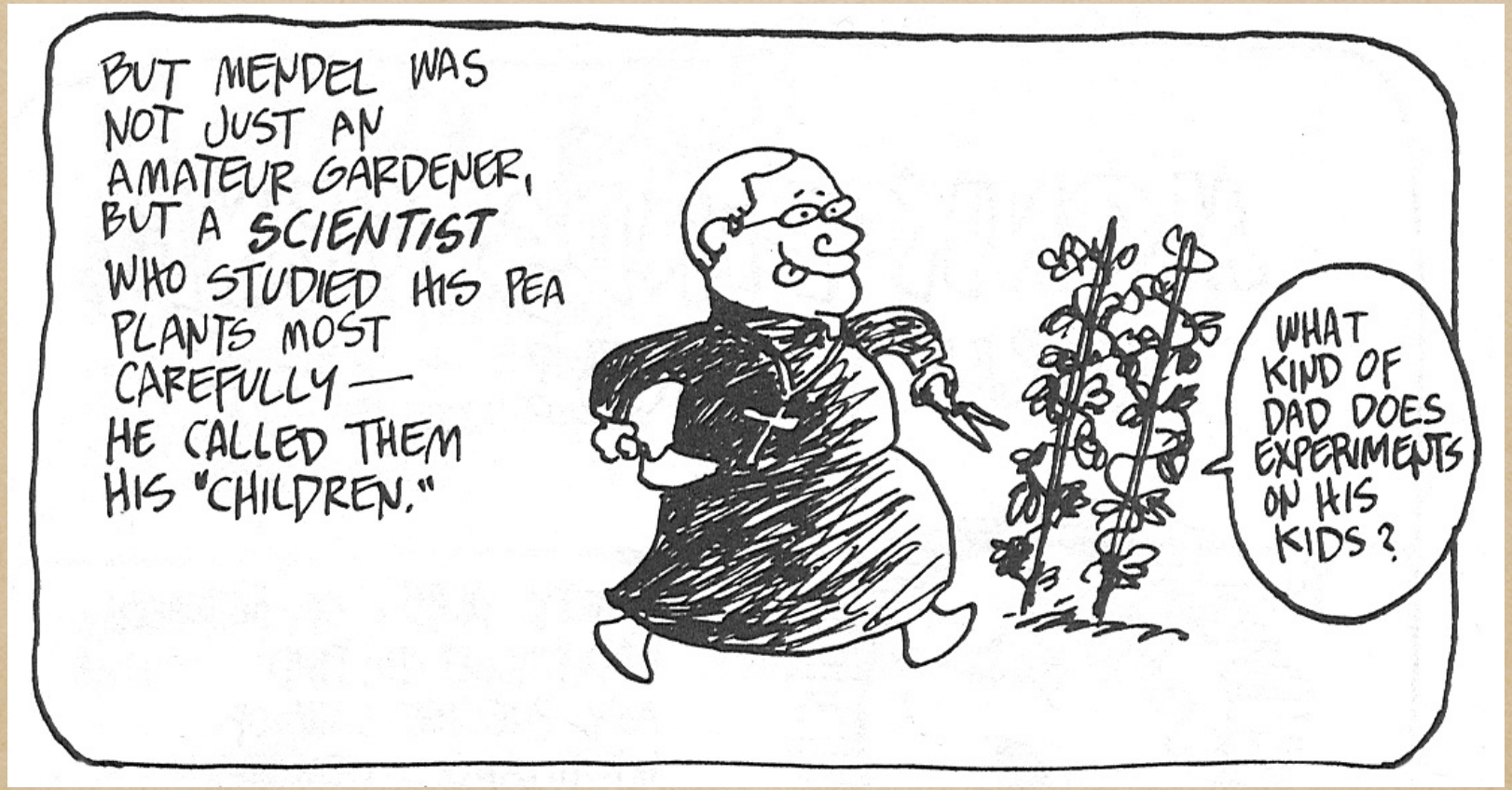
Wilkes University

CS 340 - Fall 2019

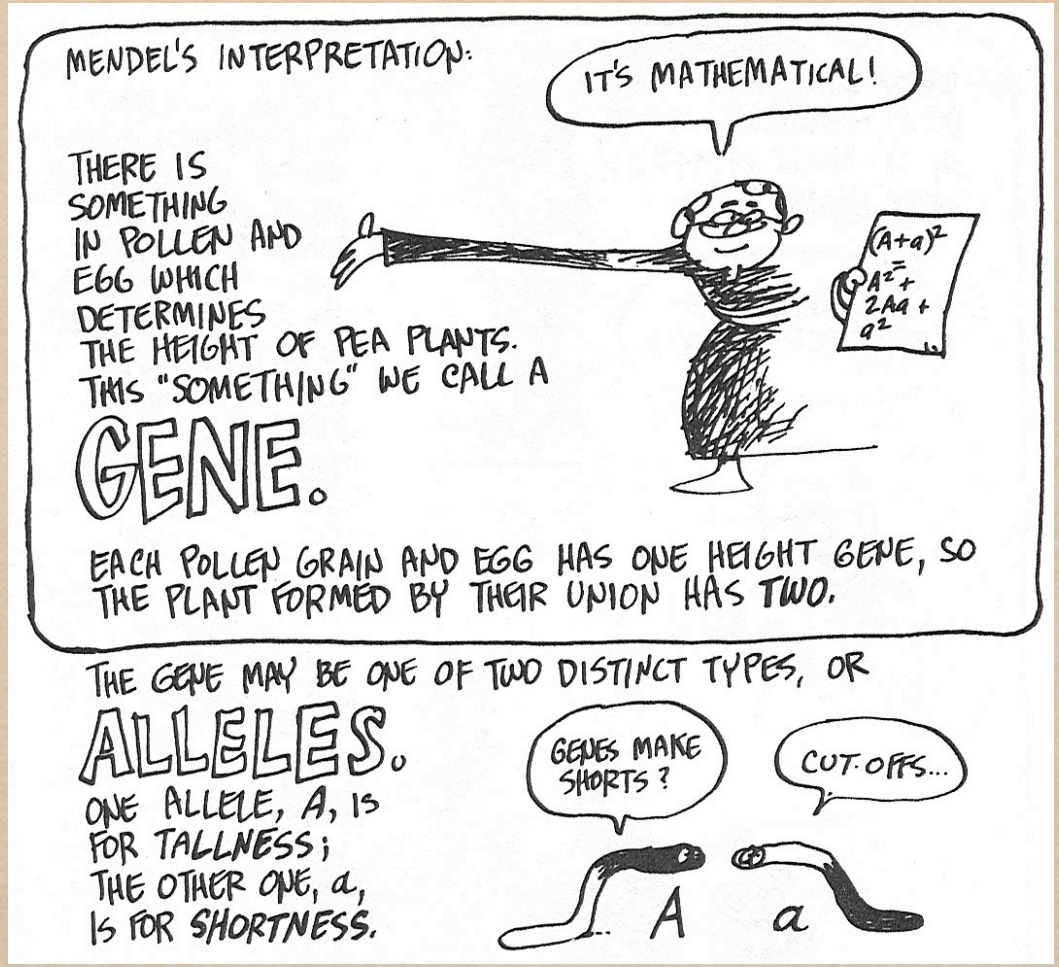
GREGOR MENDEL

(1822 - 1884) WAS AN AUGUSTINIAN MONK FROM BRÜNN, AUSTRIA. IN HIS SPARE TIME, MENDEL BRED PEA PLANTS IN THE MONASTERY GARDENS.

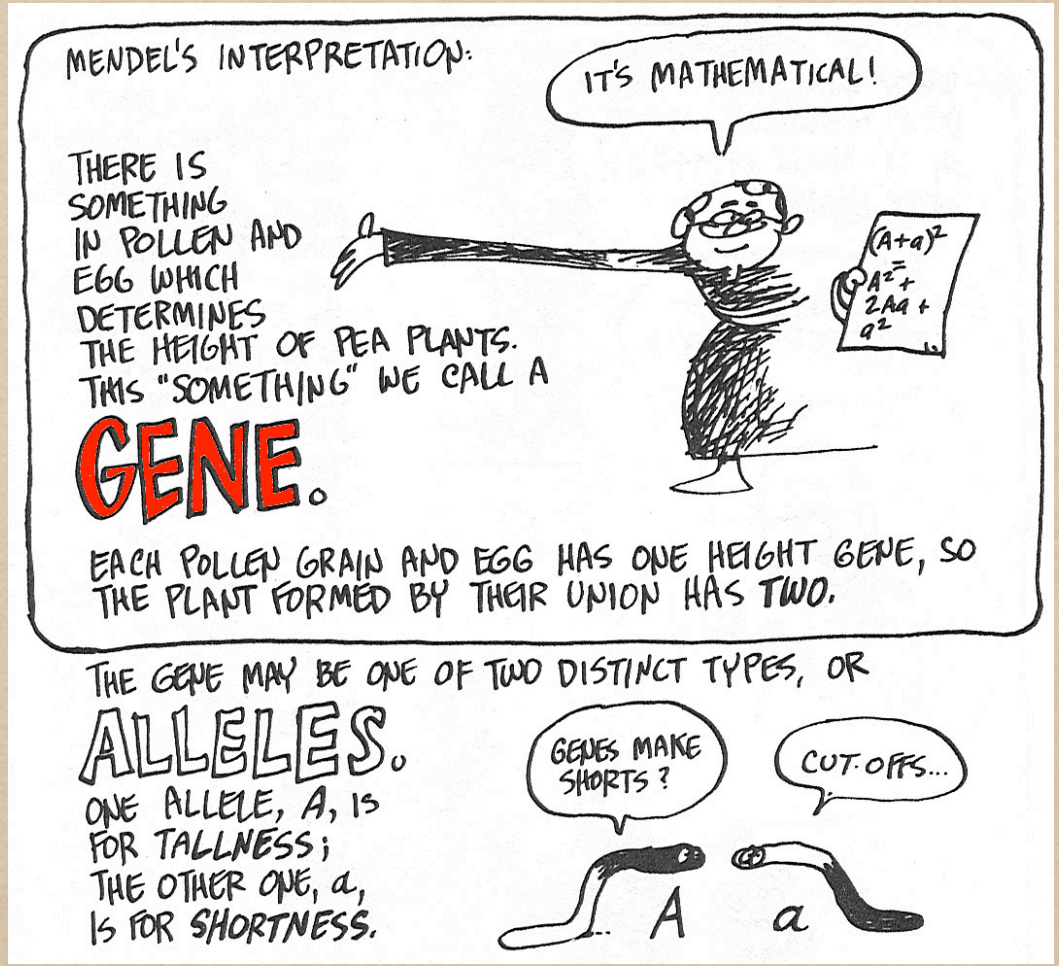
[Gonick 1991]


BUT MENDEL WAS NOT JUST AN AMATEUR GARDENER, BUT A SCIENTIST WHO STUDIED HIS PEA PLANTS MOST CAREFULLY — HE CALLED THEM HIS "CHILDREN."
WHAT KIND OF DAD DOES EXPERIMENTS ON HIS KIDS?

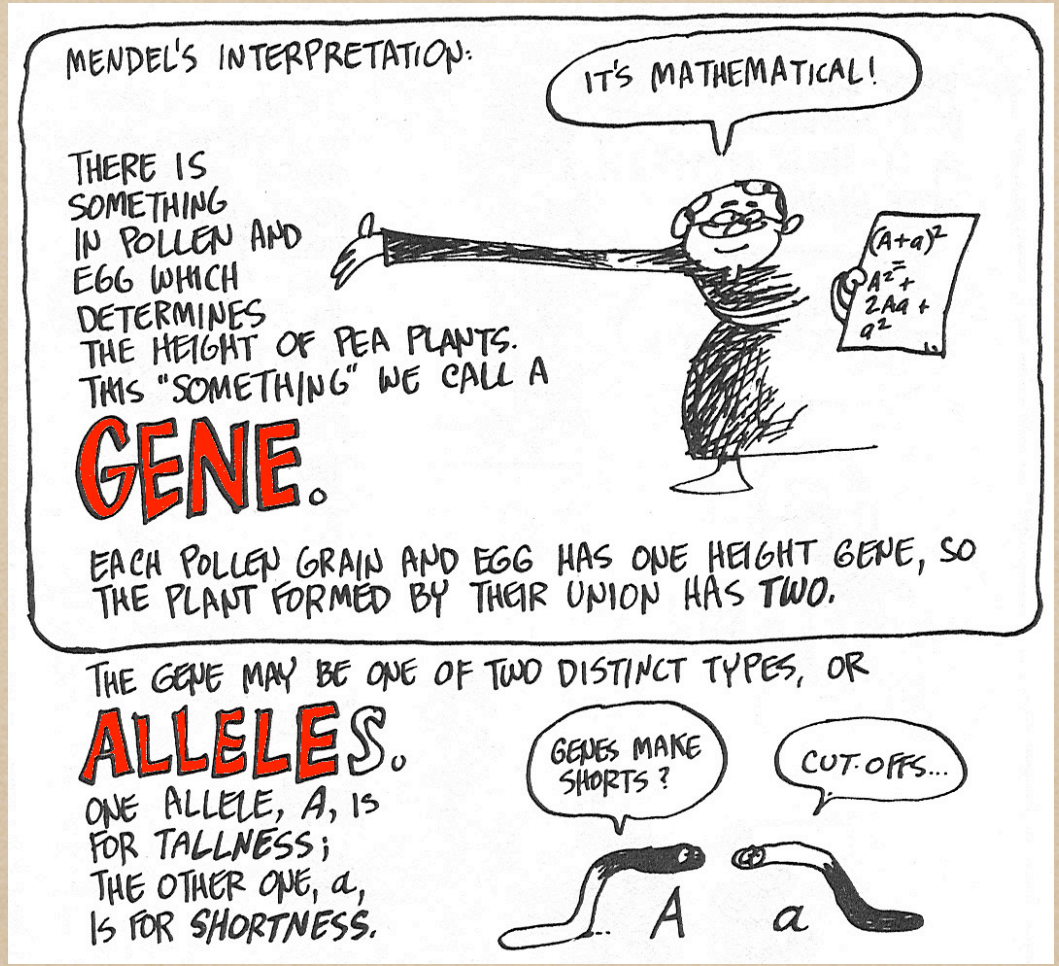
[Gonick 1991]


MENDEL'S INTERPRETATION:
IT'S MATHEMATICAL!
THERE IS SOMETHING IN POLLEN AND EGG WHICH DETERMINES THE HEIGHT OF PEA PLANTS. THIS "SOMETHING" WE CALL A
GENE.
$(A+a)^2$
$A^2 + 2Aa + a^2$
EACH POLLEN GRAIN AND EGG HAS ONE HEIGHT GENE, SO THE PLANT FORMED BY THEIR UNION HAS TWO.
THE GENE MAY BE ONE OF TWO DISTINCT TYPES, OR
ALLELES.
ONE ALLELE, A, IS FOR TALLNESS; THE OTHER ONE, a, IS FOR SHORTNESS.
GENES MAKE SHORTS?
CUT-OFFS...
A
a

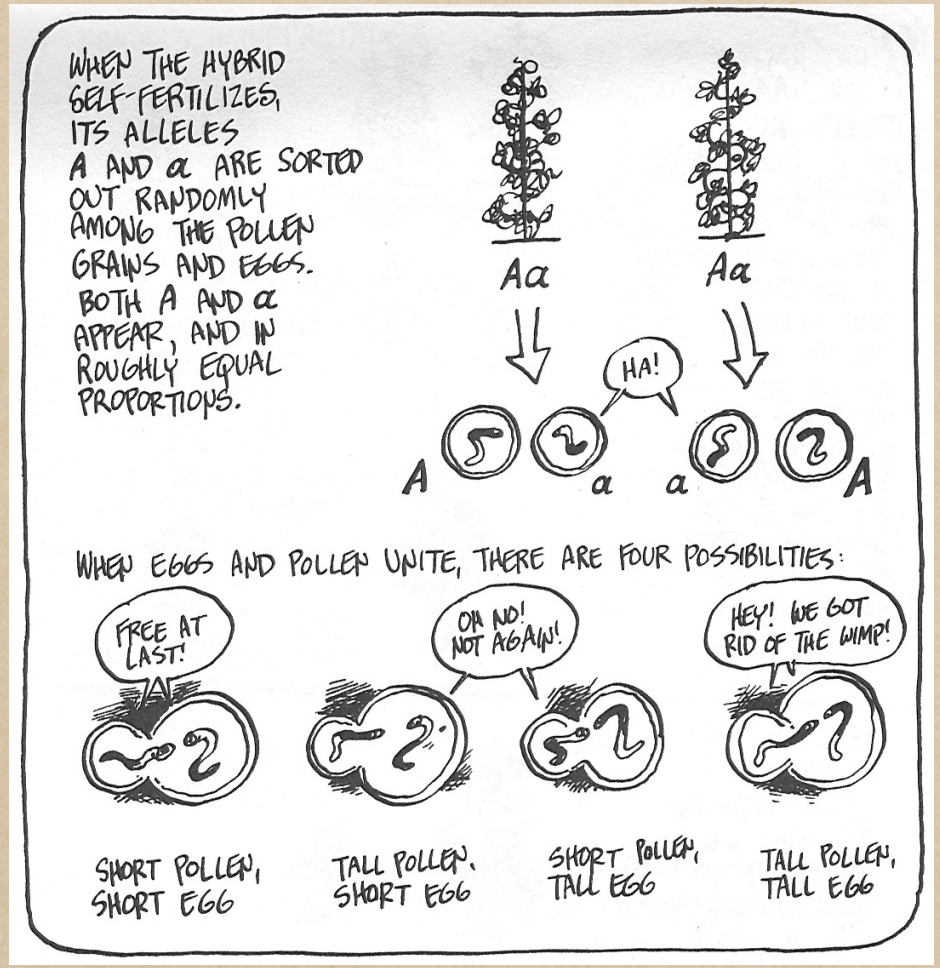
[Gonick 1991]


MENDEL'S INTERPRETATION:
IT'S MATHEMATICAL!
THERE IS SOMETHING IN POLLEN AND EGG WHICH DETERMINES THE HEIGHT OF PEA PLANTS. THIS "SOMETHING" WE CALL A
GENE.
$(A+a)^2$
$A^2 + 2Aa + a^2$
EACH POLLEN GRAIN AND EGG HAS ONE HEIGHT GENE, SO THE PLANT FORMED BY THEIR UNION HAS TWO.
THE GENE MAY BE ONE OF TWO DISTINCT TYPES, OR
ALLELES.
ONE ALLELE, A, IS FOR TALLNESS; THE OTHER ONE, a, IS FOR SHORTNESS.
GENES MAKE SHORTS?
CUT-OFFS...
A
a

[Gonick 1991]


MENDEL'S INTERPRETATION:
IT'S MATHEMATICAL!
THERE IS SOMETHING IN POLLEN AND EGG WHICH DETERMINES THE HEIGHT OF PEA PLANTS. THIS "SOMETHING" WE CALL A
GENE.
$(A+a)^2$
$A^2 + 2Aa + a^2$
EACH POLLEN GRAIN AND EGG HAS ONE HEIGHT GENE, SO THE PLANT FORMED BY THEIR UNION HAS TWO.
THE GENE MAY BE ONE OF TWO DISTINCT TYPES, OR
ALLELES.
ONE ALLELE, A, IS FOR TALLNESS; THE OTHER ONE, a, IS FOR SHORTNESS.
GENES MAKE SHORTS?
CUT-OFFS...
A
a

[Gonick 1991]

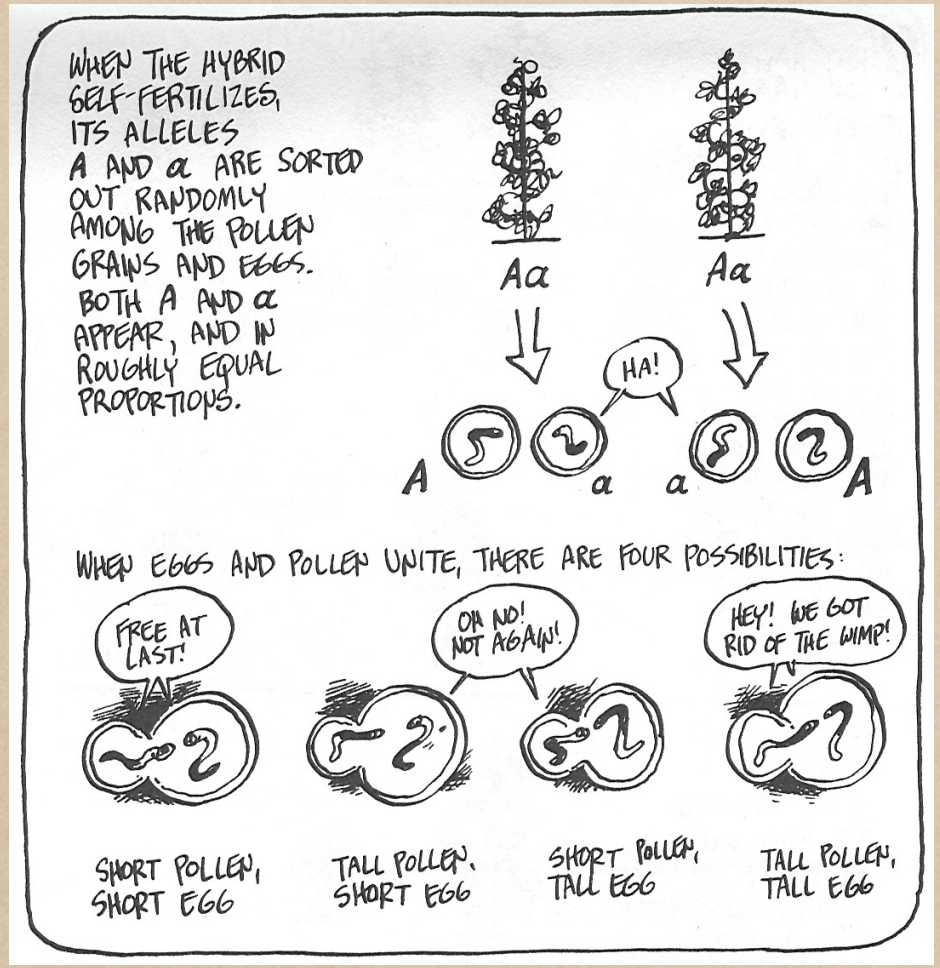


Individuals

Chromosomes

Genes

[Gonick 1991]



Individuals

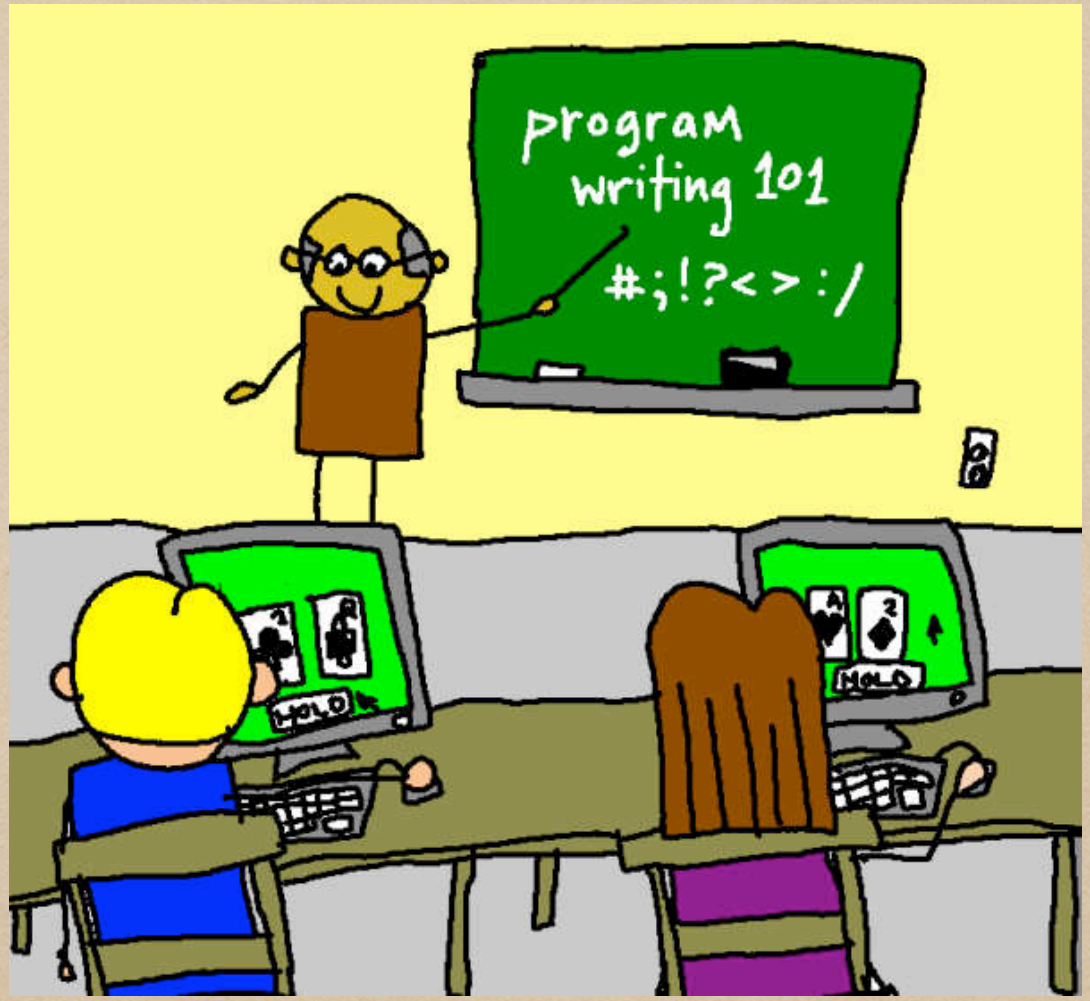
Chromosomes

Genes

Code:

IF aa THEN
be short
ELSE
be tall

[Gonick 1991]


program
writing 101
#;!?<>:/
HOLD
HOLD

First Steps

Pick a problem

- e.g. play blackjack

First Steps

Design a representation

00000000000000000000

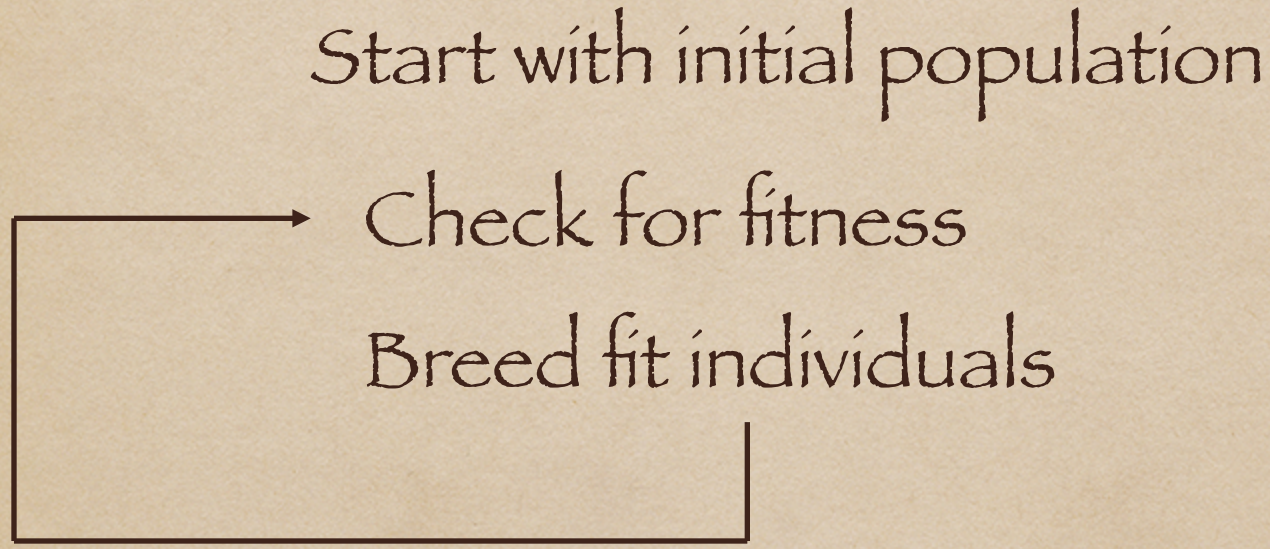
0 = stay 1 = hit

First Steps

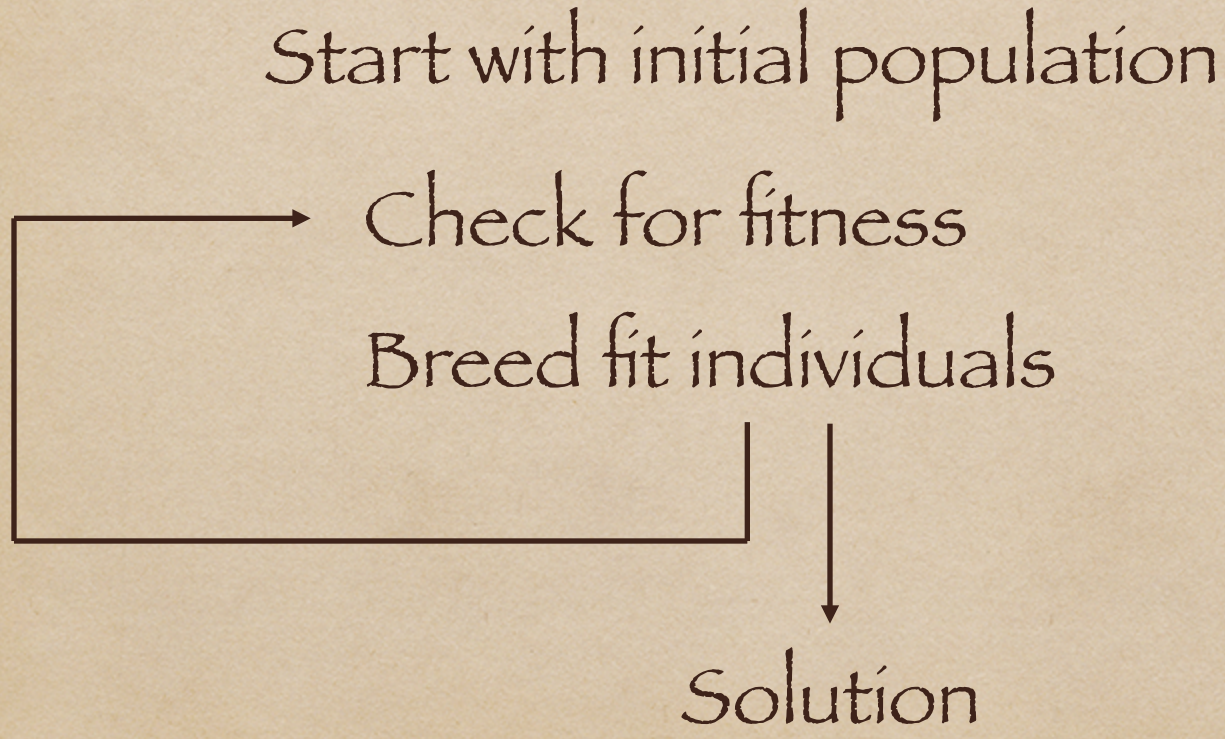
Design a fitness function

e.g. actually use
representation
to play Black Jack

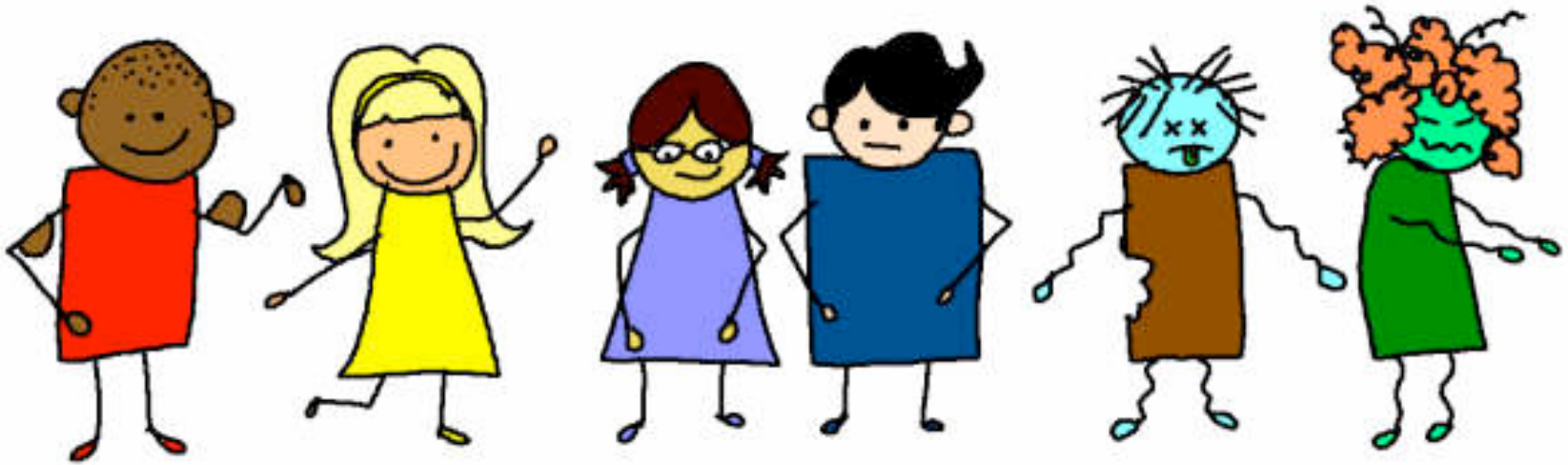
Basic Genetic Algorithm



Basic Genetic Algorithm



Which individuals to breed?

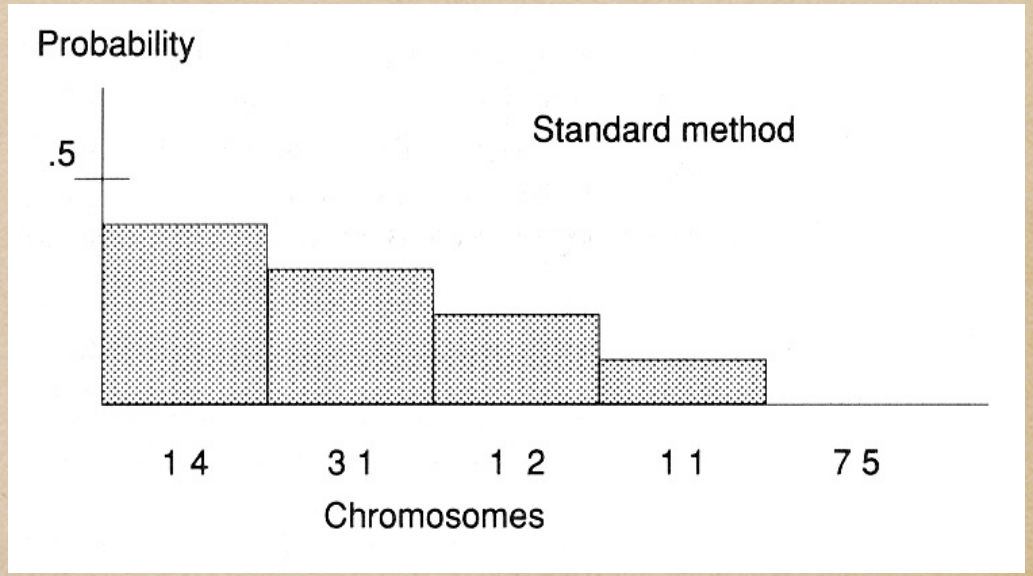


Which individuals to breed?

- Most fit
- Ranked by fitness
- Most Diverse

Standard Method

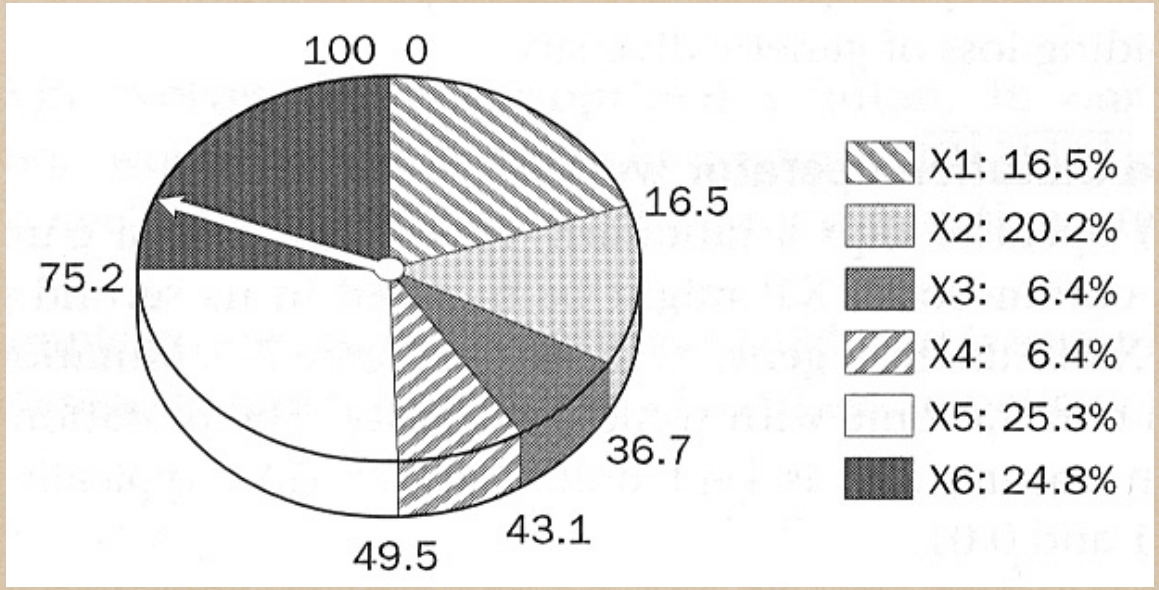
$$F_i = \frac{Q_i}{\sum_j Q_j}$$



Fitness equated
with relative quality

[Winston 1992]

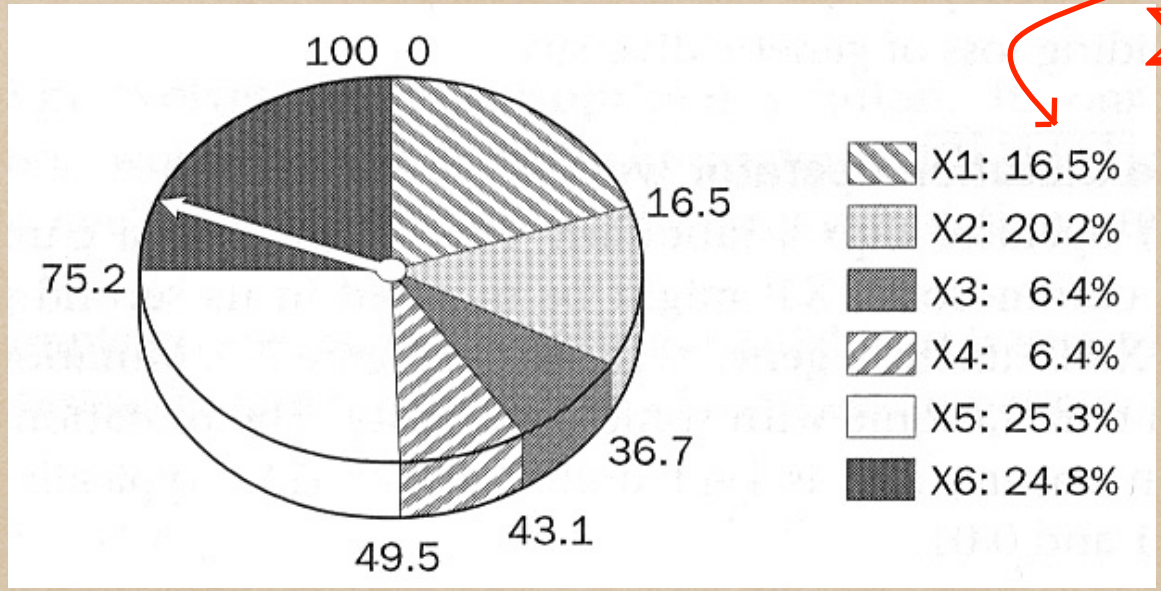
Standard Method



Randomly select individuals with relative quality

[Negnevitsky 2002]

Standard Method

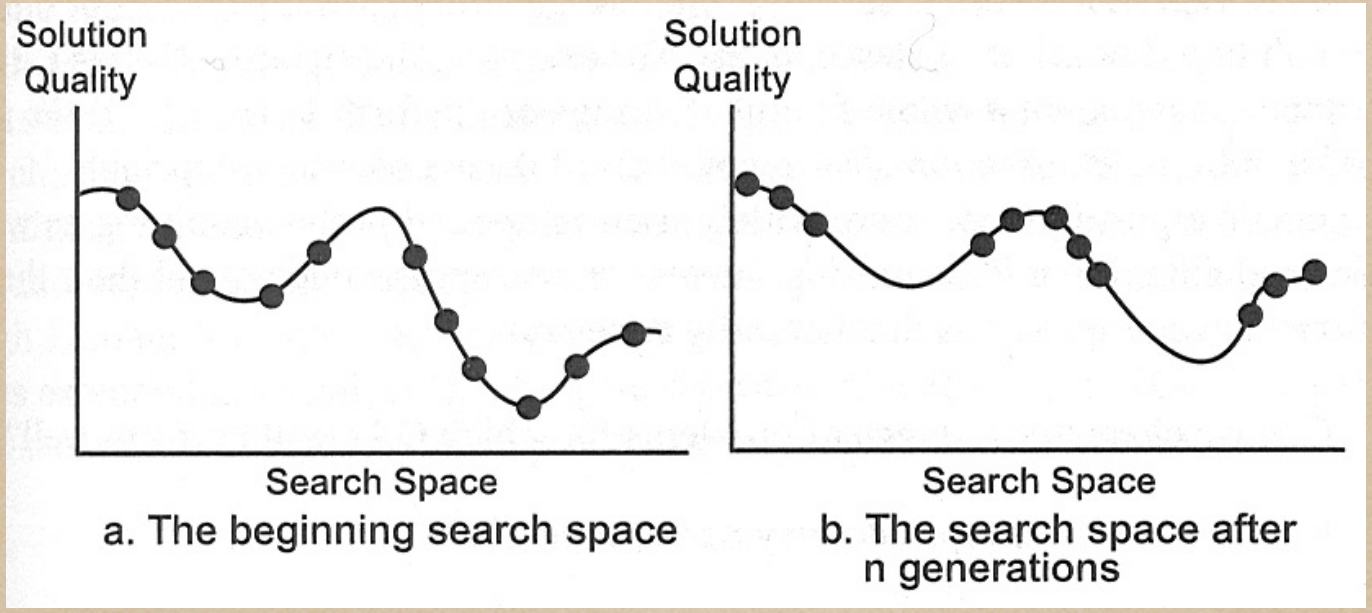


$$\frac{Q_i}{\sum_j Q_j}$$

Randomly select individuals with relative quality

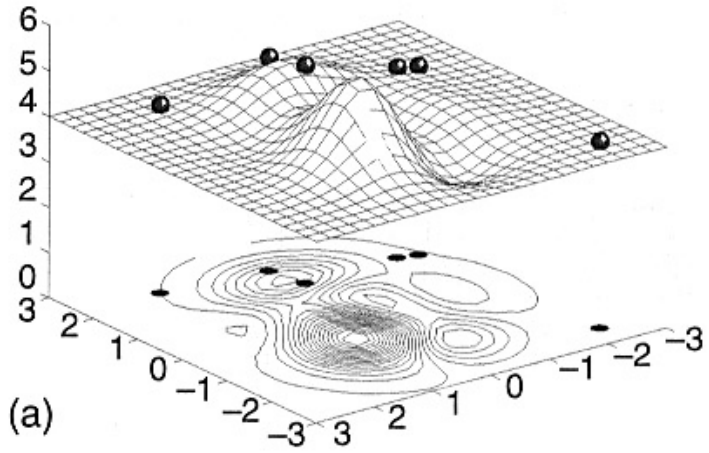
[Negnevitsky 2002]

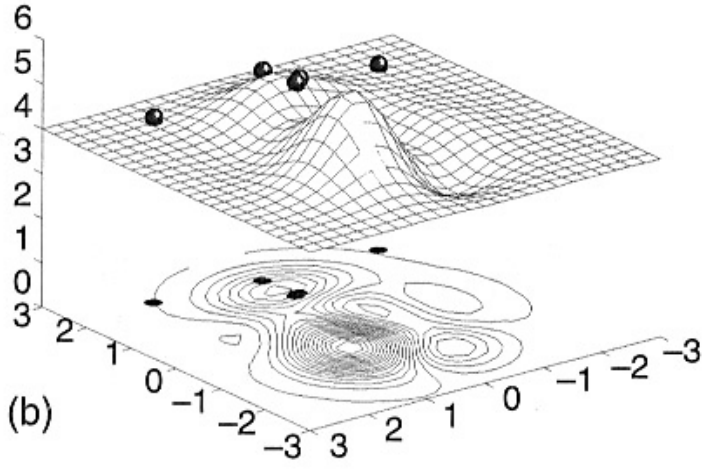
Like Parallel Hill Climbing

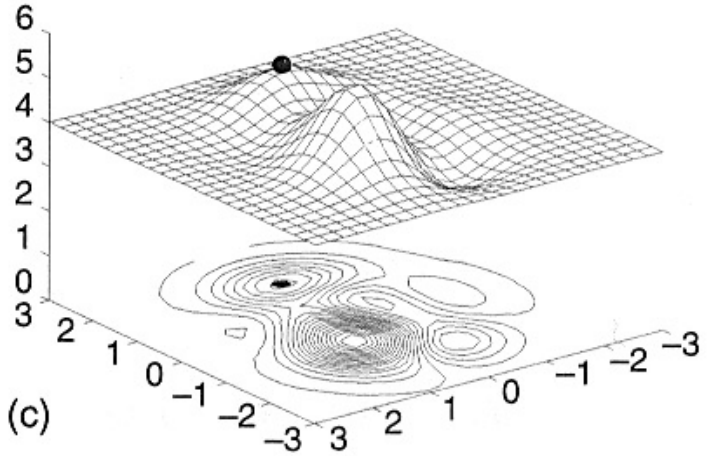


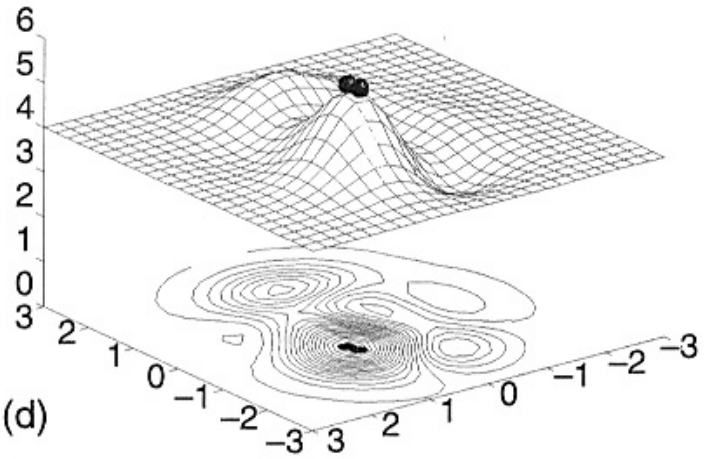
Fitness landscape

- Differs from heuristic state space search
  - no estimates of "effort" to move to new state
  - no strict ordering of states
- Has peaks, valleys, discontinuities
  - both real and introduced
  - e.g binary vs. gray coding





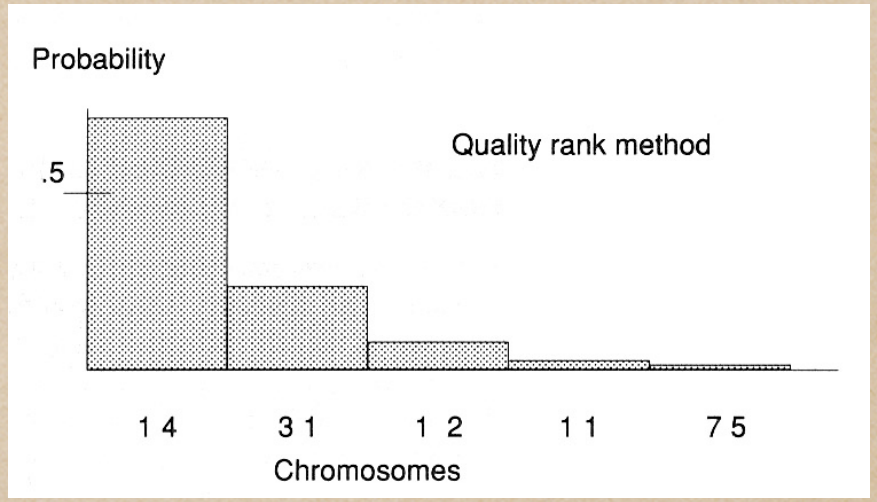




[Negnevitsky 2002]

Quality Rank Method

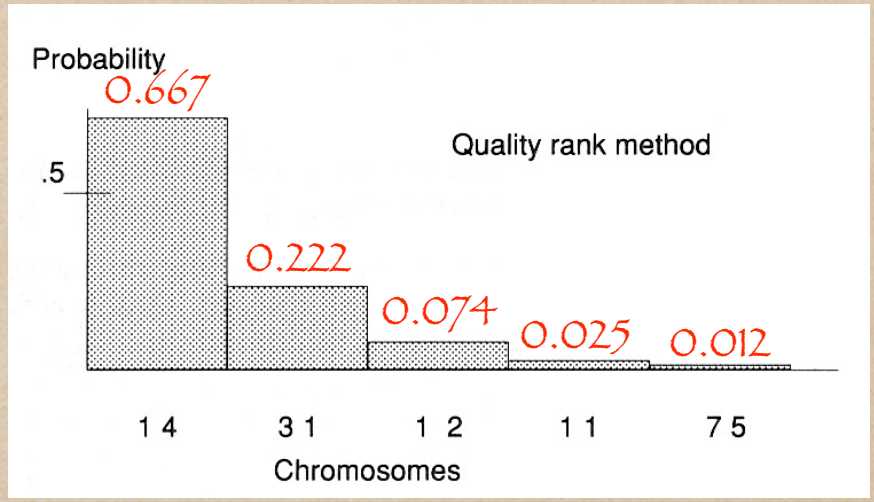
Rank each with fixed Probability



[Winston 1992]

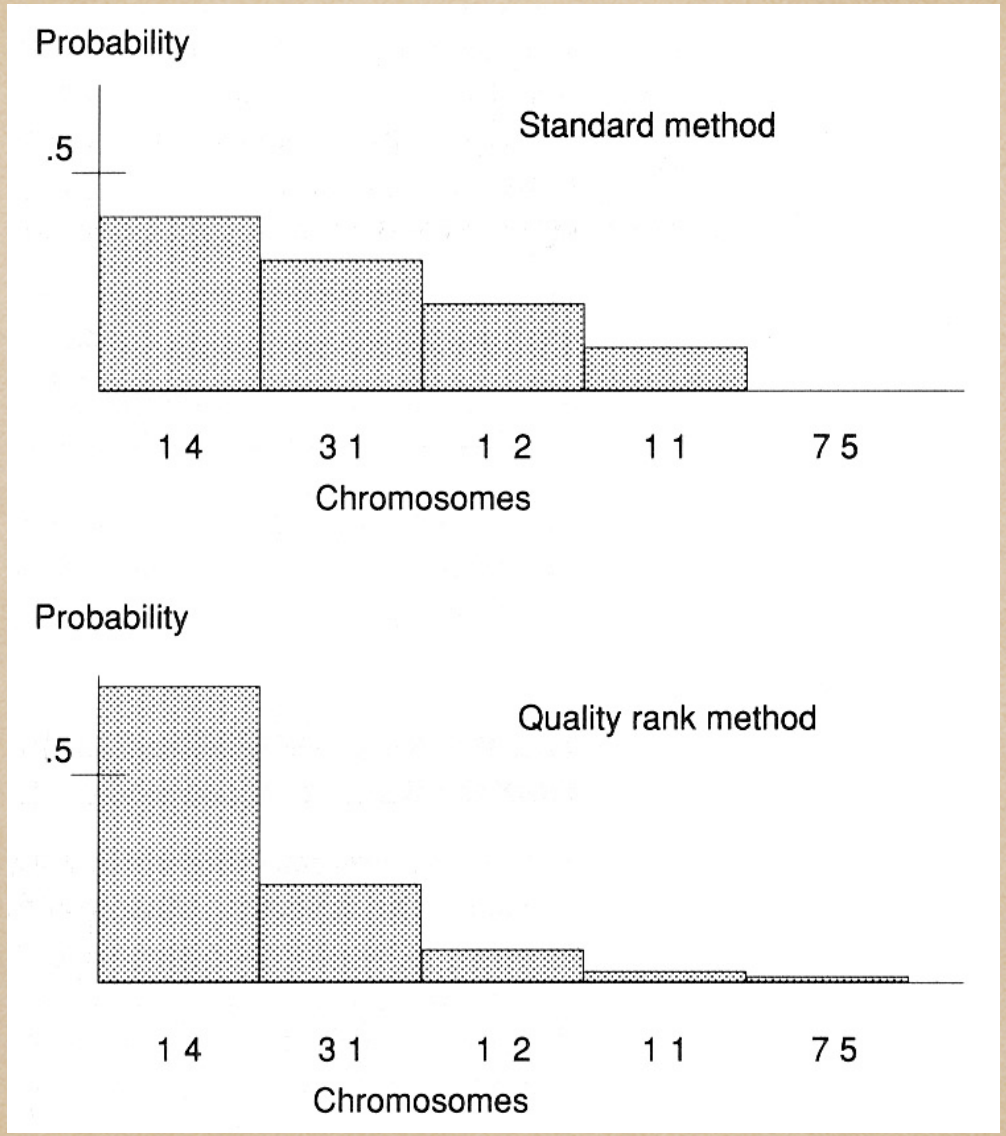
Quality Rank Method

Rank each with fixed Probability



Here 0.667 and then 1/3 of the remaining

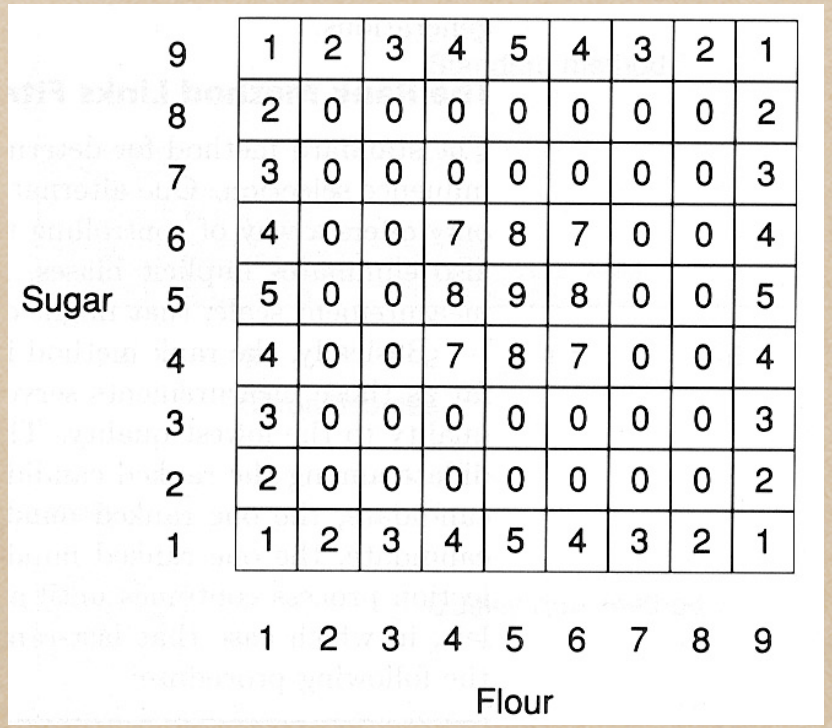
[Winston 1992]



[Winston 1992]


FLOUR
SUGAR
MAKE CAKE

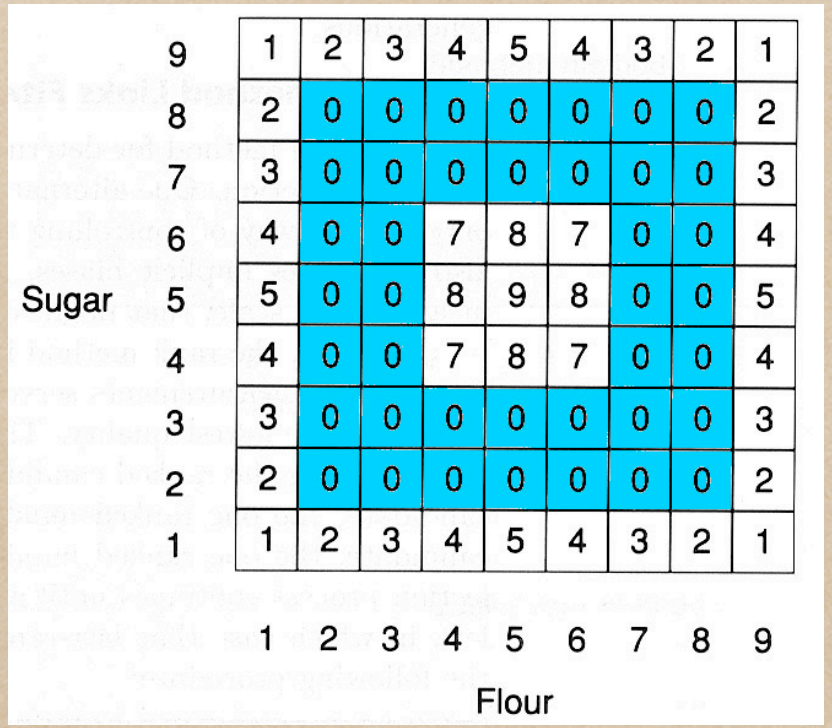
Example fitness function



Sugar	Flour 1	2	3	4	5	6	7	8	9
9	1	2	3	4	5	4	3	2	1
8	2	0	0	0	0	0	0	0	2
7	3	0	0	0	0	0	0	0	3
6	4	0	0	7	8	7	0	0	4
5	5	0	0	8	9	8	0	0	5
4	4	0	0	7	8	7	0	0	4
3	3	0	0	0	0	0	0	0	3
2	2	0	0	0	0	0	0	0	2
1	1	2	3	4	5	4	3	2	1

[Winston 1992]

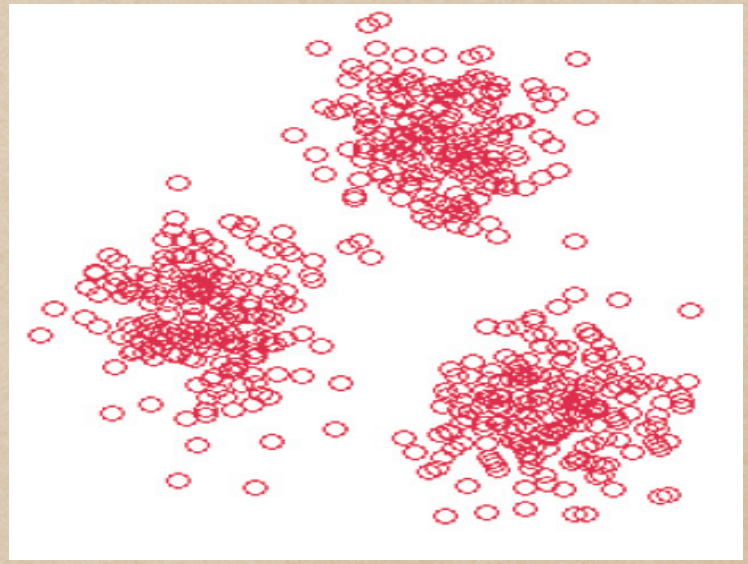
A fitness moat



[Winston 1992]

Rank Space Method

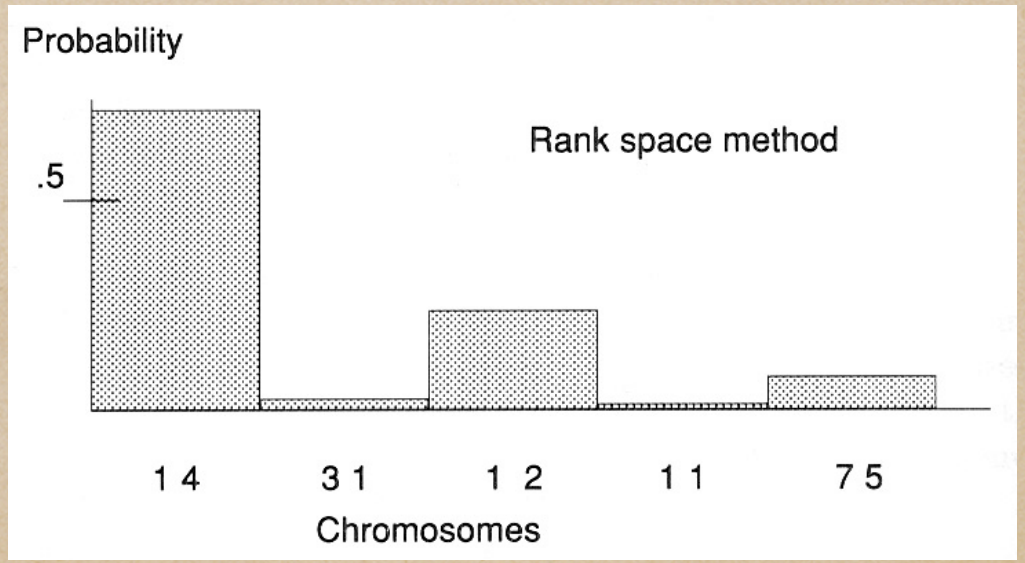
$$D_i \approx \sum_i \frac{1}{d_i^2}$$



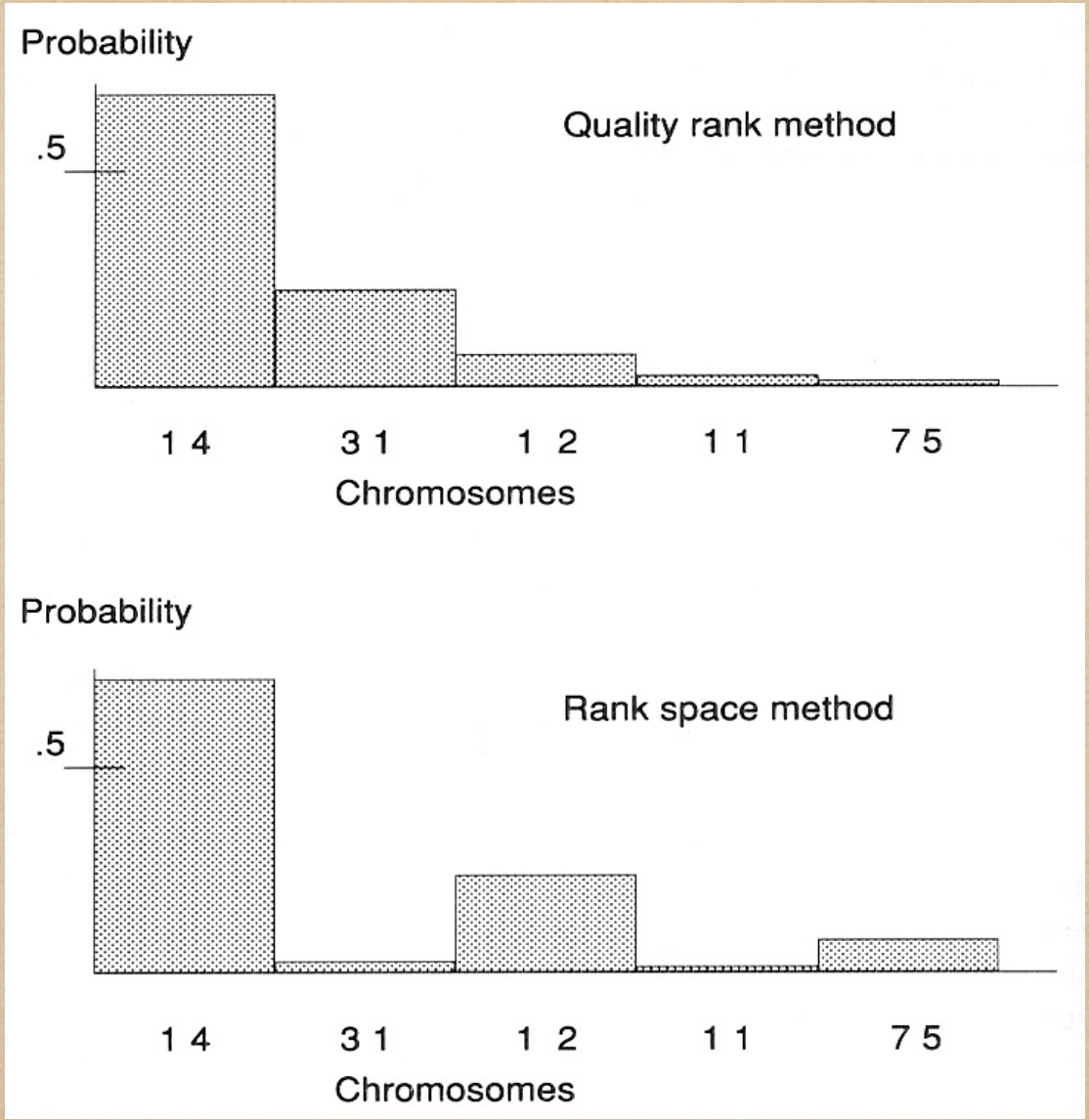
Also compute a diversity rank

Rank Space Method

$$\frac{Q_i}{\sum_j Q_j} + \sum_i \frac{1}{d_i^2} \approx \text{Combined Fitness}$$



[Winston 1992]



[Winston 1992]

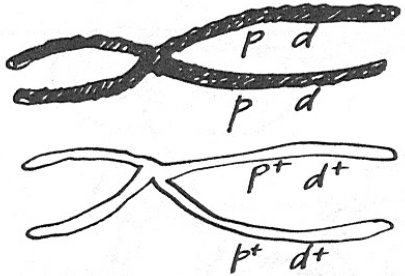
New Individuals?

During reproduction there are two genetic operators:

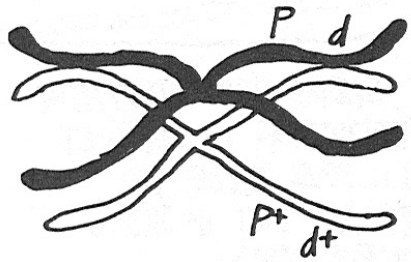
Crossover

Mutation

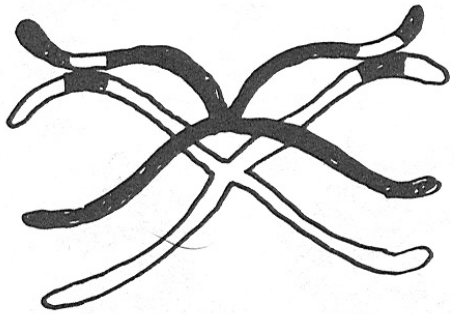
DURING MEIOSIS, HOMOLOGUES LINE UP WITH CORRESPONDING ALLELES OPPOSITE ONE ANOTHER.



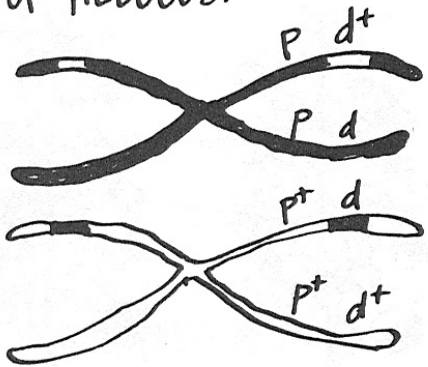
AT CERTAIN POINTS, SEEMINGLY "CHOSEN" AT RANDOM, THE CHROMOSOMES TOUCH:



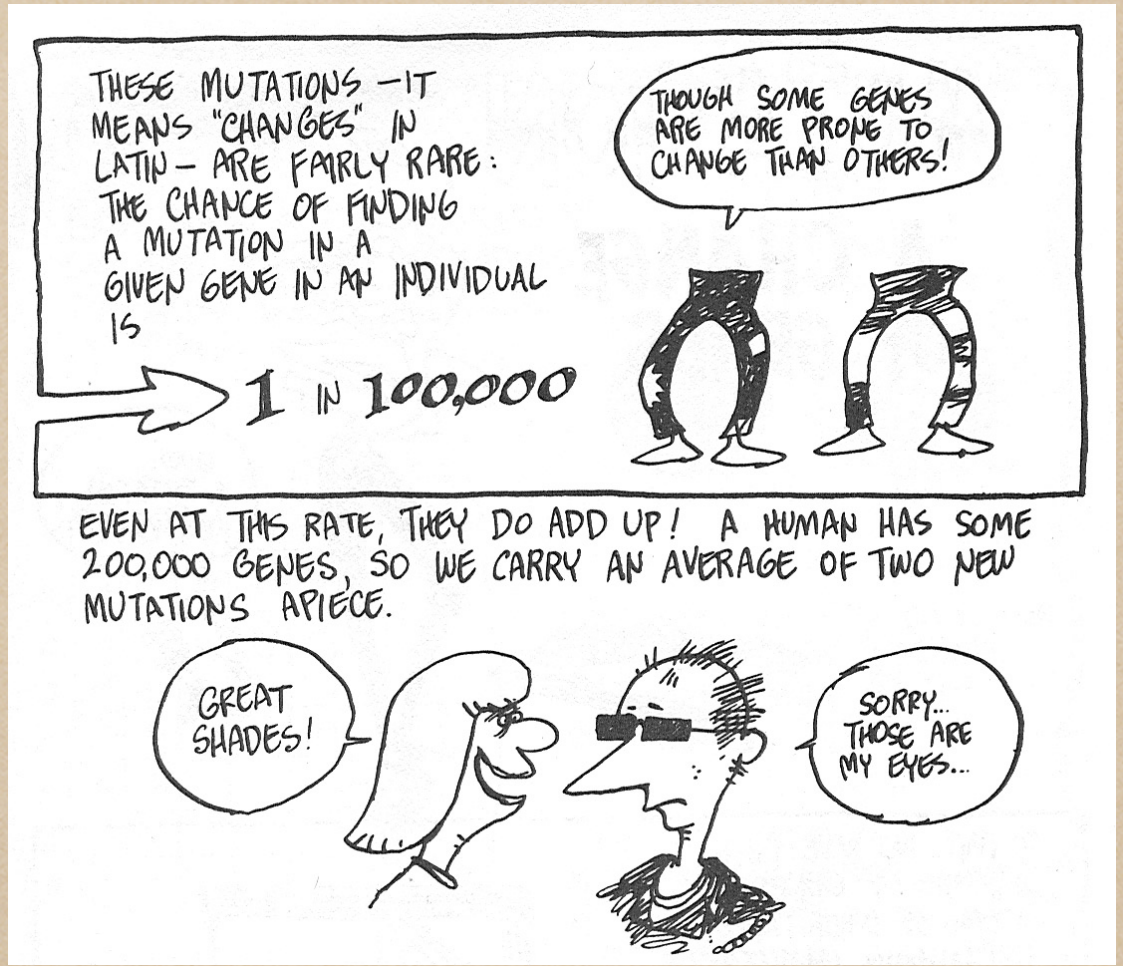
SOME SEGMENTS CROSS OVER:



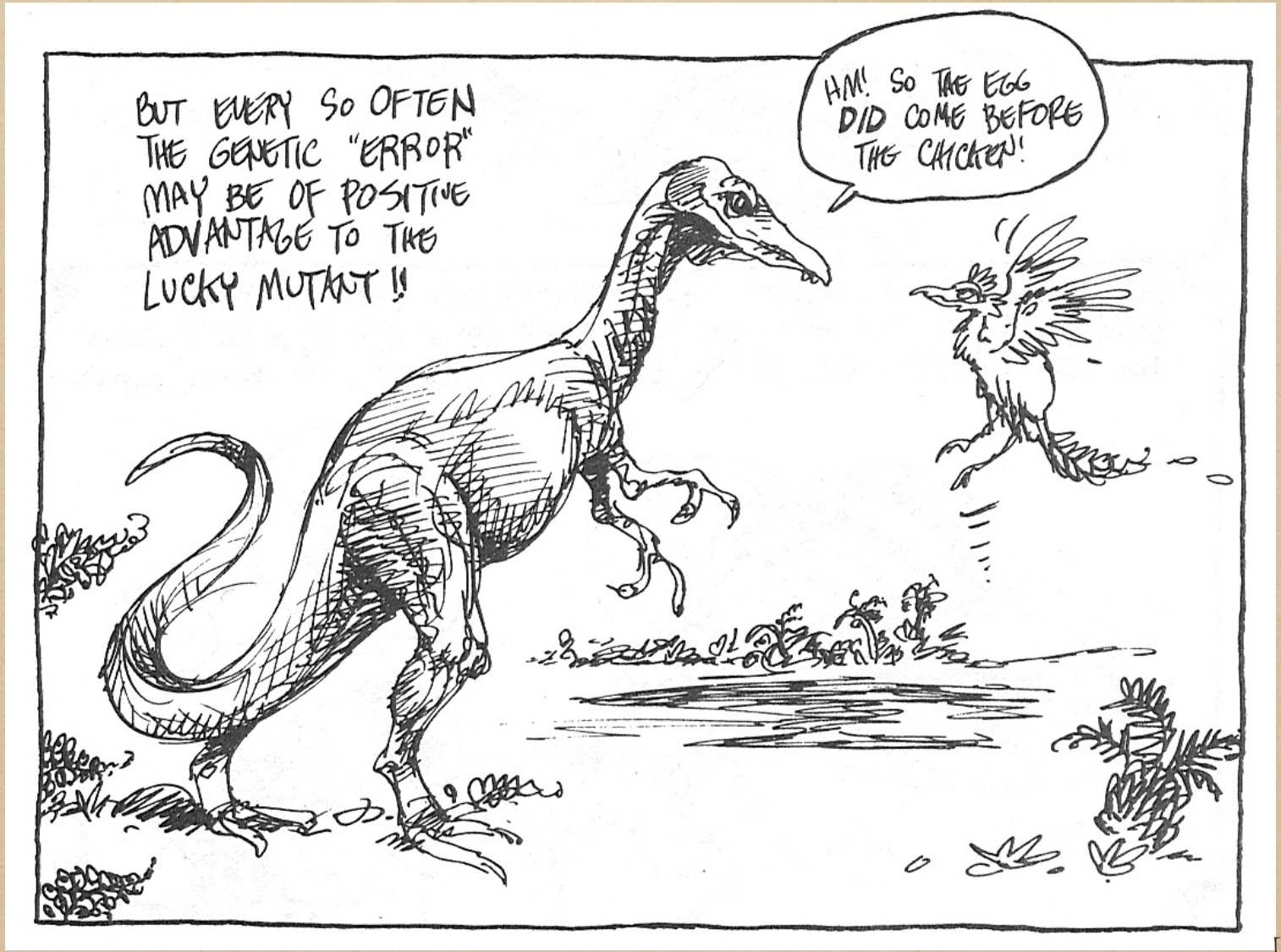
WHEN THEY SEPARATE, THEY HAVE NEW COMBINATIONS OF ALLELES.



[Gonick 1991]


THESE MUTATIONS – IT MEANS "CHANGES" IN LATIN – ARE FAIRLY RARE: THE CHANCE OF FINDING A MUTATION IN A GIVEN GENE IN AN INDIVIDUAL IS
1 IN 100,000
THOUGH SOME GENES ARE MORE PRONE TO CHANGE THAN OTHERS!
EVEN AT THIS RATE, THEY DO ADD UP! A HUMAN HAS SOME 200,000 GENES, SO WE CARRY AN AVERAGE OF TWO NEW MUTATIONS APIECE.
GREAT SHADES!
SORRY... THOSE ARE MY EYES...

[Gonick 1991]


BUT EVERY SO OFTEN THE GENETIC "ERROR" MAY BE OF POSITIVE ADVANTAGE TO THE LUCKY MUTANT !!
HM! SO THE EGG DID COME BEFORE THE CHICKEN!

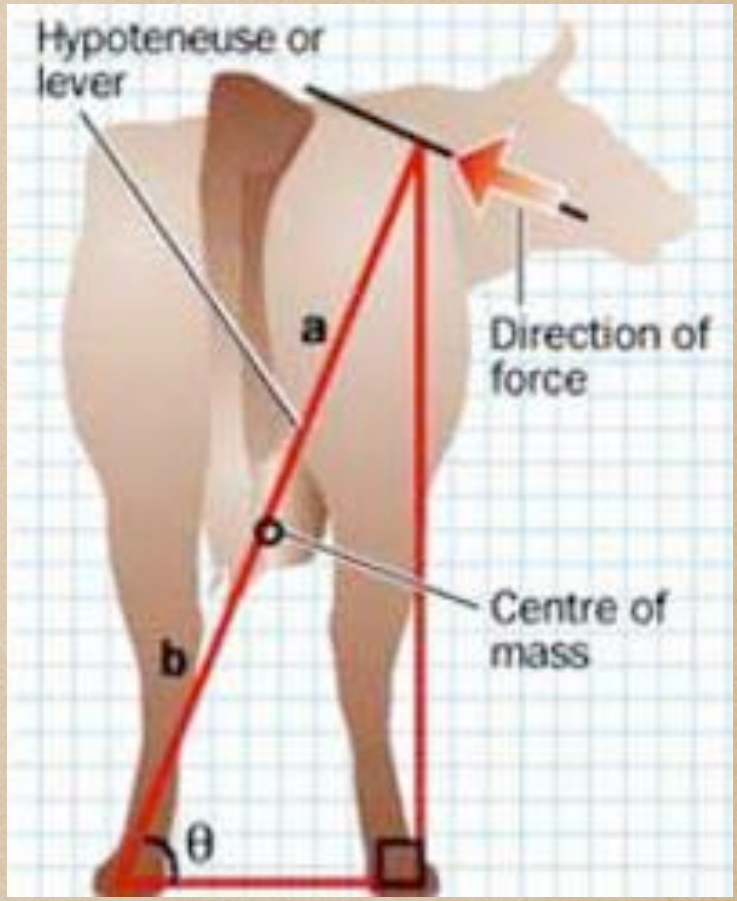
[Gonick 1991]

[http://www.norcalblogs.com/bumpkins/]

[AP Photo/Shane Jennings]

"Cows are heavy, man!" said Peter Park, a college student who tried to cow-tip on a trip to central Illinois. "There were four of us, but we had to run at the cow from 30 feet away and push as hard as we could to get it to even notice us. It woke up and looked at us like we were a bunch of flies, and walked away."

[http://seedmagazine.com/content/article/udder_impossibility/]



[http://www.englishwithrae.com/]

Fitness landscape

- Differs from heuristic state space search
  - no estimates of "effort" to move to new state
  - no strict ordering of states
- Has peaks, valleys, discontinuities
  - both real and introduced
  - e.g binary vs. gray coding

Binary vs. Gray Coding

0000 ------ 0000

0001 ------ 0001

0010 ------ 0011

0011 ------ 0010

0100 ------ 0110

0101 ------ 0111

0110 ------ 0101

0111 ------ 0100

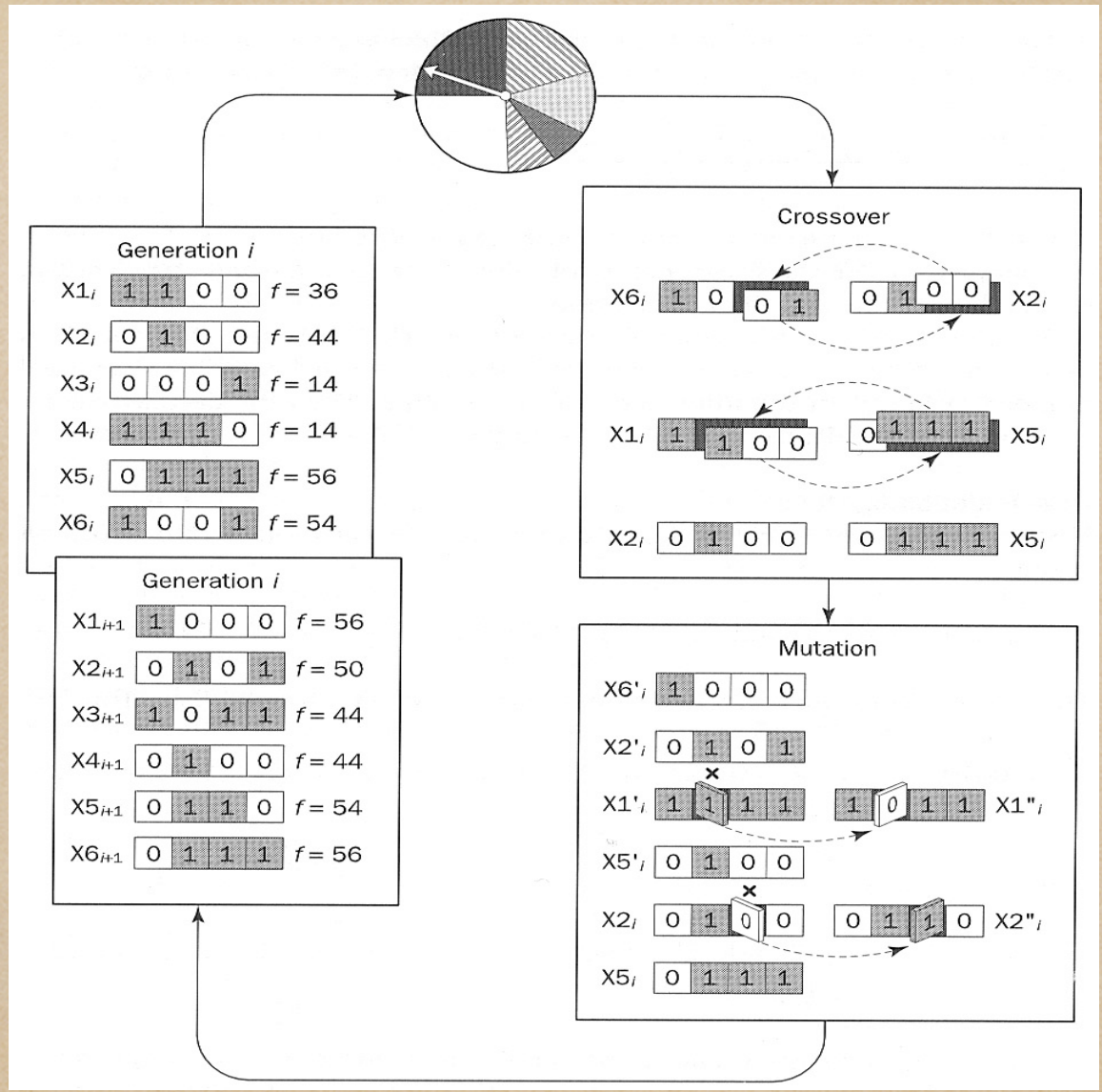
1000 ------ 1100

Binary vs. Gray Coding

0000	------	0000
0001	------	0001
0010	------	0011
0011	------	0010
0100	------	0110
0101	------	0111
0110	------	0101
0111	------	0100
1000	------	1100

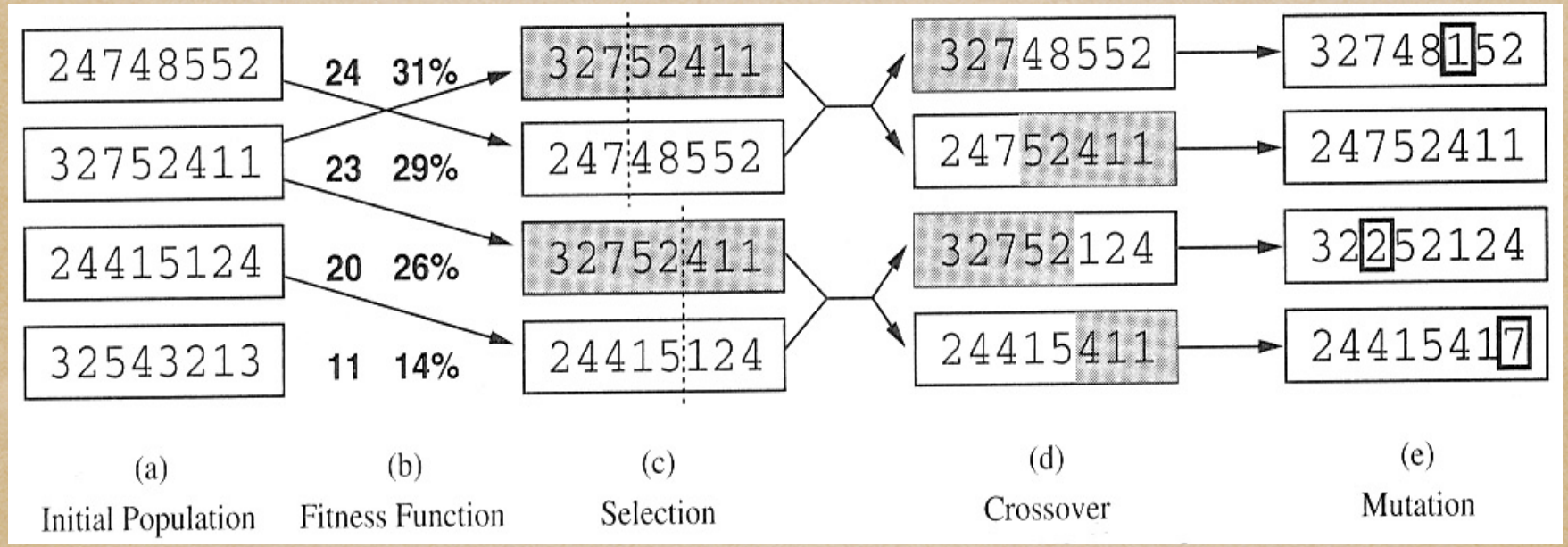
Hamming Distance

4 versus 1!

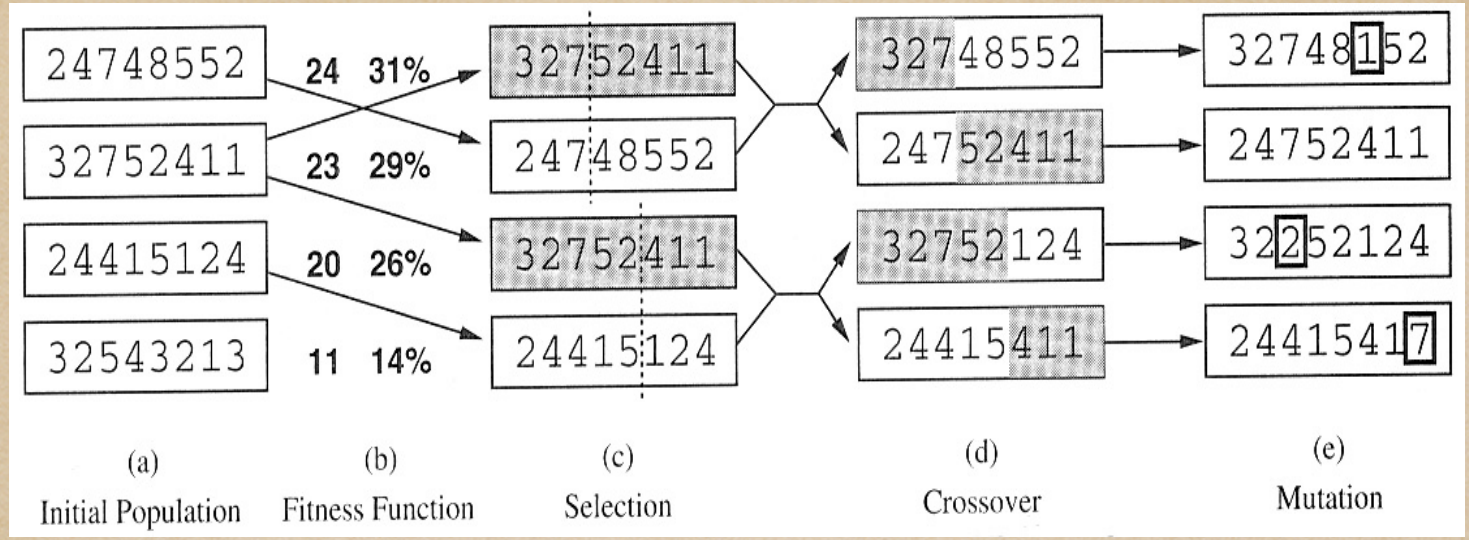


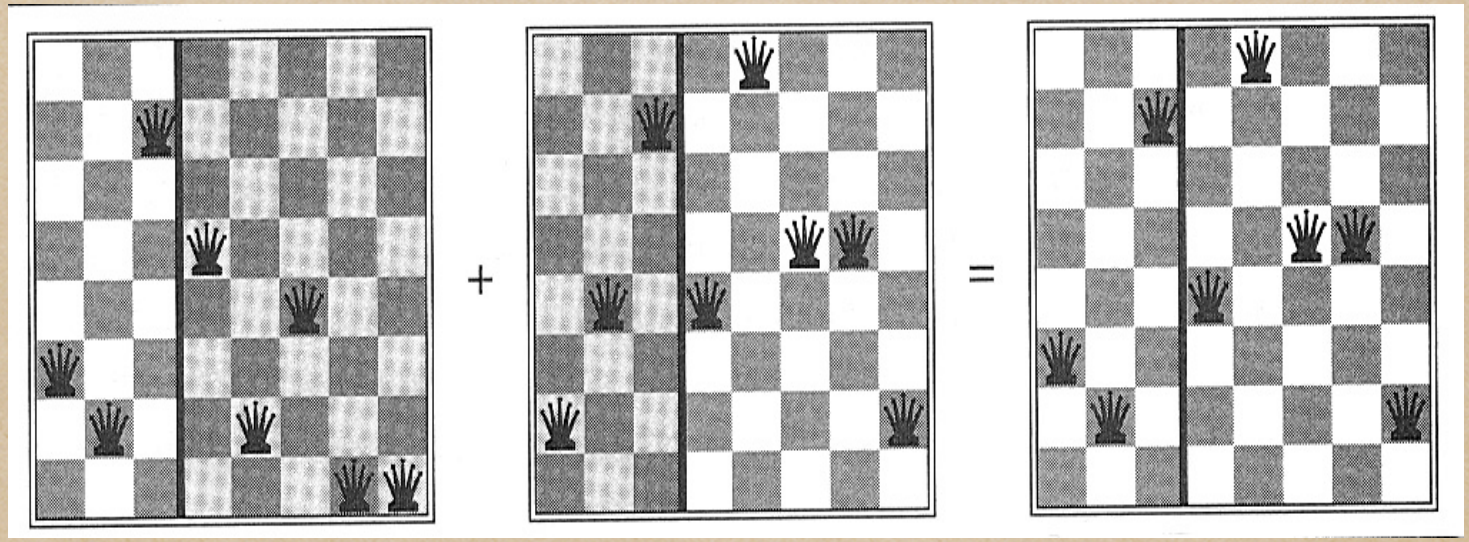
[Negnevitsky 2002]

Example

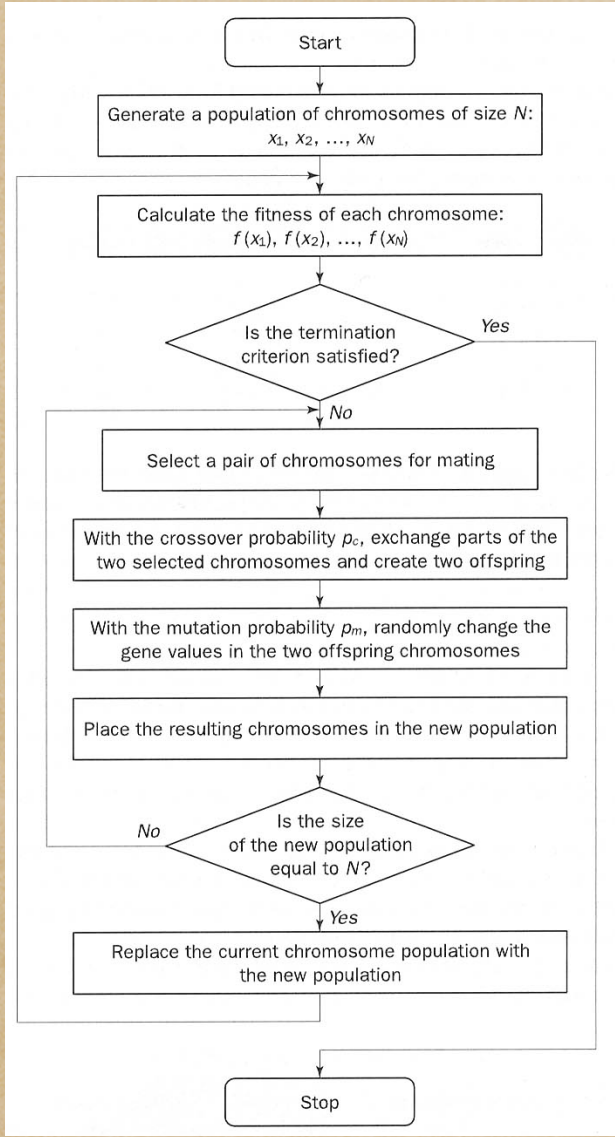


[Russel & Norvig 2003]





[Russel & Norvig 2003]




```
procedure GeneticAlgorithm

{ t = 0;
  initialize P(t);
  evaluate chromosomes in P(t);
  while (not satisfied) do
   { t += 1;
     select P(t) from P(t-1);
     recombine chromosomes in P(t);
     evaluate chromosomes in P(t);
   }
}
```

[Negnevitsky 2002, Grefenstette 1990]

the GENESIS system

Genes can be bit sequence or multiple floats
Can specify or randomly initialize population
Selection by standard or ranked method
You write the evaluation function in C

```
double eval(str, length, vect, genes)
  char str[];          /* bits as string  */
  int length;          /* length of str   */
  double vect[];       /* float genes     */
  int genes;           /* length of vect  */

{ /* your function here */ }
```


program
writing 101
#;!?<>
i didn't
know the
comp. could
play...
COMP
WINS!
HOLD
COMPUTER
WINS!
HOLD

Using Genesis for BlackJack

BJ eval function

- Use bit vector strategy

  1111111111111111000000

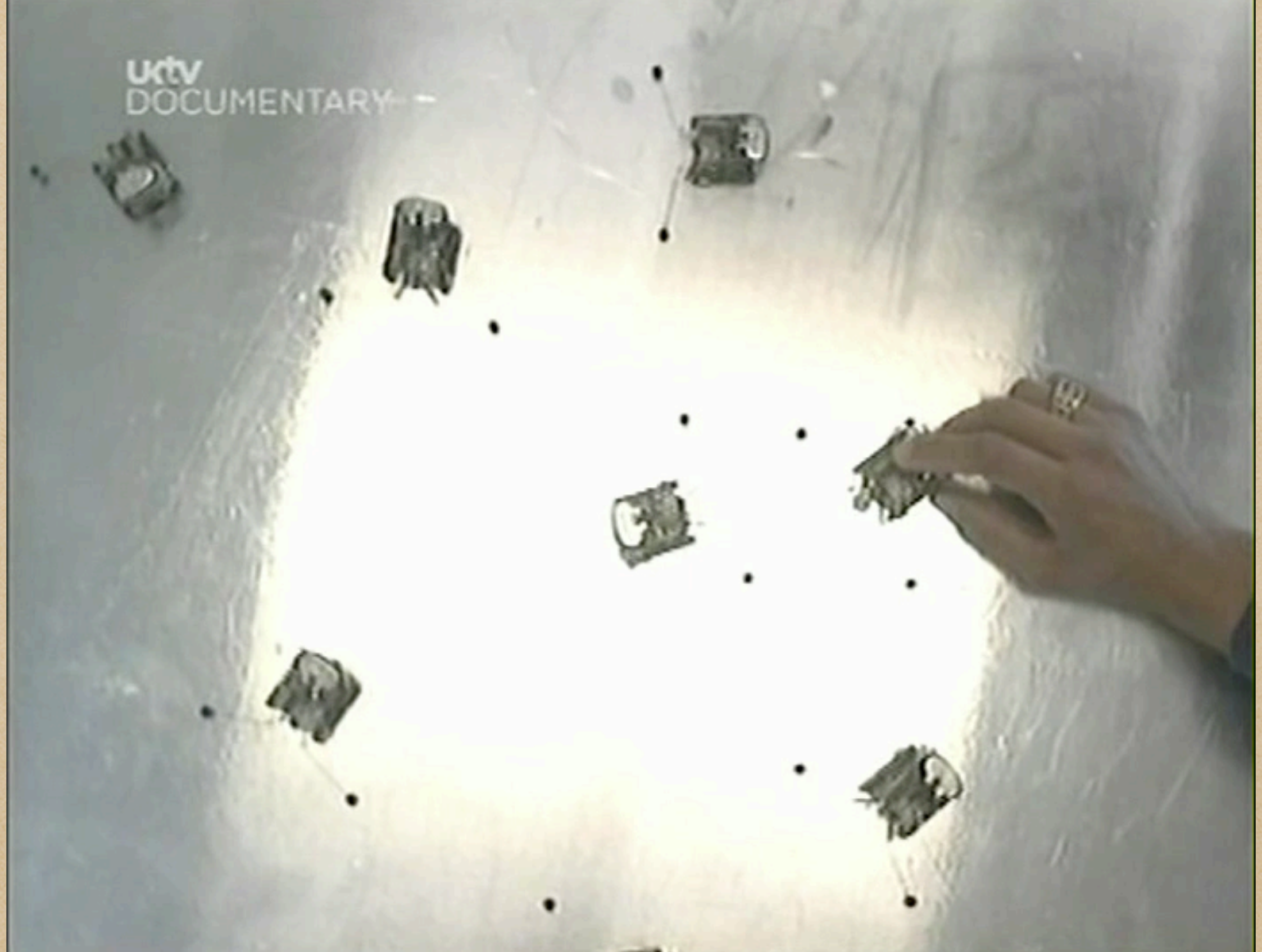
- Play BJ against dealer
- Report losing %age

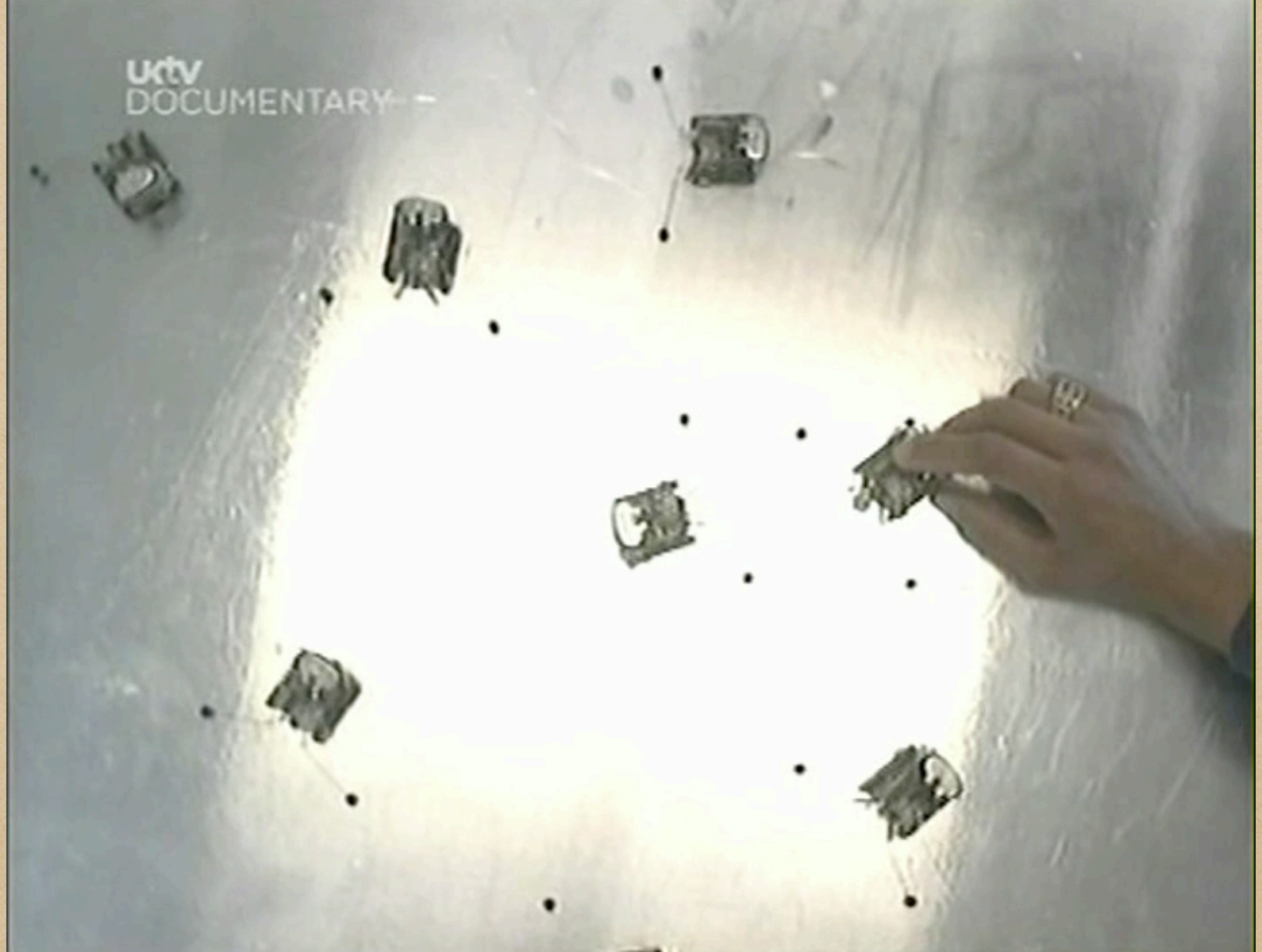
Genesis Setup

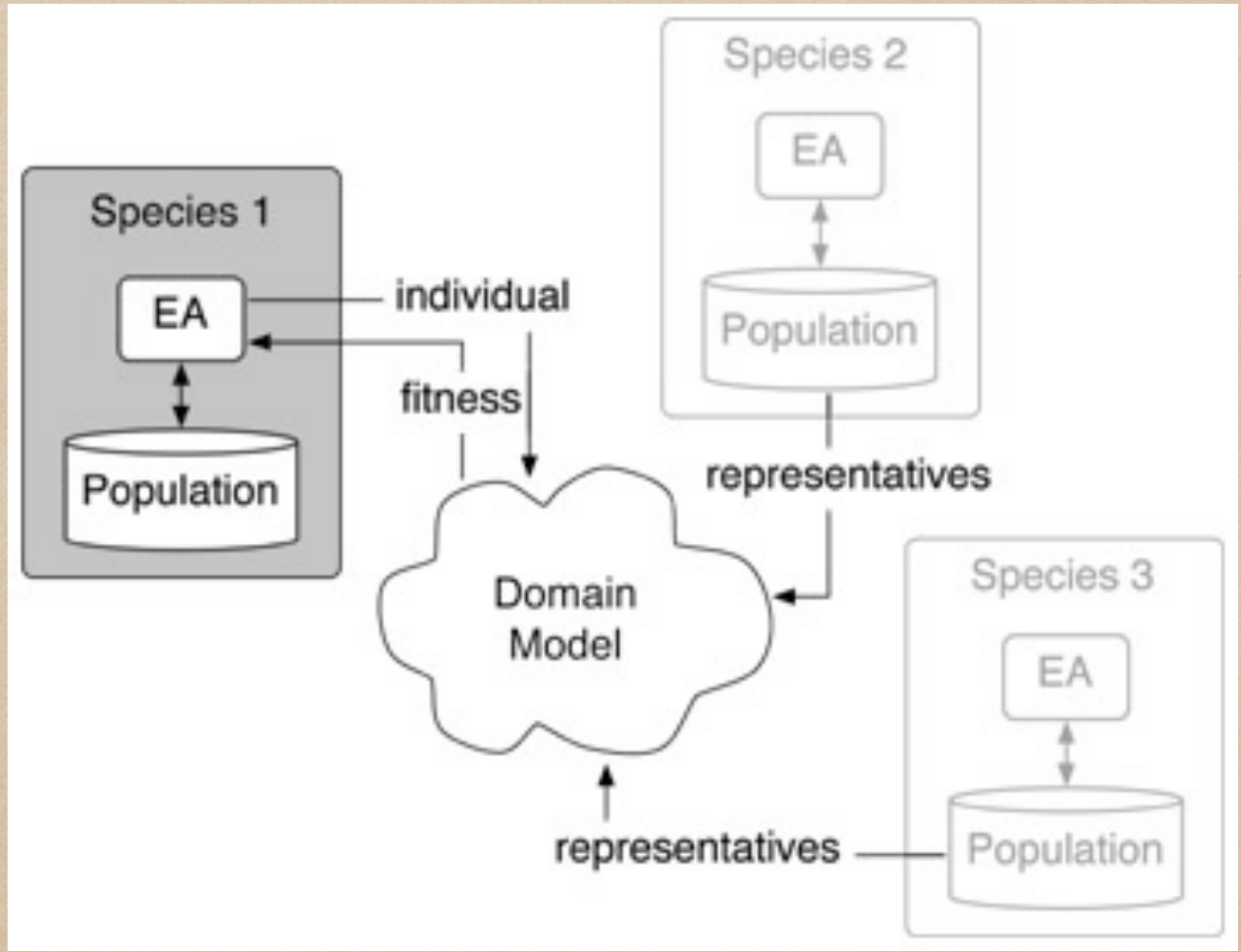
```
File suffix []: bj
Floating point representation [y]: n
Experiments [1]:
Total Trials [1000]: 5000
Population Size [50]:
Structure Length [30]: 22
Crossover Rate [0.6]:
Mutation Rate [0.001]: 0.01
Generation Gap [1.0]:
Scaling Window [5]:
Report Interval [100]:
Structures Saved [10]:
Max Gens w/o Eval [2]:
Dump Interval [0]: 5
Dumps Saved [0]: 5
Options [cel]:
Random Seed [123456789]:
Rank Min [0.75]:
```

Individuals After 150 Generations

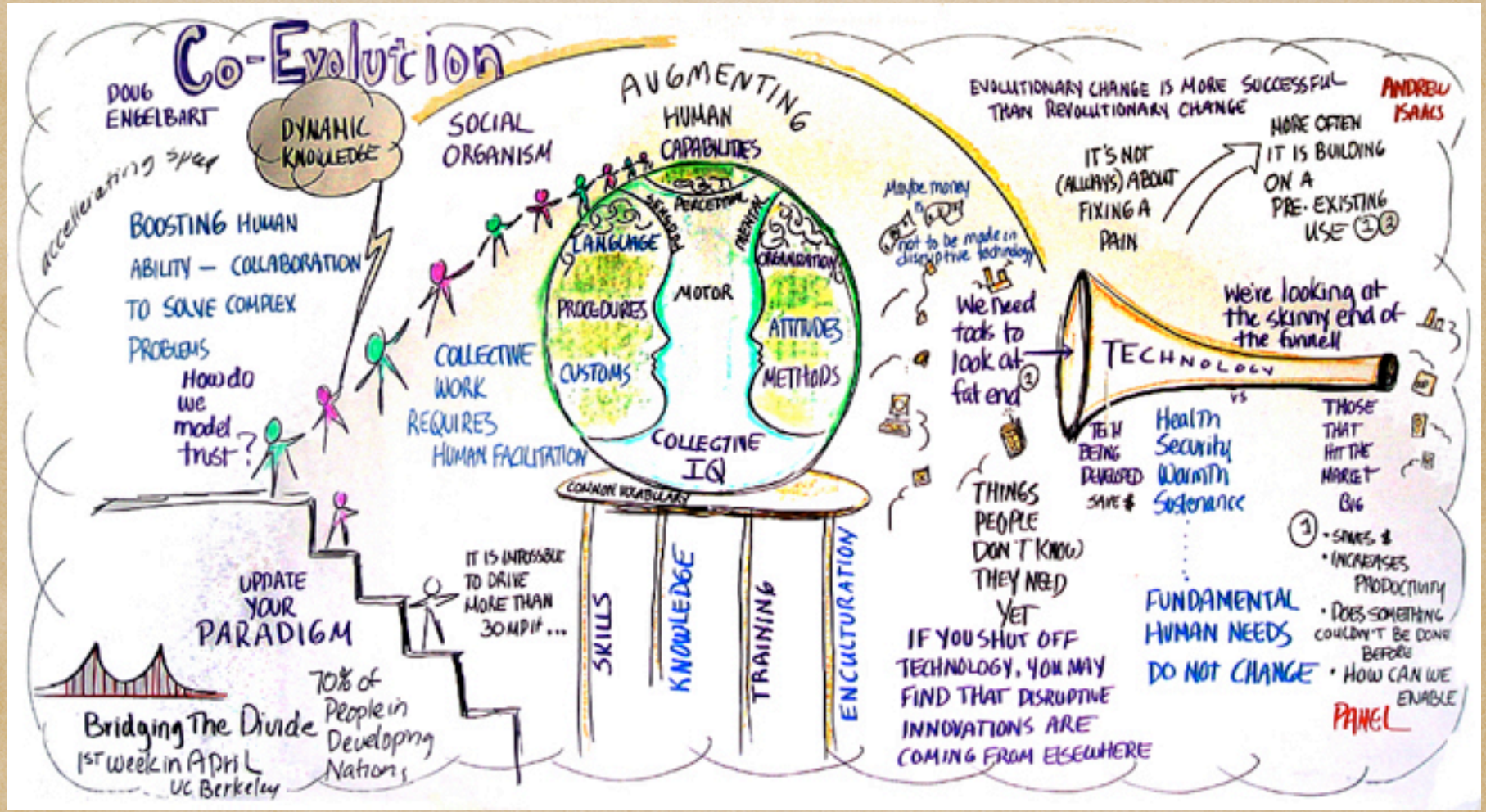
000010001111011100000	4.5000e-01	111	2551
001011001111011100000	4.6000e-01	133	2692
100011011110111100000	4.6000e-01	74	1831
000011001111011100000	4.3000e-01	83	2018
000011001110111100000	4.4000e-01	3	122
100000010111001010000	4.4000e-01	18	500
100000010111001000000	4.6000e-01	43	1114
100000010111111000000	4.7000e-01	33	863
100011001110111100000	4.5000e-01	51	1309
100011011010011000000	4.6000e-01	92	2219


uktv
DOCUMENTARY


uktv
DOCUMENTARY



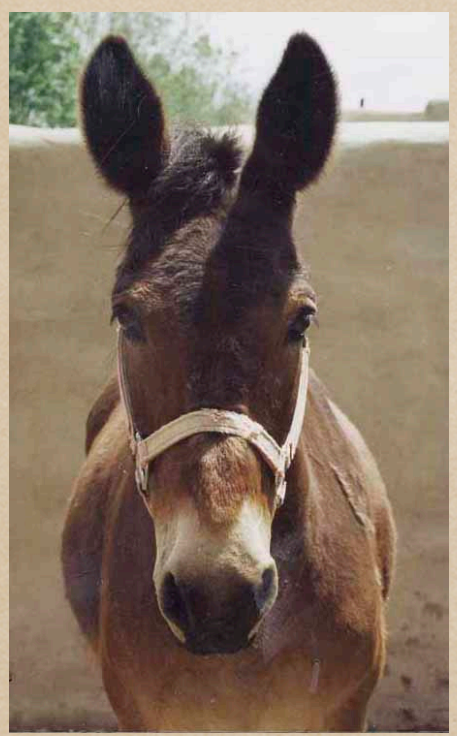
[http://www.nrl.navy.mil/aic/iss/aas/Co-evolution.php]

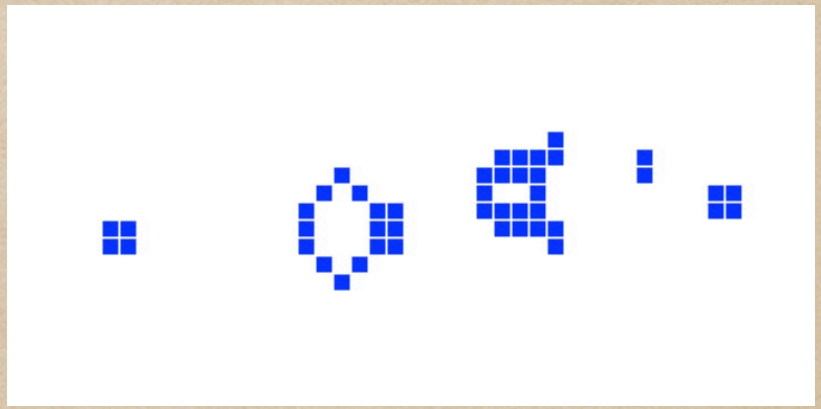


[http://www.almaden.ibm.com/coevolution/gallery.shtml]

[http://www.almaden.ibm.com/coevolution/gallery.shtml]

Which is alive?





Suppose we wanted to learn something about right triangles.

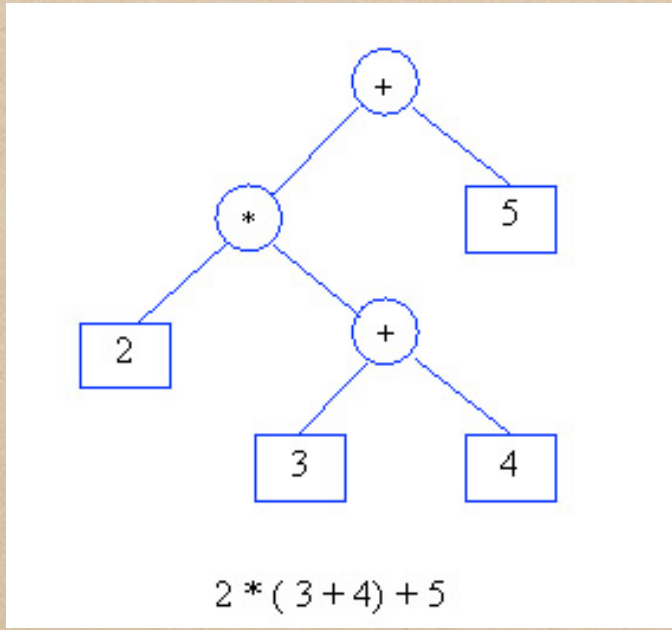
Side *a*	Side *b*	Hypotenuse *c*	Side *a*	Side *b*	Hypotenuse *c*
3	5	5.830952	12	10	15.620499
8	14	16.124515	21	6	21.840330
18	2	18.110770	7	4	8.062258
32	11	33.837849	16	24	28.844410
4	3	5.000000	2	9	9.219545

That is, learn a function that takes a and b and predicts the hypotenuse.

[Negnevitsky 2002]

Genetic Programming

Representing Equations



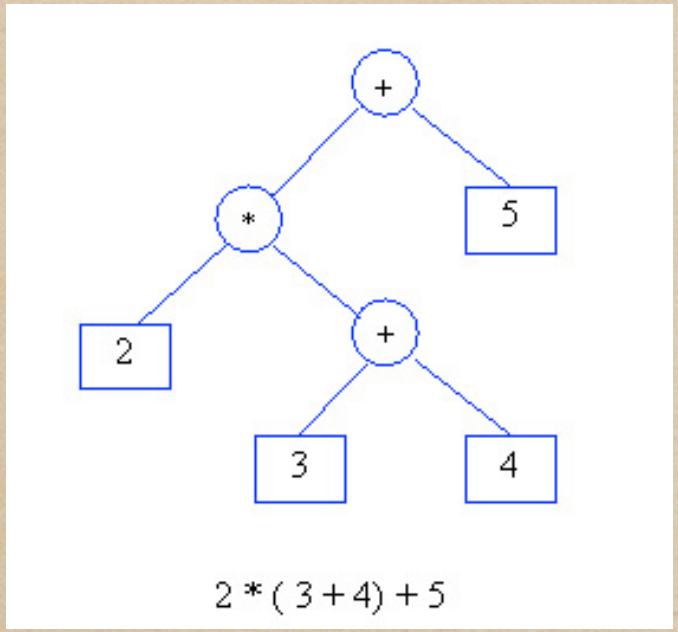
Simple expression tree

(could use variables a & b)

Abstract Syntax Tree

(parse tree / compilers)

Representing Equations



Simple expression tree

(could use variables a & b)

Abstract Syntax Tree

(parse tree / compilers)

(+ (* 2 (+ 3 4)) 5)

LISP S-expressions

$$(/\ (-\ (sqrt\ (+\ (*\ a\ a)\ (-\ a\ b)))\ a)\ (*\ a\ b)),$$

which is equivalent to

$$\frac{\sqrt{a^2 + (a - b)} - a}{ab},$$

and

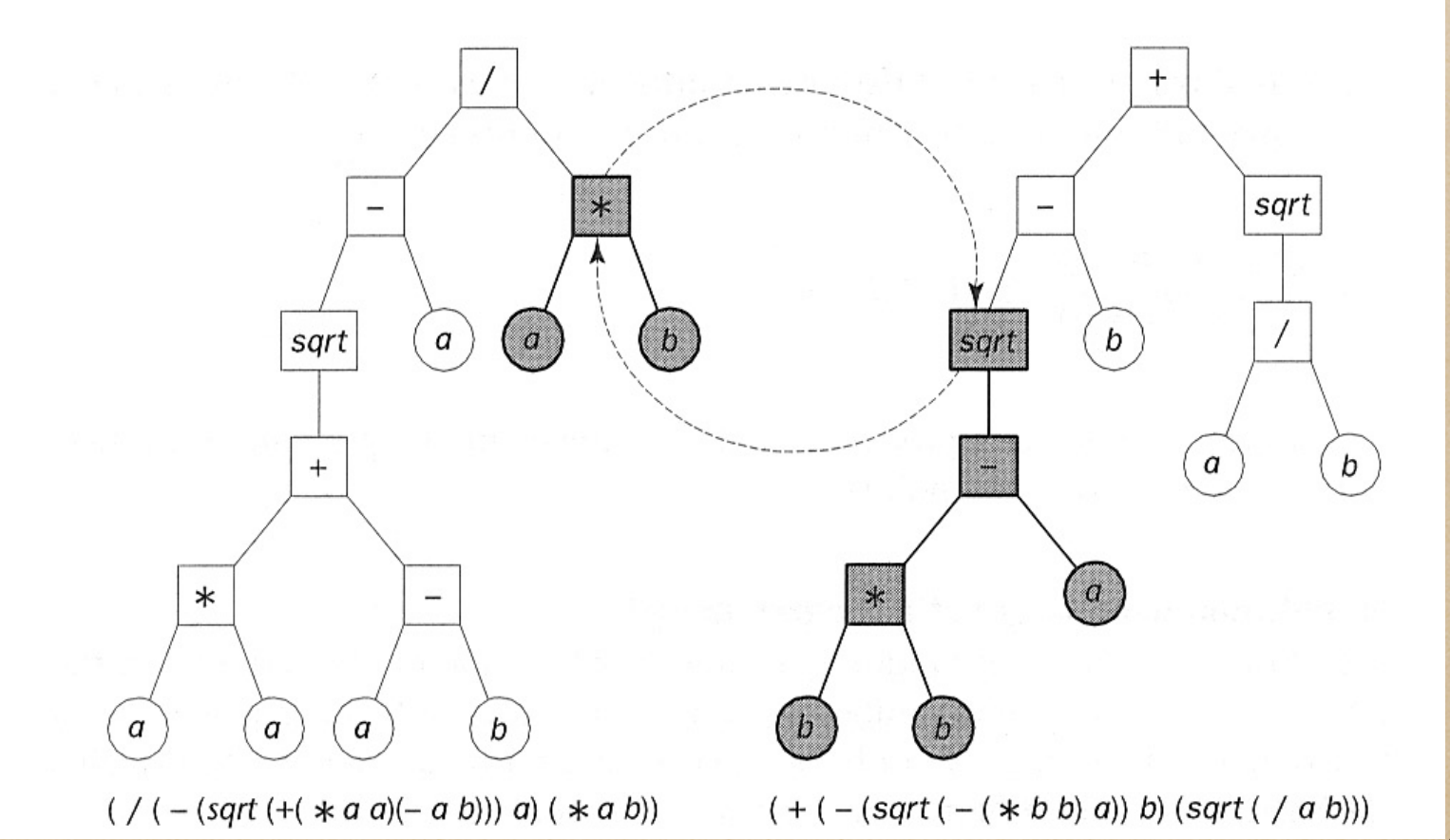
$$(+\ (-\ (sqrt\ (-\ (*\ b\ b)\ a))\ b)\ (sqrt\ (/\ a\ b))),$$

which is equivalent to

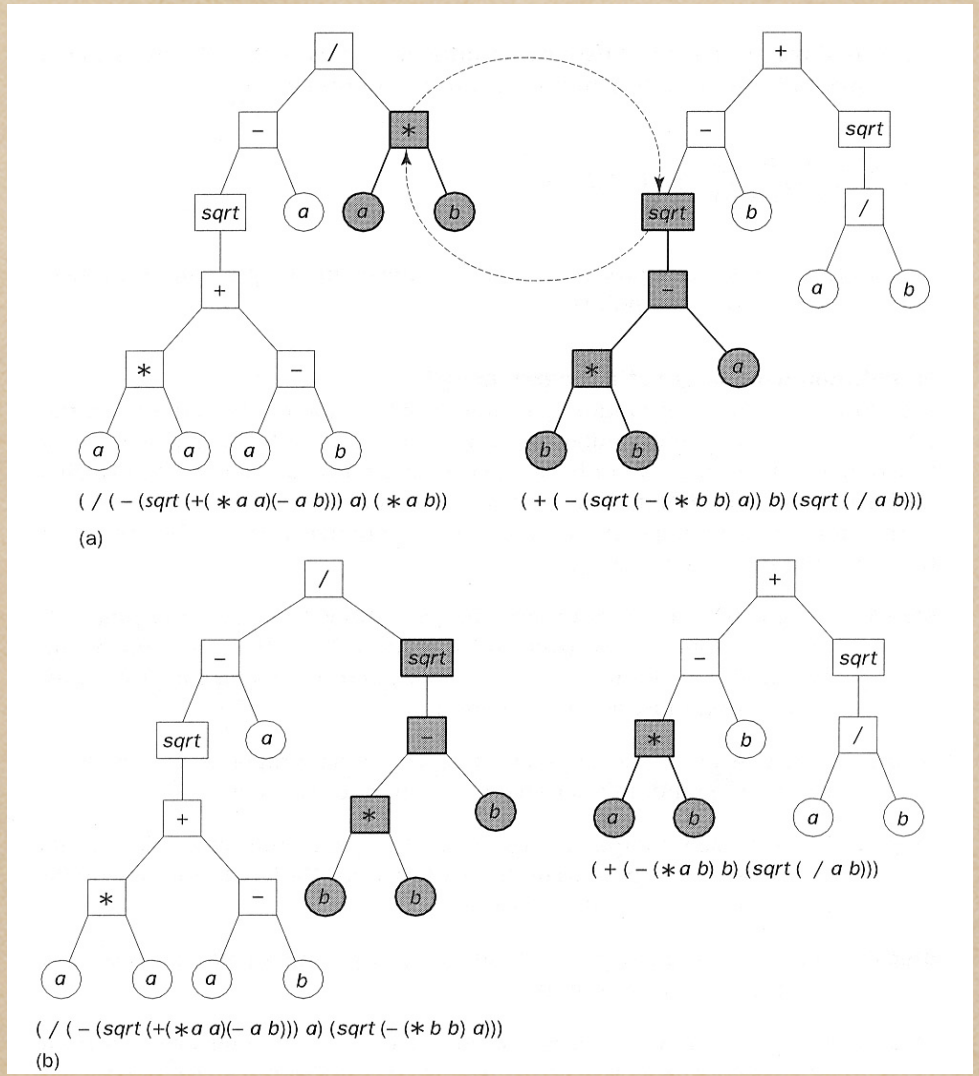
$$\left(\sqrt{b^2 - a} - b\right) + \sqrt{\frac{a}{b}}.$$

[Negnevitsky 2002]

Crossover

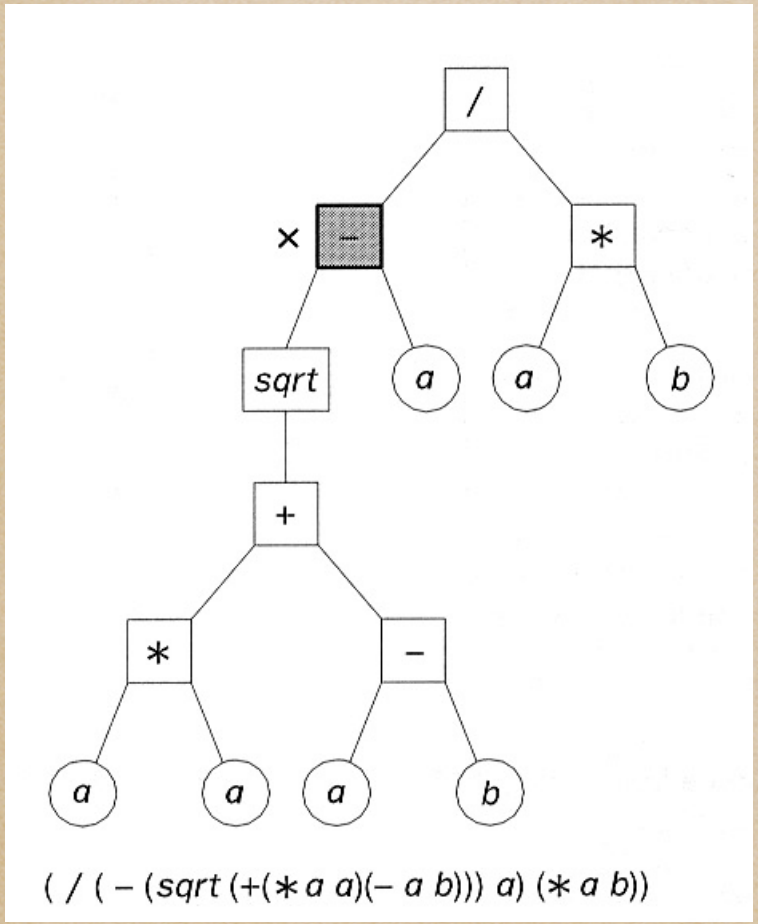


[Negnevitsky 2002]



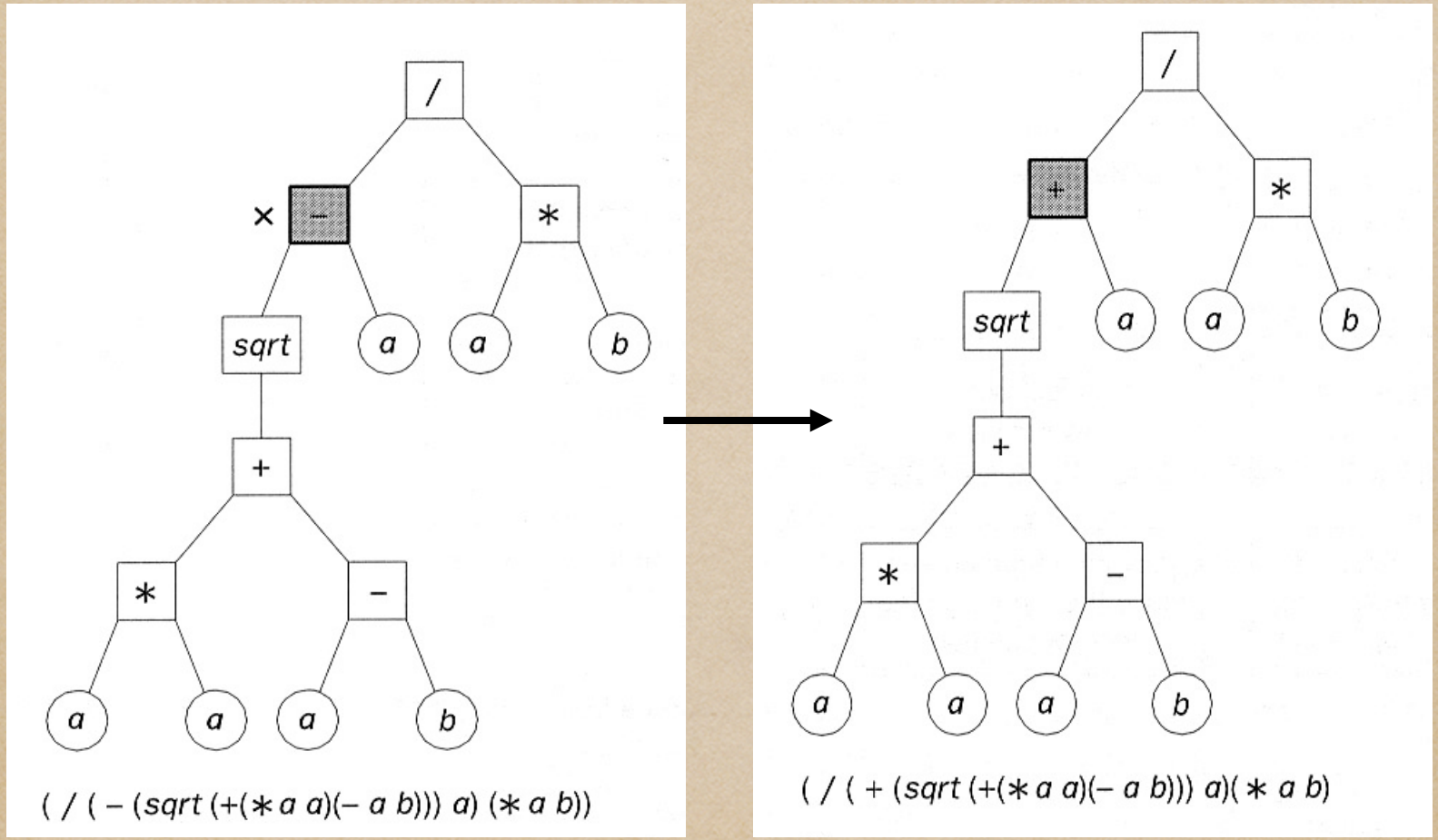
[Negnevitsky 2002]

Mutation - operators



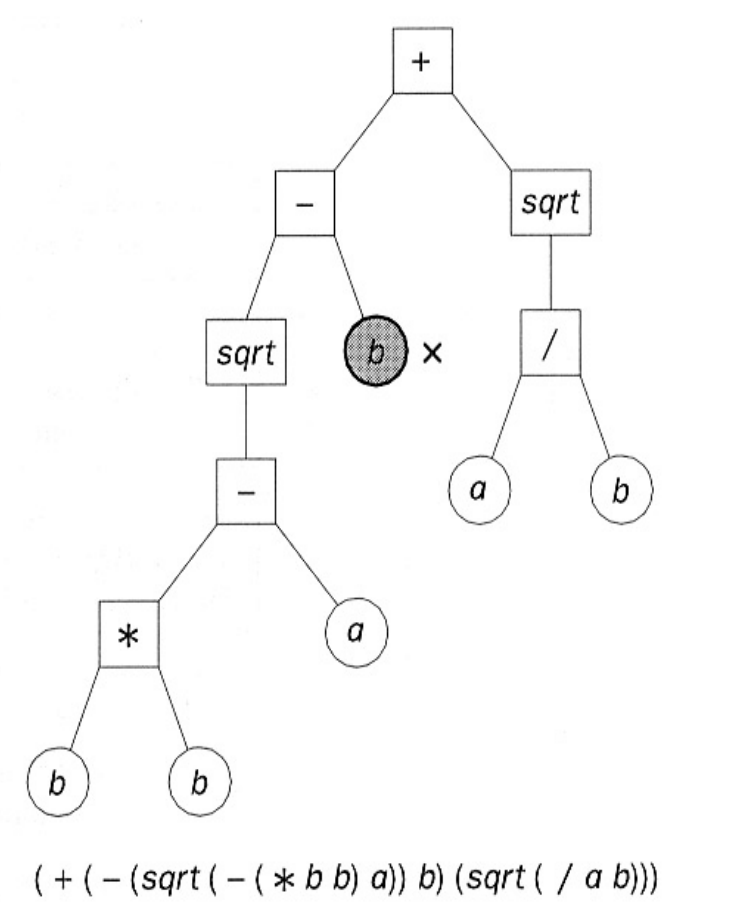
[Negnevitsky 2002]

Mutation - operators



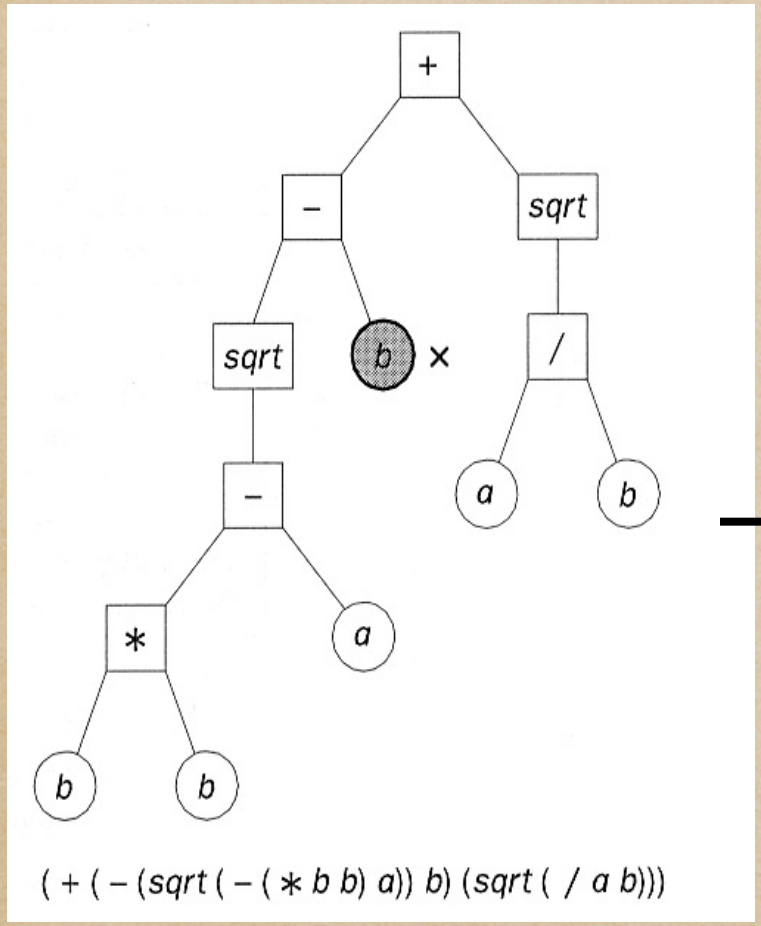
[Negnevitsky 2002]

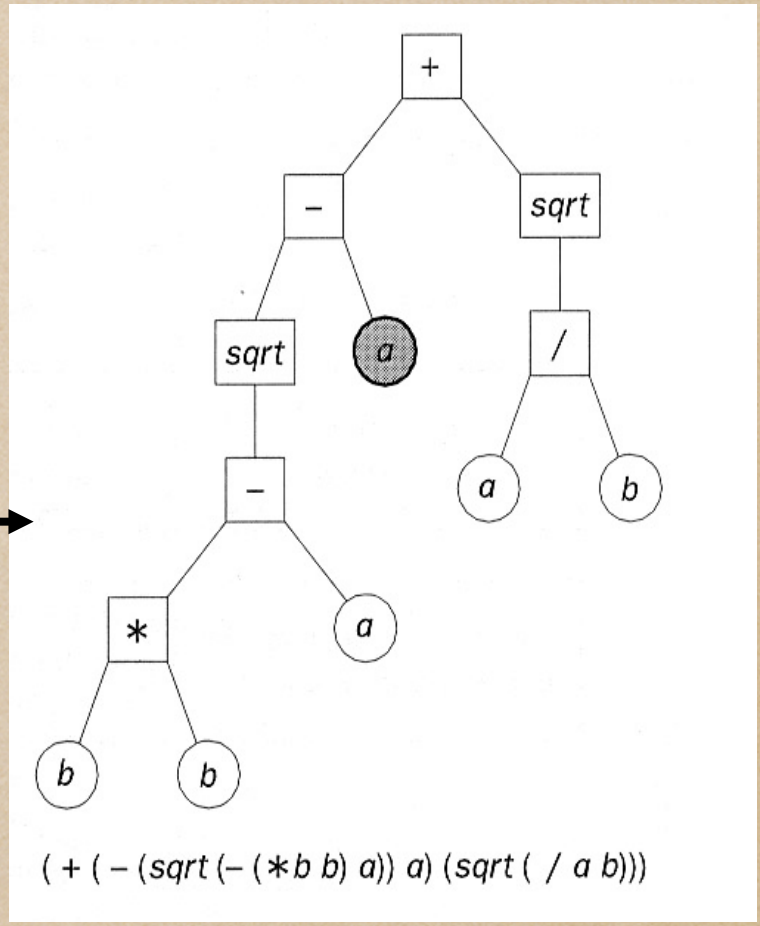
Mutation - operands



[Negnevitsky 2002]

Mutation - operands





[Negnevitsky 2002]

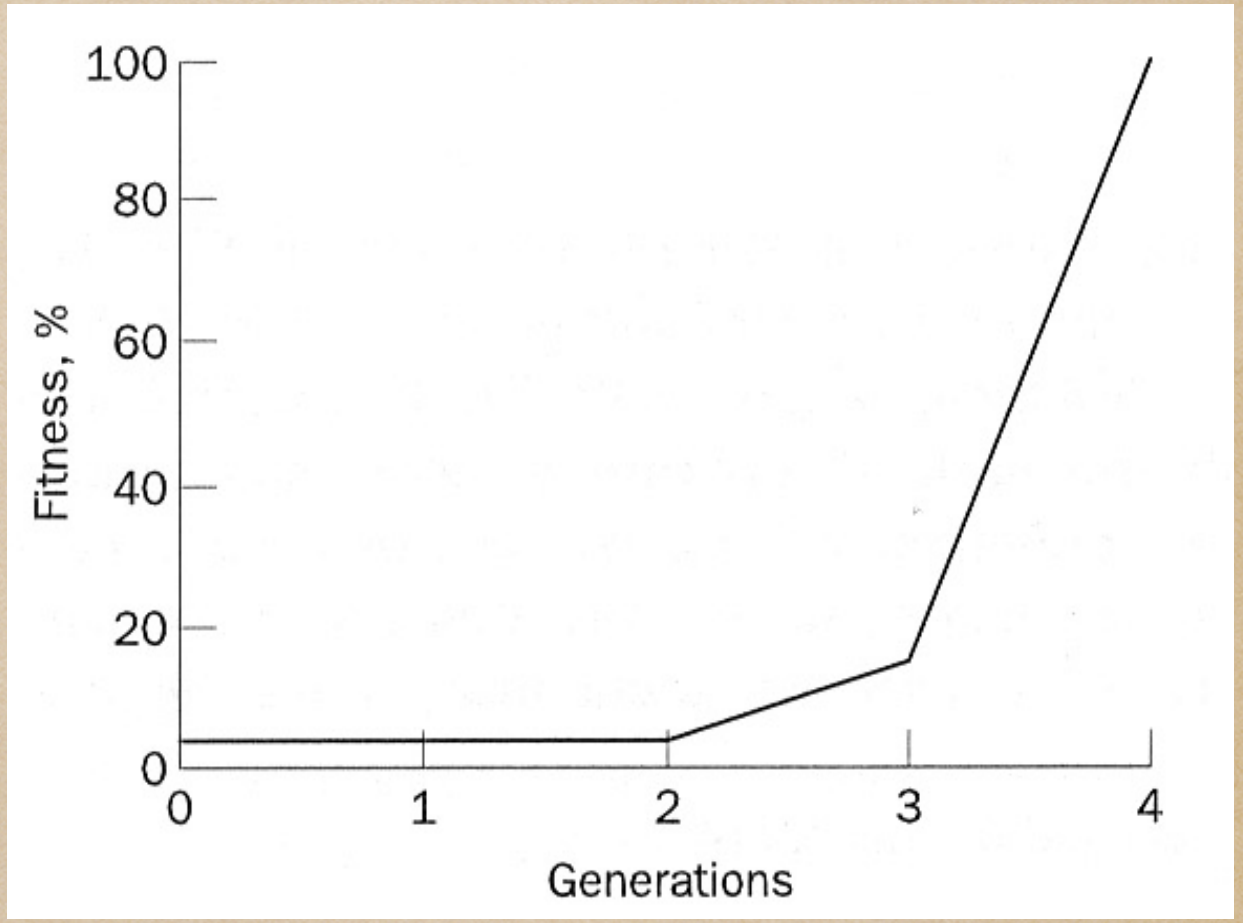
Fitness function

Side *a*	Side *b*	Hypotenuse *c*	Side *a*	Side *b*	Hypotenuse *c*
3	5	5.830952	12	10	15.620499
8	14	16.124515	21	6	21.840330
18	2	18.110770	7	4	8.062258
32	11	33.837849	16	24	28.844410
4	3	5.000000	2	9	9.219545

Test on sample instances,
compute similarity to desired value.

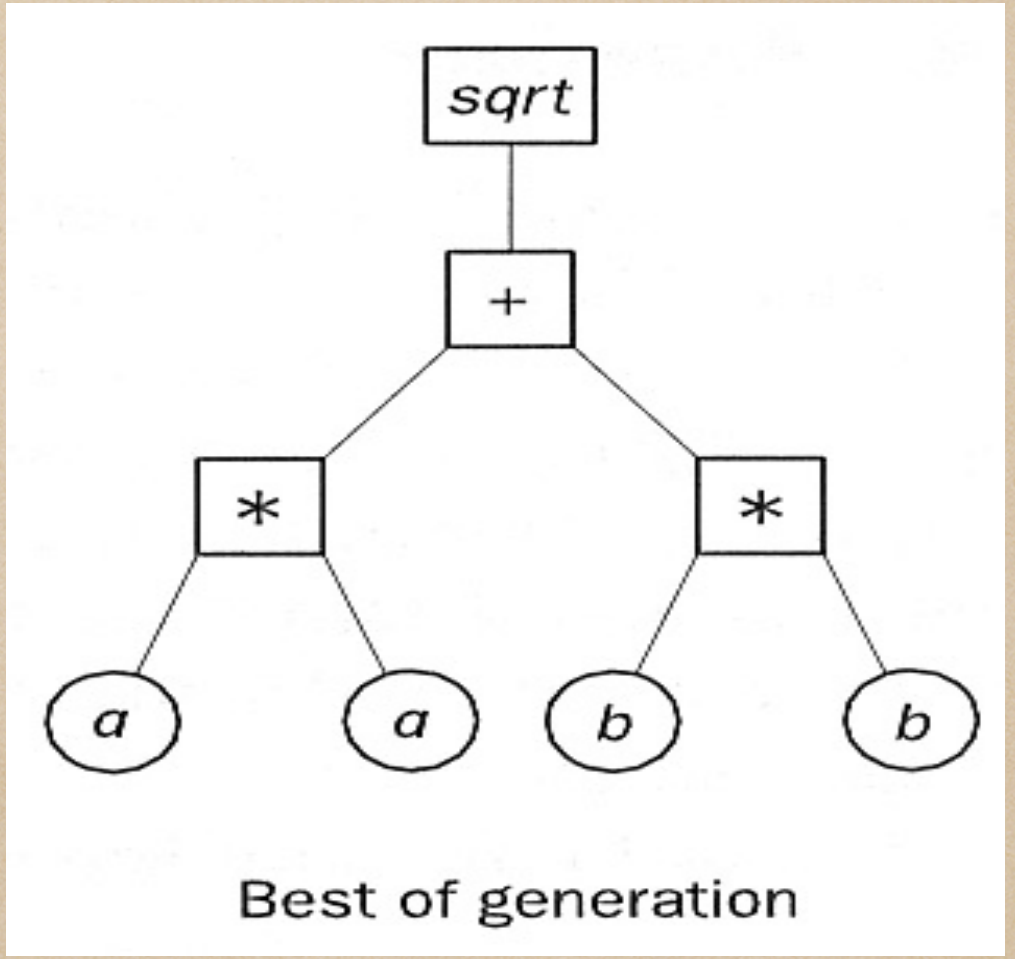
[Negnevitsky 2002]

Results



[Negnevitsky 2002]

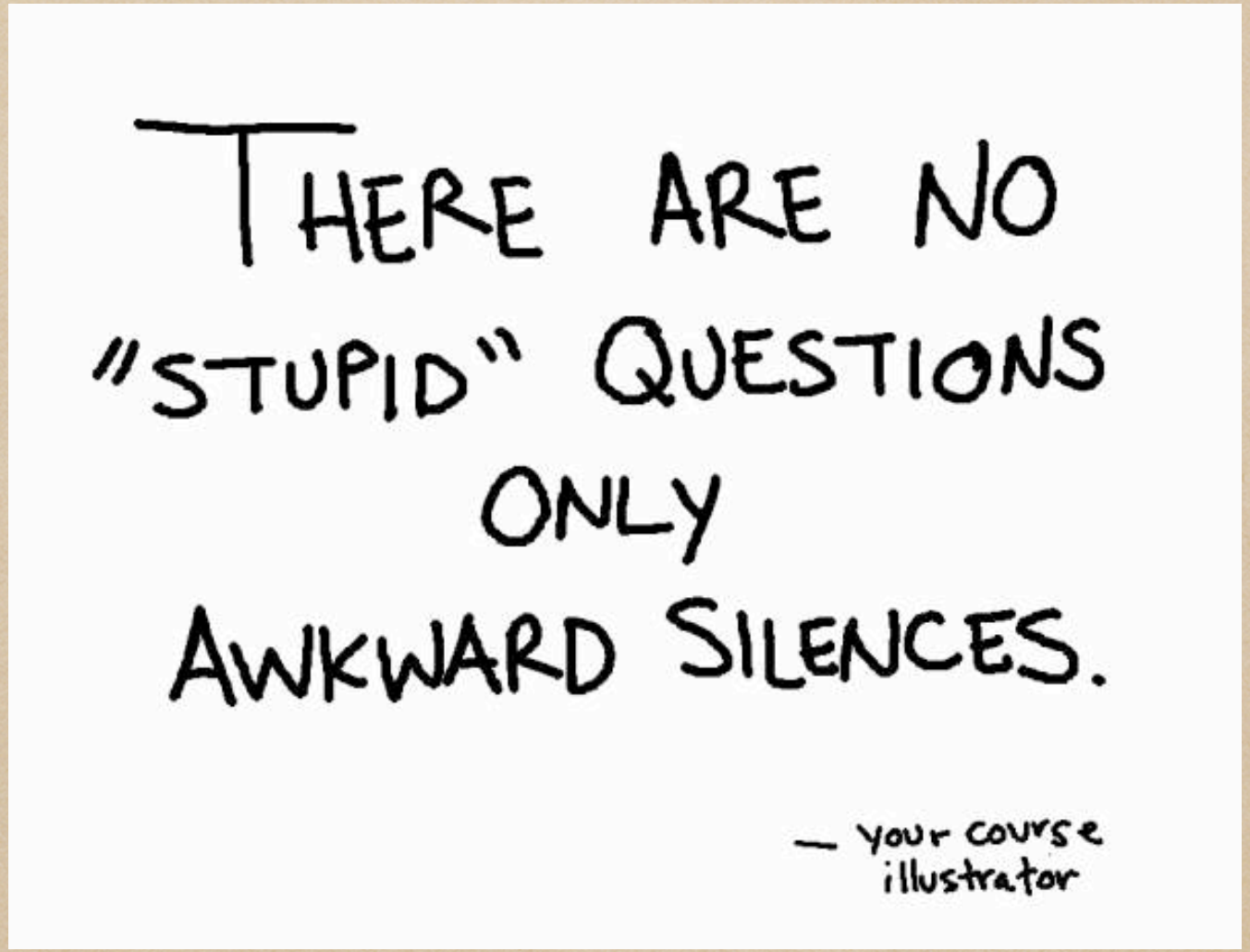
Results



[Negnevitsky 2002]

(Online) Readings

- Why GA work
- Case Studies


THERE ARE NO
"STUPID" QUESTIONS
ONLY
AWKWARD SILENCES.
— your course
illustrator

Study Questions

Draw a flowchart implementing the main steps in a genetic algorithm.

Explain crossover and mutation on both bit strings and Lisp S-expressions.

What is a schema? How does Holland's Schema Theorem explain why genetic algorithms work?

Do genetic algorithms cover the complete search space?