Stochastic Search. Part III

Genetic Algorithms

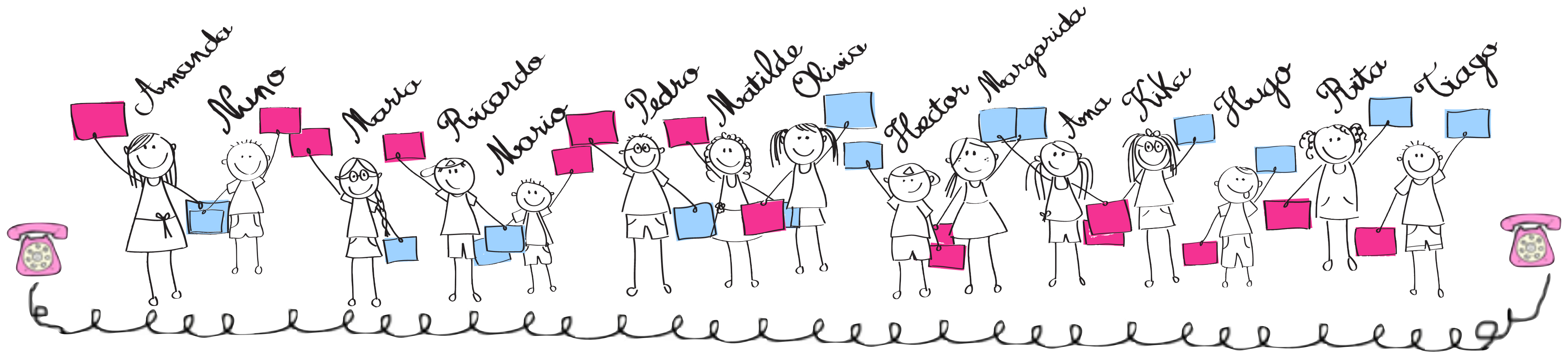
Manuel Pita | April 2020

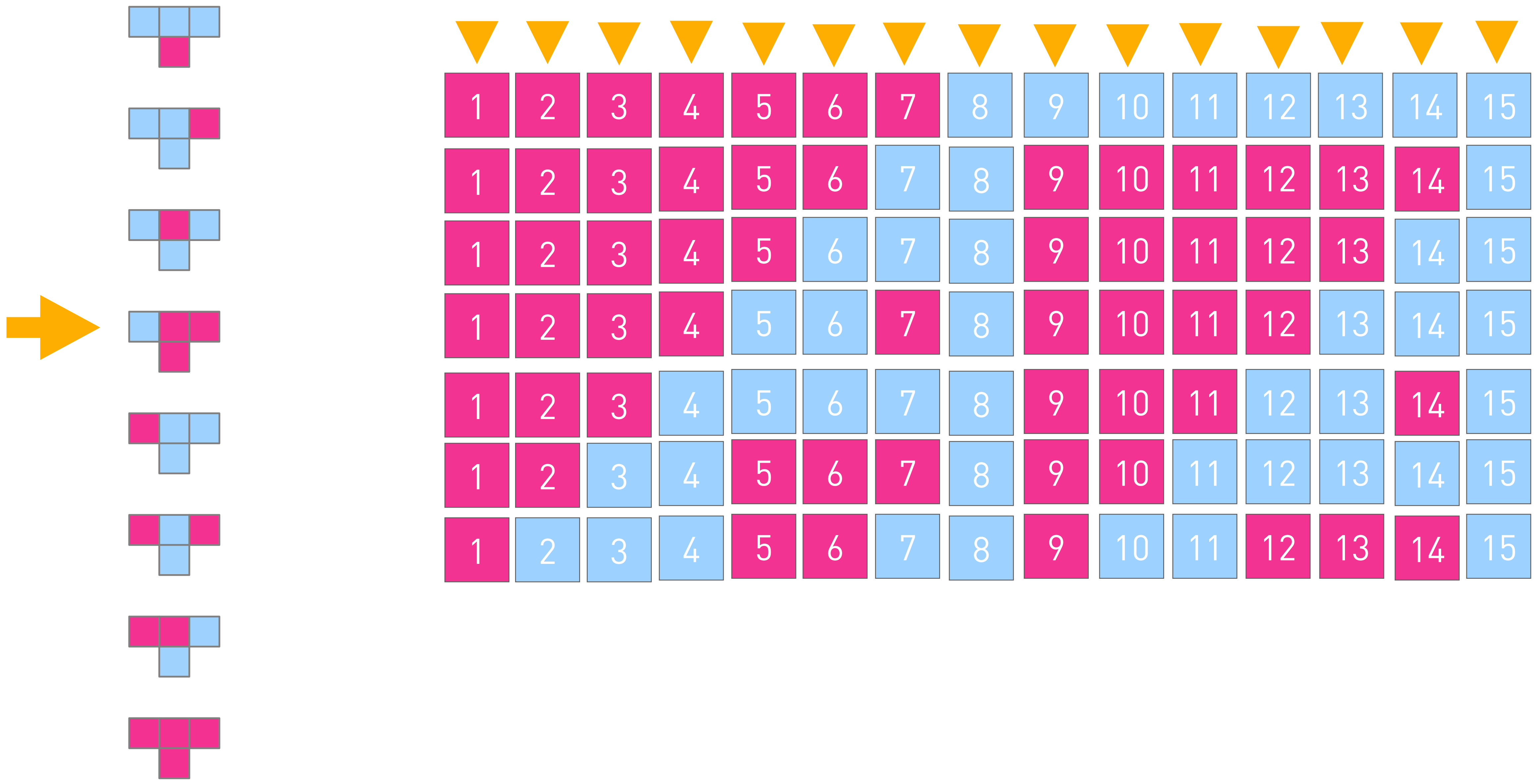
UNIVERSIDADE LUSÓFONA

Today

- Recap: cellular automata
- Recap: Collective information processing and the majority classification task
- Recap: Stochastic Search **Exploration** vs. **Intensification**
- A note about representation in Genetic Algorithms
- The Crossover and Mutation operators
- The basic Genetic Algorithm

Cellular Automata





- The card shown at **t+1** always depends on the neighbourhood at time **t**
- Time is discrete, tic, toc, tic, toc…
- Notice that over time cells can communicate (space-time computation)
- CAs converge to repeating patterns
- Python implementation is easy, but we have to be careful with the boundaries: start working on it, because this will be part of your first assessed project.

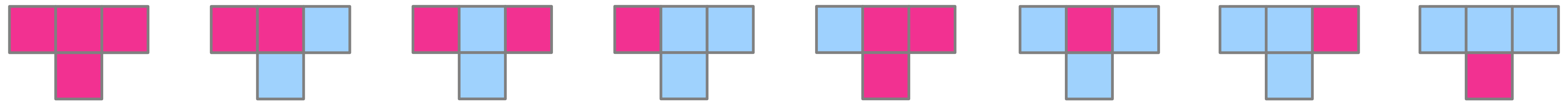
CA Rules

Lets look at cellular automata rules more closely

Lets assume pink:0 and blue: 1

Lexicographical order of binary numbers, from zero to seven

000 001 010 011 100 101 110 111



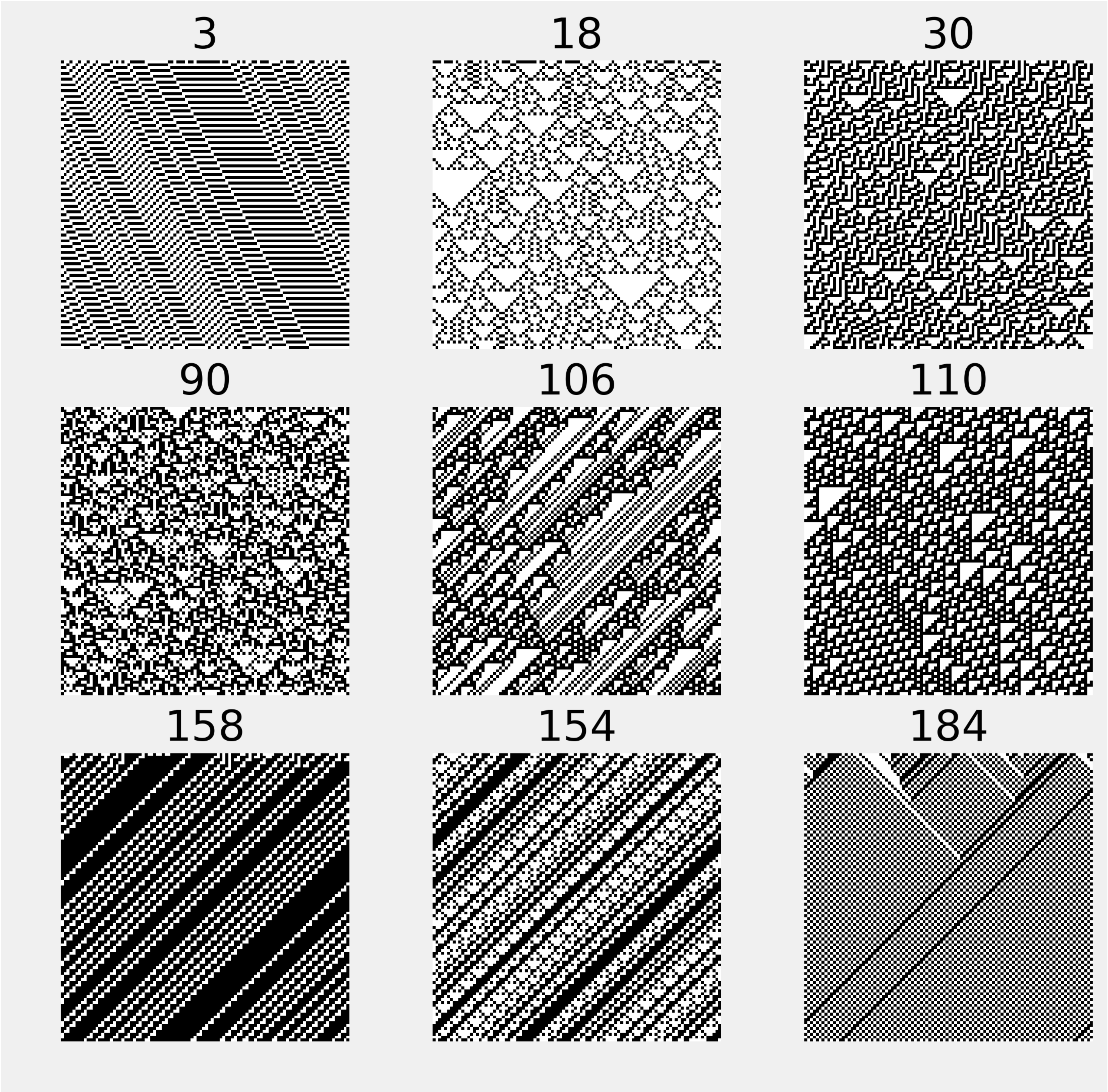
0 1 1 1 0 1 1 0

We can pack our CA rule in a 8-digit binary number

Any combination of 8 binaries is a valid r=1 rule
This specific binary is 110 in decimal, so this rule is
called the Elementary rule 110

CA Rules

The patterns of other elementary rules



CA Rules

Lets look at cellular automata rules more closely

r=1
n = 2r+1 = 3

Rule size? 2^3 = 8
How many different rules? 2^8 = 256

r=2
n = 2r+1 = 5

Rule size? 2^5 = 32
How many different rules? 2^32 (!)

r=3
n = 2r+1 = 7

Rule size? 2^7 = 128
How many different rules? 2^128 (!!)

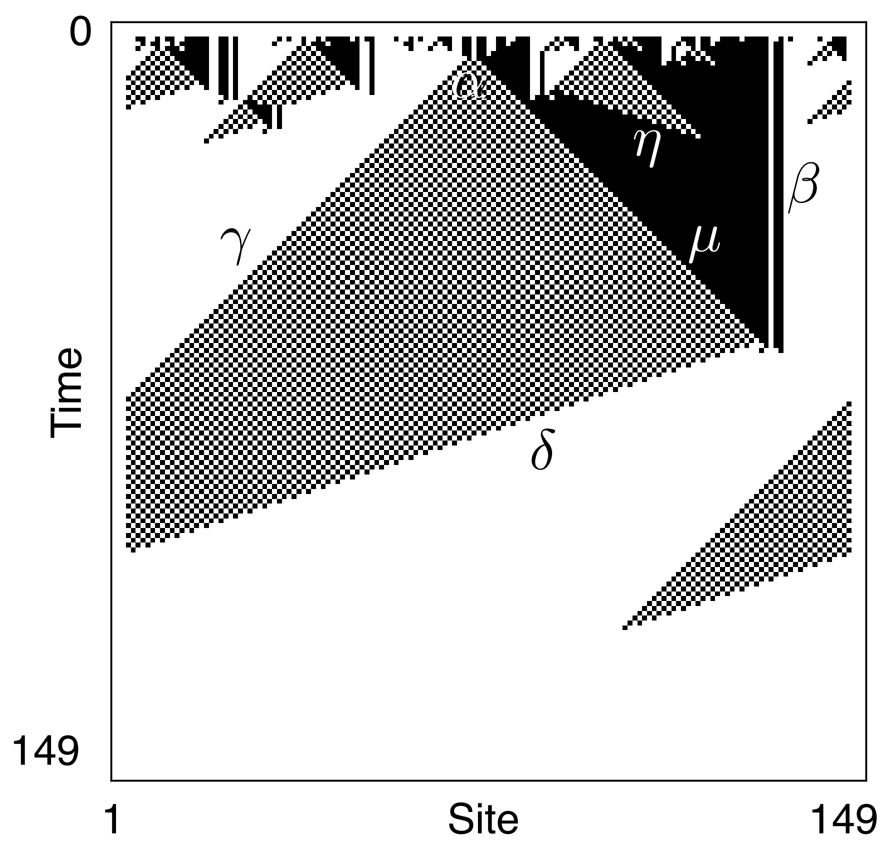
The Majority Classification Task

From CA rules, to collective information processing

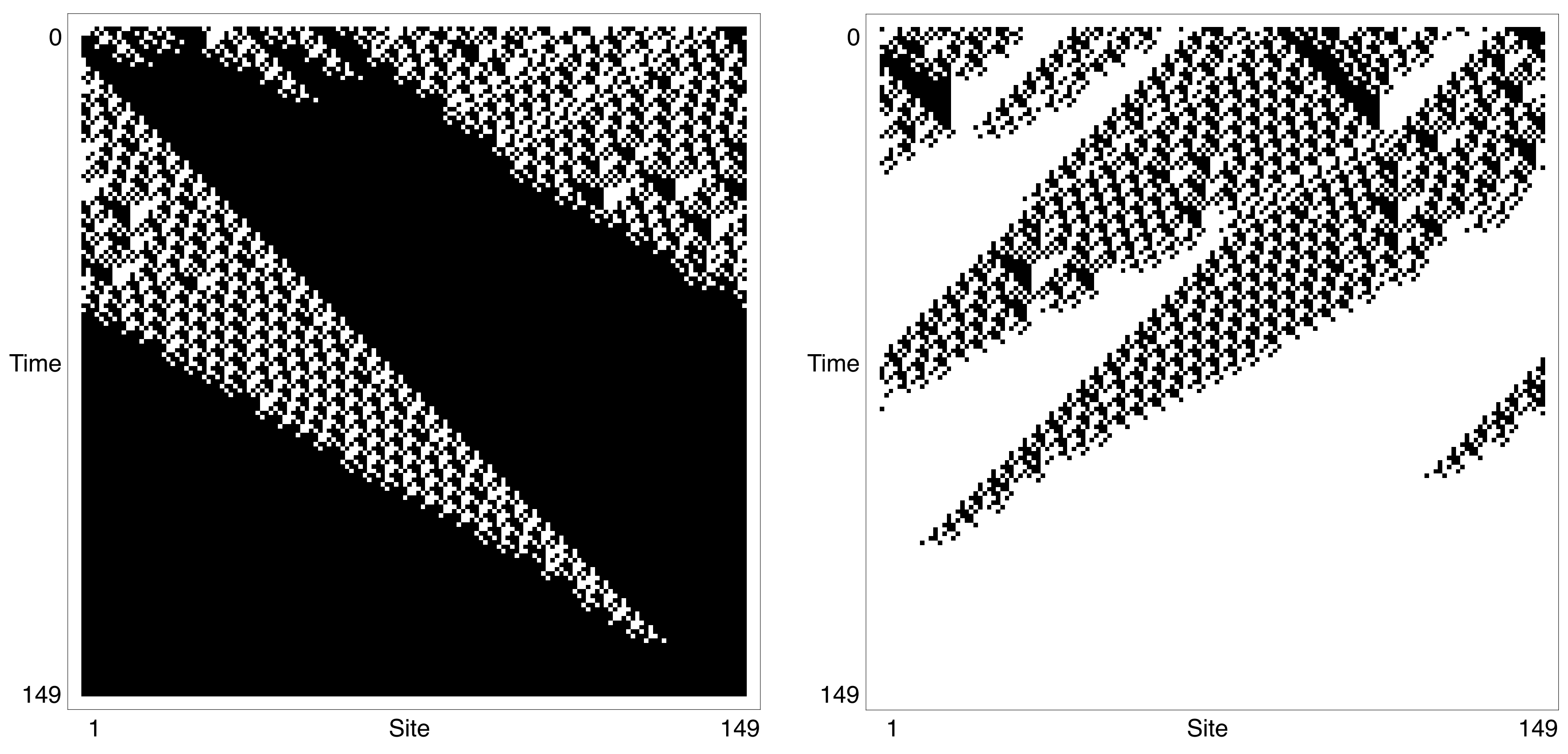
- We have an odd number of players
- Whatever random selection of cards there is always one majority colour
- Can we figure out a rule that makes all players show the majority colour?
- Why is this *complex*?
- What do we know about this special game?
    - The rule needs to consider three (r=3) neighbours on either side
    - There is no perfect solution
- What is the number of possible rules we have in our search space? Think!!

```
r = 3 -> n = 7
Rule size = 2^7 = 128
Possible rules = 2^128 (!)
```

GKL rule



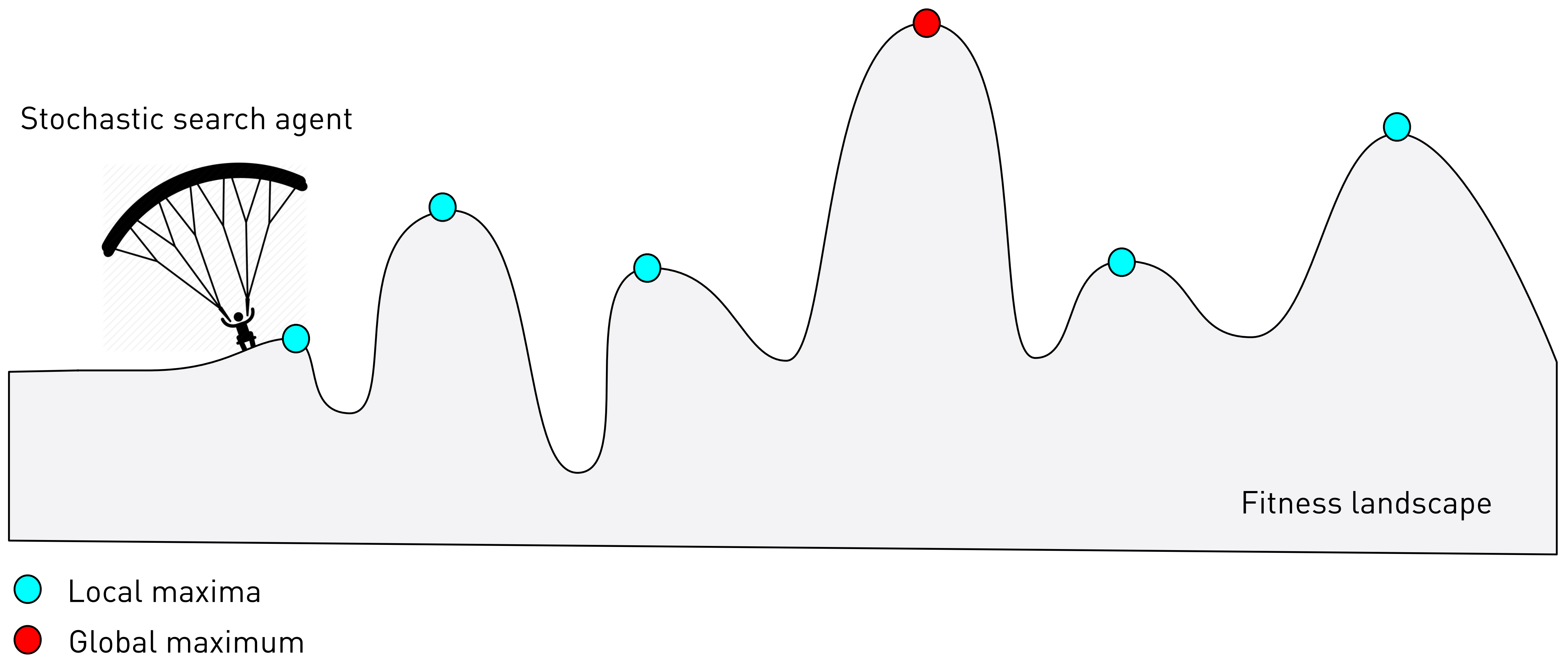
D'Oliveira rule

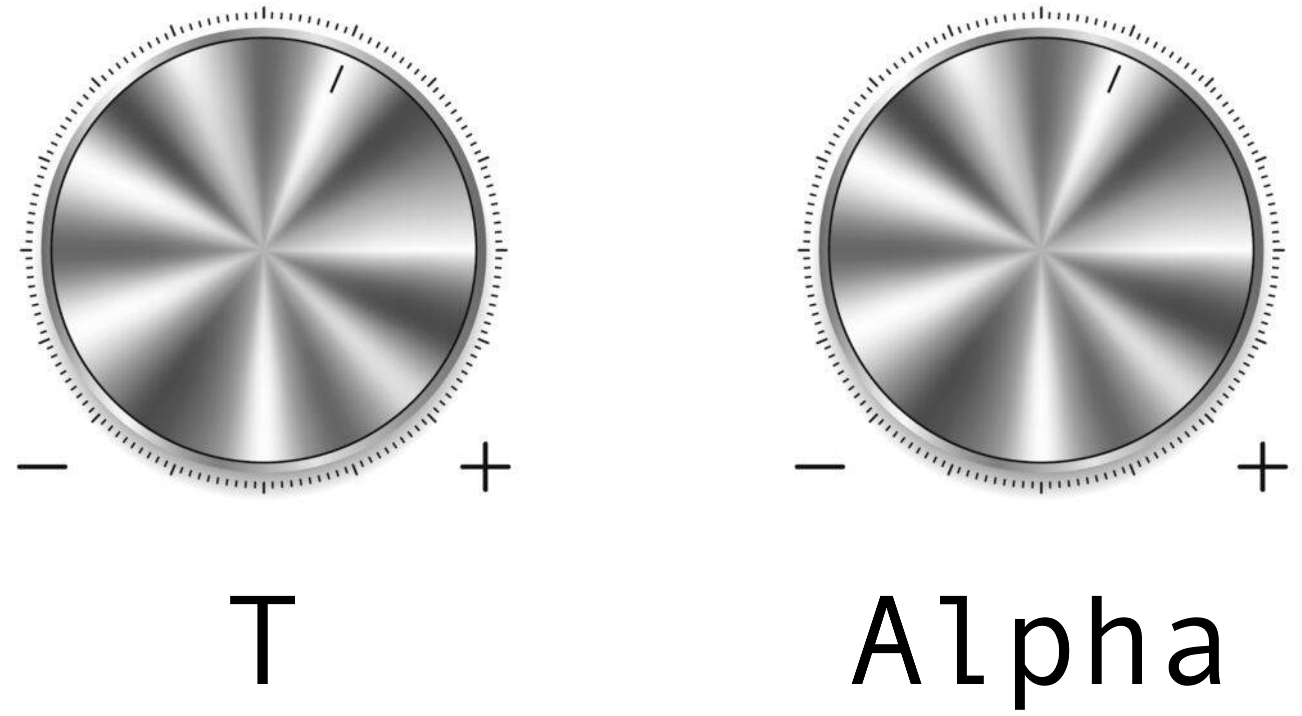


Stochastic Search

Controlling exploration and intensification

- Recall our parachuting search agent





Higher T: more exploration

Higher alpha : more intensification

Do poll if available

Side Story

Does language determine how we think about time?

- Remember how the bees last week used the sun as a base reference for communication?
- Some humans do something similar, but not us.
- Let's watch the video.

https://www.youtube.com/watch?v=mUjHE0feqFc

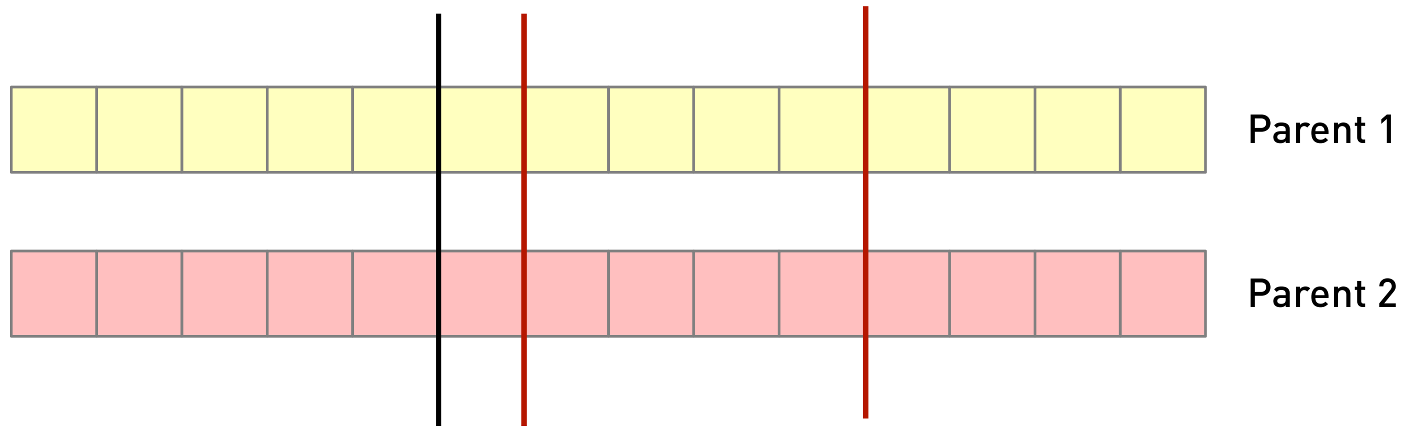
Genotypes and Phenotypes

- What do we refer to as *genotype*?
  - Arrays of discrete features (like *genes* in DNA)
  - Genes have a discrete set of *alleles*
- What are *phenotypes* in this context?
  - They are what we measure in our fitness function

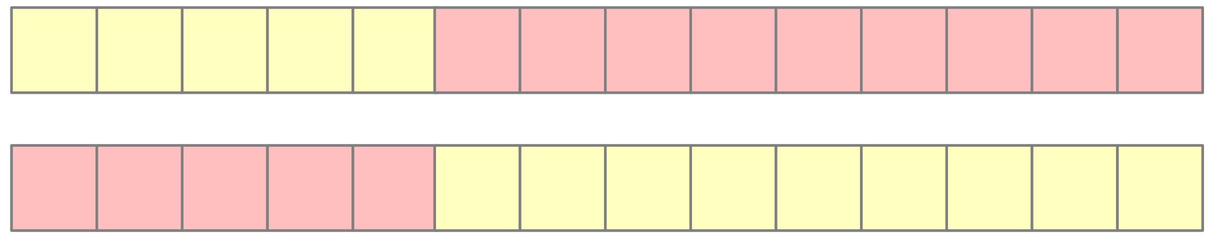
```
f(x): 2*x1 + 3*x2 + x3/2 + x4/3 = 10

[x1,x2,x3,x4]

Fitness = 1 - Abs(f(x) - 10)/10

What about in the majority classification task?

CA rules of size 128

Fitness = proportion of cells in state_majority after M iterations
```

In a broad sense, the term "genotype" refers to the genetic makeup of an organism; in other words, it describes an organism's complete set of genes. In a more narrow sense, the term can be used to refer to the alleles, or variant forms of a gene, that are carried by an organism. Humans are diploid organisms, which means that they have two alleles at each genetic position, or locus, with one allele inherited from each parent. Each pair of alleles represents the genotype of a specific gene. For example, in sweet pea plants, the gene for flower color has two alleles. One allele codes for purple flowers and is represented by the uppercase letter F, whereas the second codes for white flowers and is represented by the lowercase letter f. A varied population of sweet pea plants could therefore feature three possible genotypes at this locus: FF, Ff, or ff. Each plant's genotype contributes to its phenotype, which, in this case, is the outward appearance of its flowers. A particular genotype is described as homozygous if it features two identical alleles and as heterozygous if the two alleles differ. The process of determining a genotype is called genotyping.

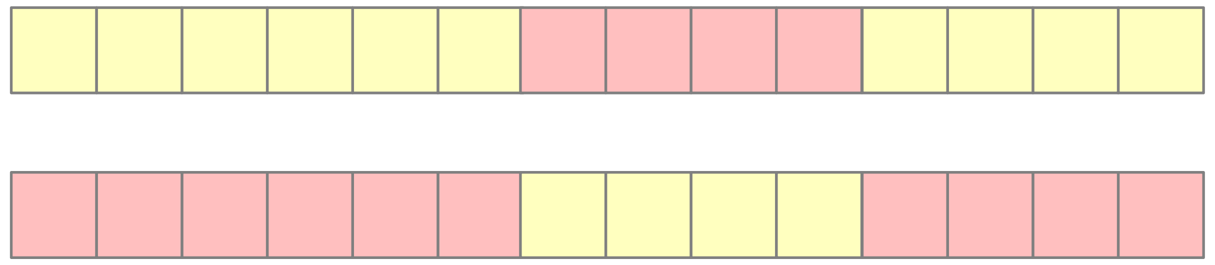
Crossover



Single point



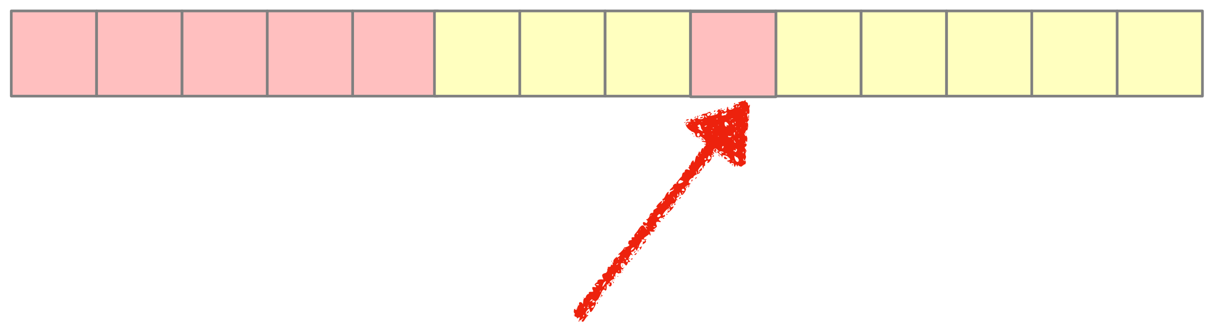
Two point



Mutation

- For each child
  - For each gene
    - Get a random number *d* between 0 and 1
    - If d < mutation rate (m) mutate, else leave as is

`m = 0.02`



`d = 0.01`

Basic Genetic Algorithm

1. Generate a random population of solutions of size P
2. Compute fitness of each individual and sort population from highest to lowest fitness
3. Select the top E individuals and transfer them without change to the next generation
4. Create the remaining P-E members as children of Elite members
    4.1 choose random pair of parents
    4.2 generate children with crossover using n crossover points
    4.3 subject each children to mutation rate m
5. Repeat until termination condition

How to terminate a Genetic Algorithm

- After a number of generations in a single *epoch*
- After achieving target fitness in single *epoch*
- The concept of multi *epoch* runs

Parameters you need to control

- Population size (P)
- Elite size (E)
- Mutation rate (m)
- Number of cross-over points