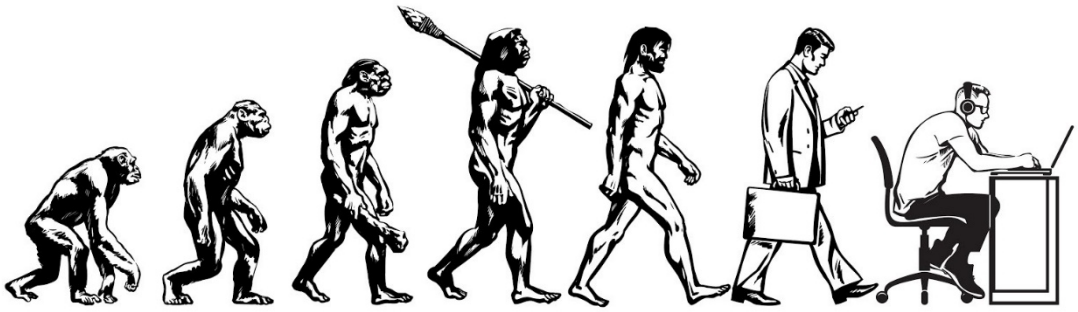
7: EVOLUTIONARY ALGORITHMS

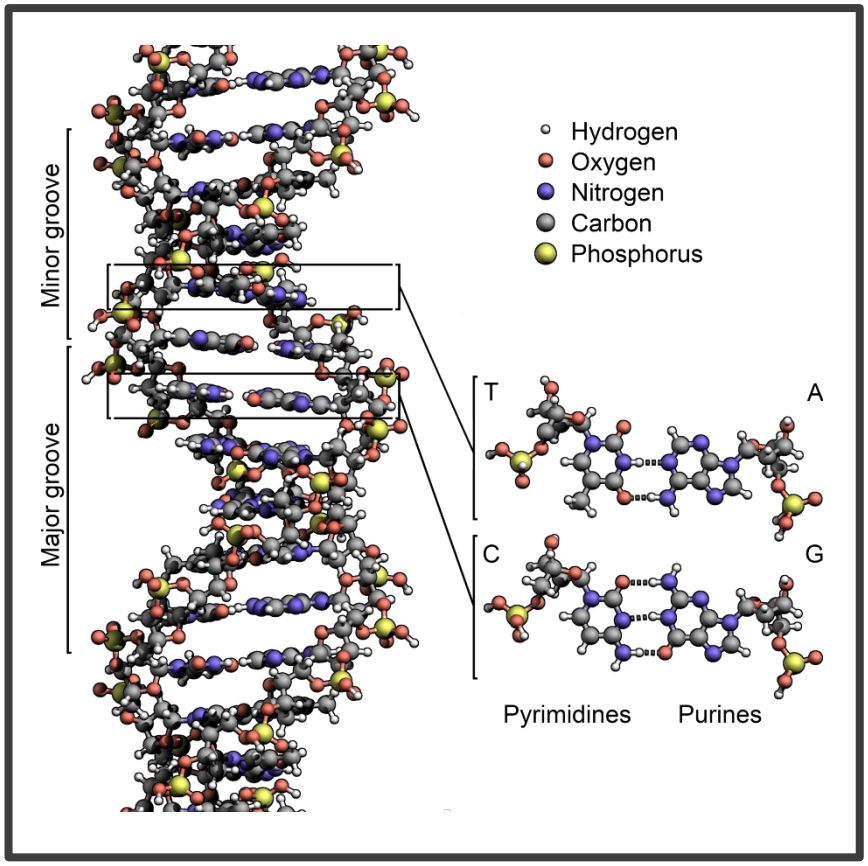
COMP704: MACHINE LEARNING

BIOLOGICAL EVOLUTION

- Biological populations survive by **reproduction**
- When organisms reproduce, they copy themselves **imperfectly**
- Offspring take on **traits** from their **parents**
- If offspring **survive** long enough to reproduce, their traits are **passed on**
- "**Survival of the fittest**" – traits which are valuable for survival tend to propagate



GENETICS FOR DUMMIES



- **DNA**: deoxyribonucleic acid
- A molecular string of **base pairs**: adenine-thymine, cytosine-guanine
- Essentially a **string** of characters A C G T
- Grouped into **genes**, which in turn are grouped into **chromosomes**

GENETICS FOR DUMMIES

- DNA constructs **proteins** through complex chemical reactions
- Proteins construct **organisms**
- DNA is **copied** when cells and organisms reproduce
- "Mistakes" sometimes happen during copying: this is called **mutation**
- In sexual reproduction, offspring take **some genes from each parent**

GENOTYPES AND PHENOTYPES

- **Genotype**: DNA
- **Phenotype**: organism
- Genotype dictates the form the phenotype takes
- Changes in the genotype cause changes in the phenotype
- This is a **one-way mapping**: changes in the phenotype do not change the genotype
  - Discredited theory of Lamarckian evolution
- Genotype is **information** (the sequence of base pairs in the genes)
- Phenotype is a **physical organism**

FITNESS

- **Fitness** is a property of the **phenotype**
- In biology: how capable the organism is of surviving and reproducing
- This directly determines how likely its genes are to survive by being passed on
- Fitness is **endogenous**: a by-product of the environment

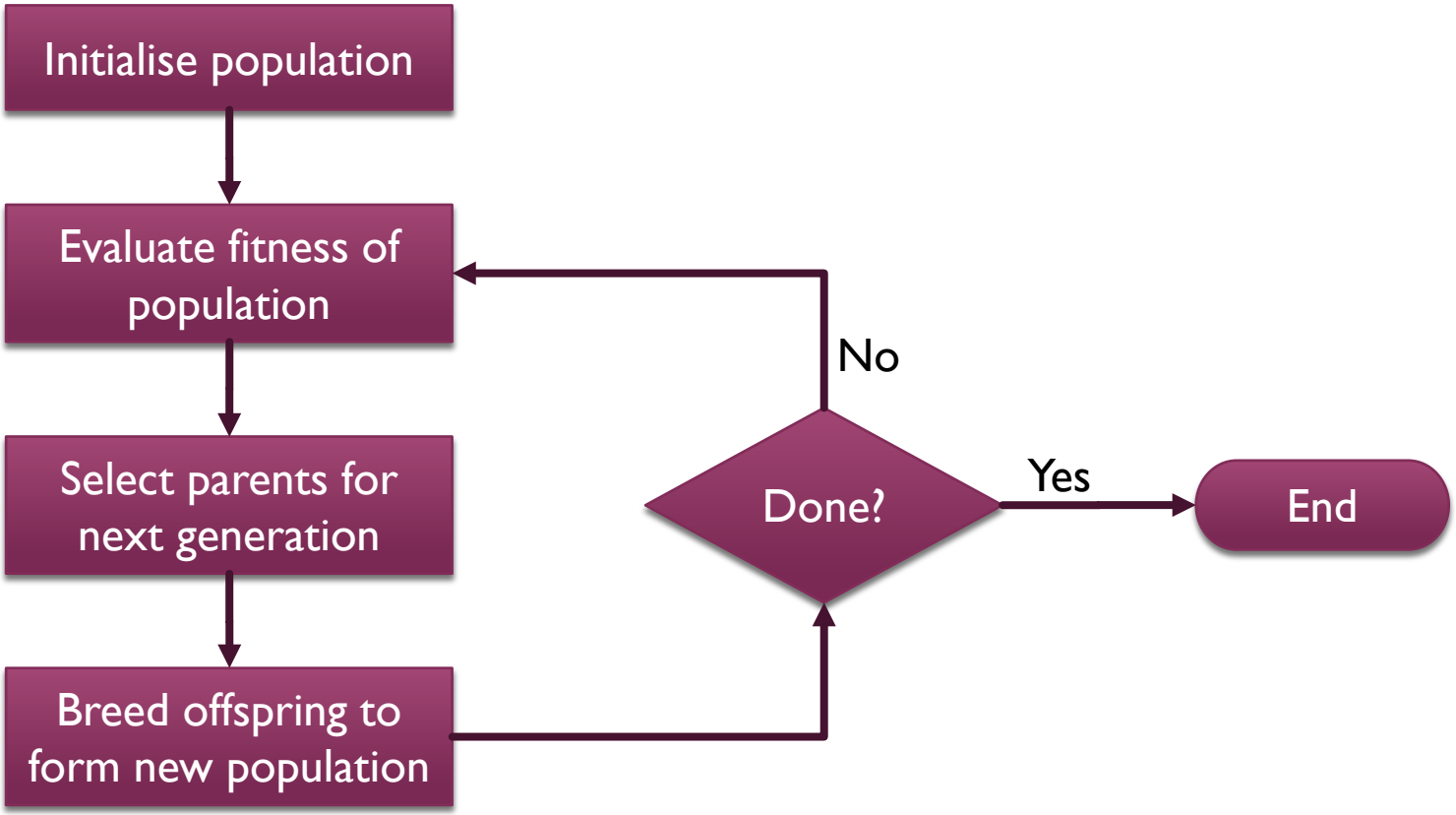
EVOLUTIONARY ALGORITHMS (EAS)

- **Inspired by** biological evolution
- A family of **population-based search** algorithms
- **Genotype**: search space representation
- **Phenotype**: solution space representation
- E.g. genotype is parameters to a content generator, phenotype is the generated content itself

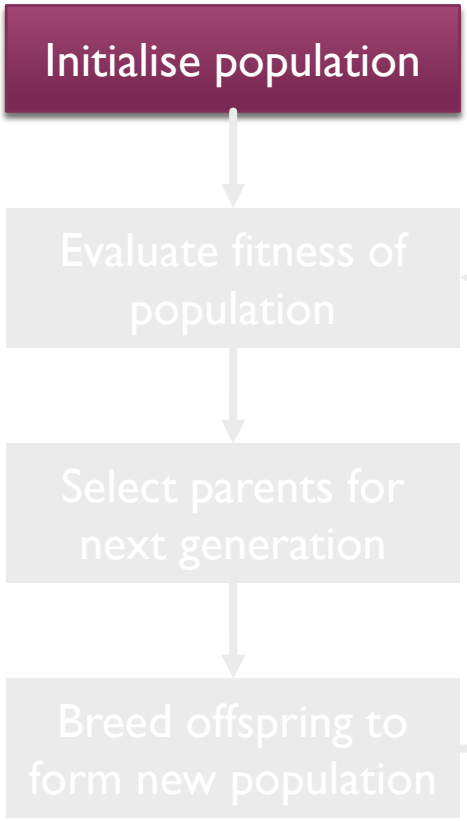
FITNESS

- Fitness is still a property of the phenotype
- However in EAs, fitness is usually **exogenous** rather than endogenous
- Fitness is not a by-product of an environment, but is an explicit **numerical function** provided by the developer

BASIC EVOLUTIONARY ALGORITHM



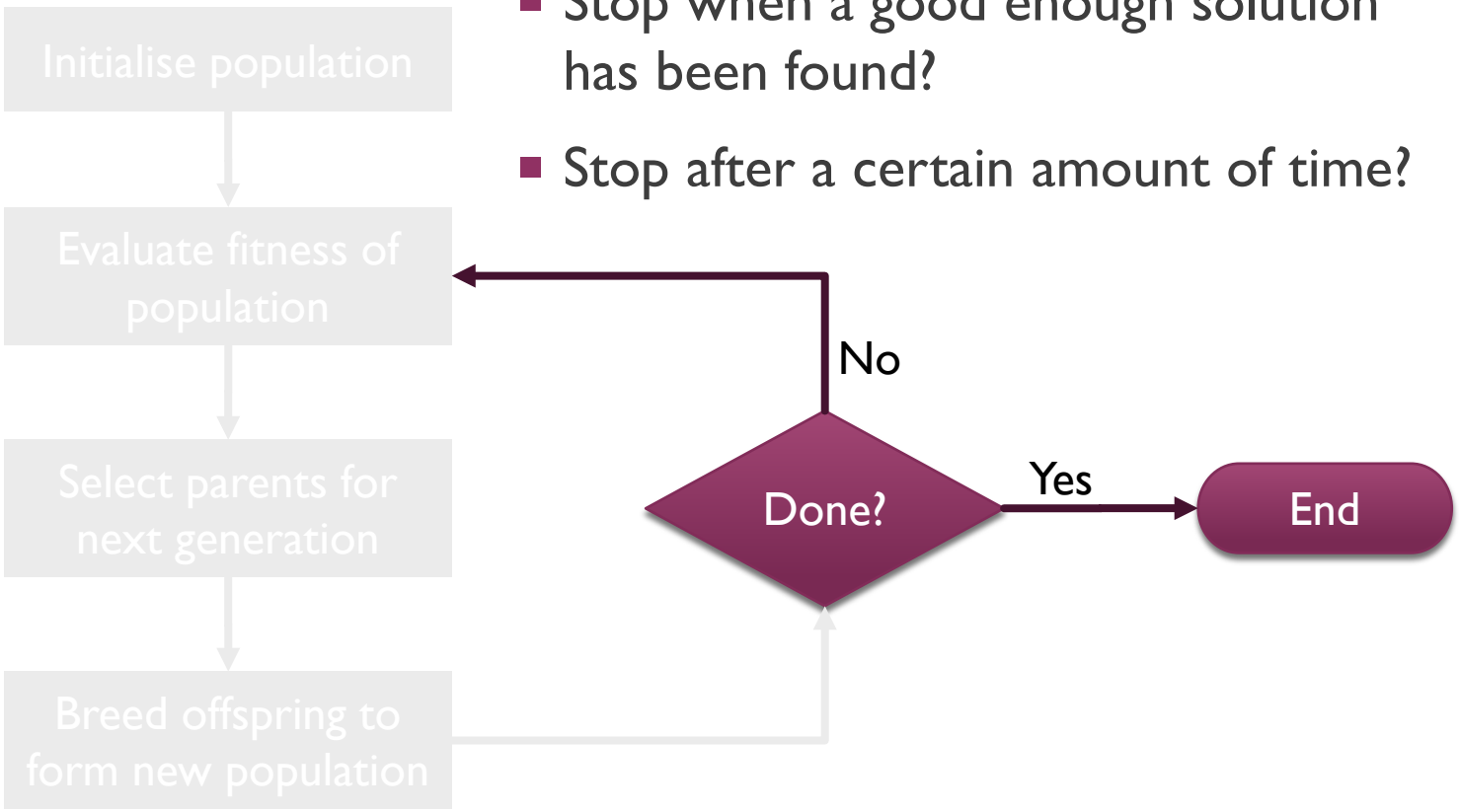
INITIALISATION



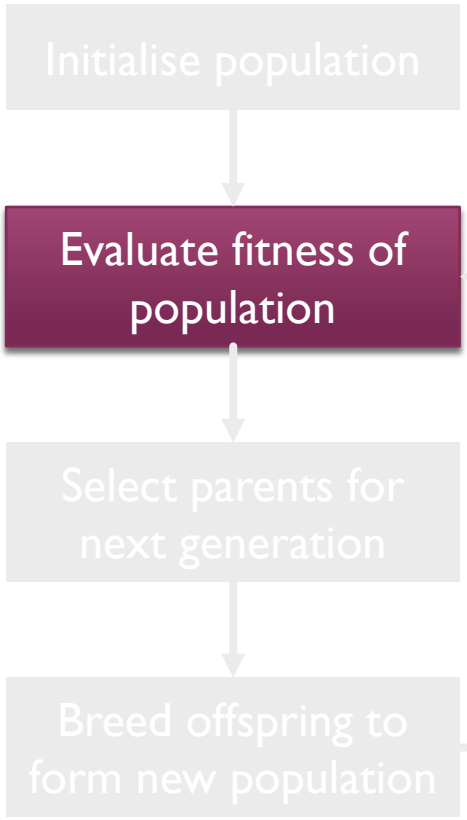
- Often initialise the population with **random** individuals
- May also use **known good** individuals (though this can **bias** the search towards those solutions too much)

STOPPING

- Stop when a good enough solution has been found?
- Stop after a certain amount of time?

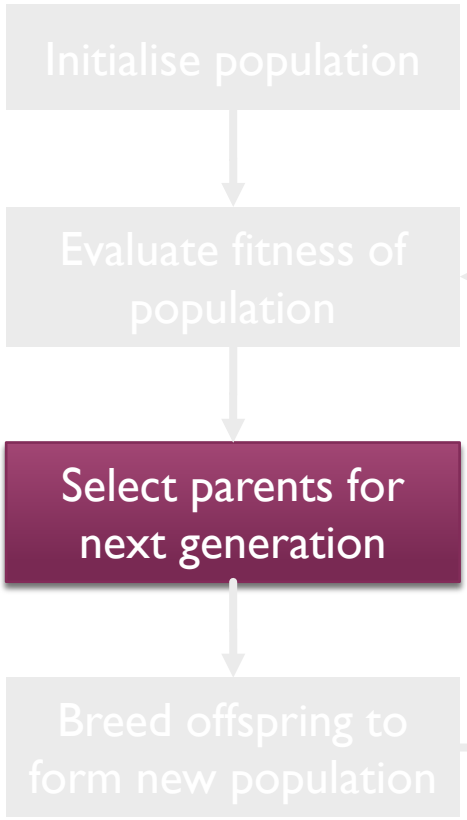


FITNESS



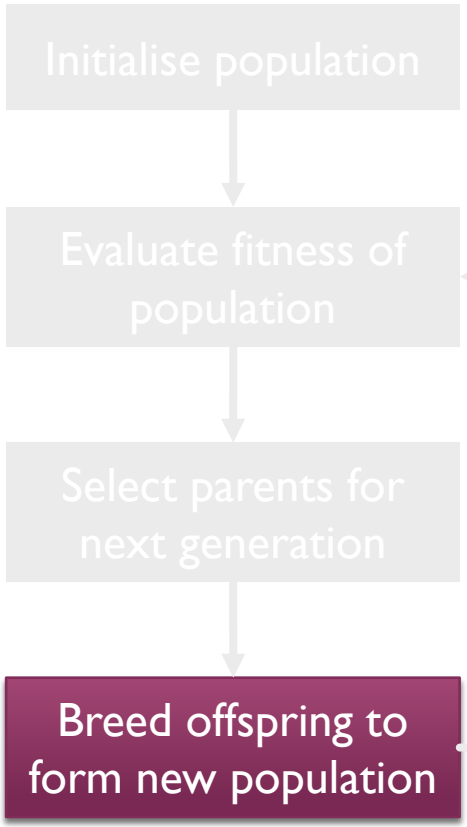
- Need a **fitness function** to evaluate individuals
- Fitness is a feature of the phenotype – probably have to map genotype to phenotype here

SELECTION



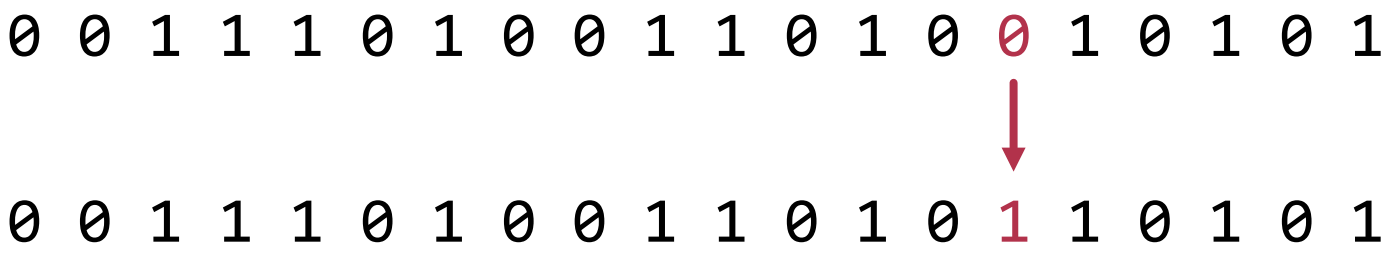
- **Deterministic**: choose the top X% by fitness score
- **Probabilistic**: choose with probability proportional to fitness
- **Tournament**: choose n at random and take the fittest
- **Artist as gardener**: skip the fitness evaluation and ask a human user to choose

REPRODUCTION



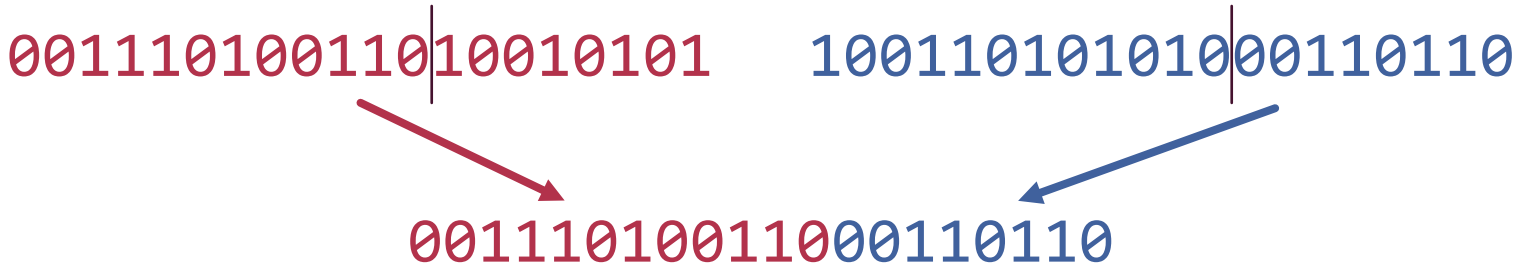
- **Mutation**: take one parent and change it slightly
- **Crossover**: take two parents and combine them
- **Elitism**: allow the fittest individuals in the current generation straight through to the next

MUTATION



- One parent
- Make a **small random change** to the genotype
- E.g. if genotype is a binary string, flip a bit
- (Or flip X% of bits, or flip each bit with probability P)

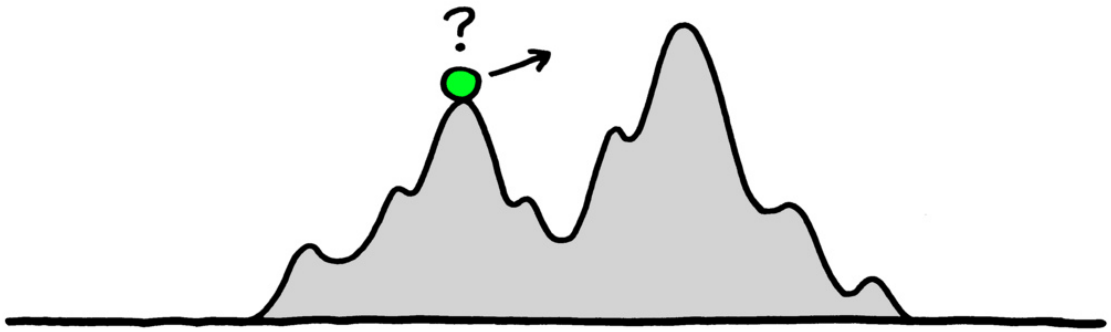
CROSSOVER



- Two parents
- **Split** the genotype at a random **crossover point**
- **Combine** part of the first parent and part of the second parent

WHY BOTH?

- **Mutation** helps avoid the problem of getting stuck in a **local optimum** – ensures **diversity**
- **Crossover** works to recombine "useful" parts of genotype into fitter individuals – "**building block hypothesis**"



WORKSHOP TOMORROW

- Assignment support!