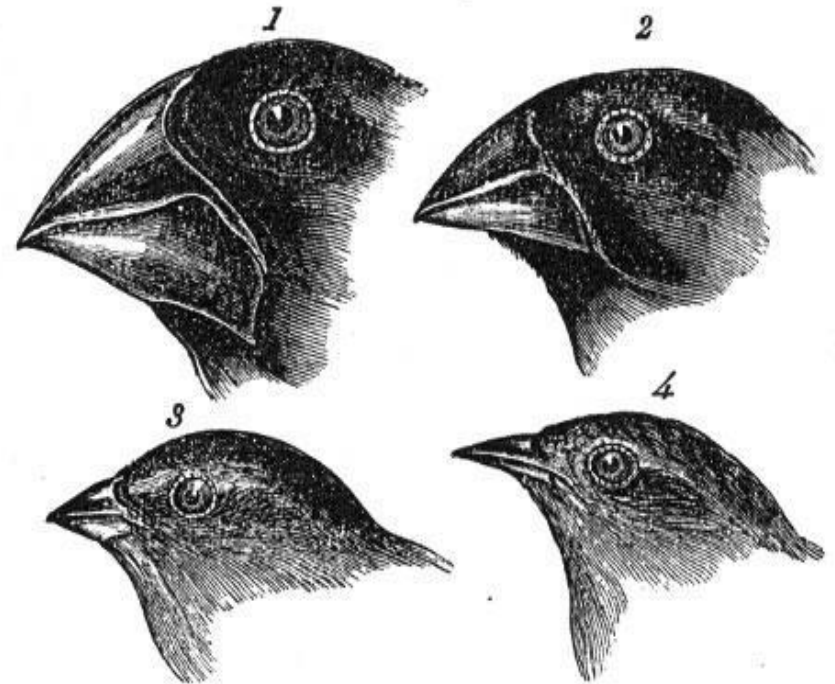# SELECTION IN E.A.
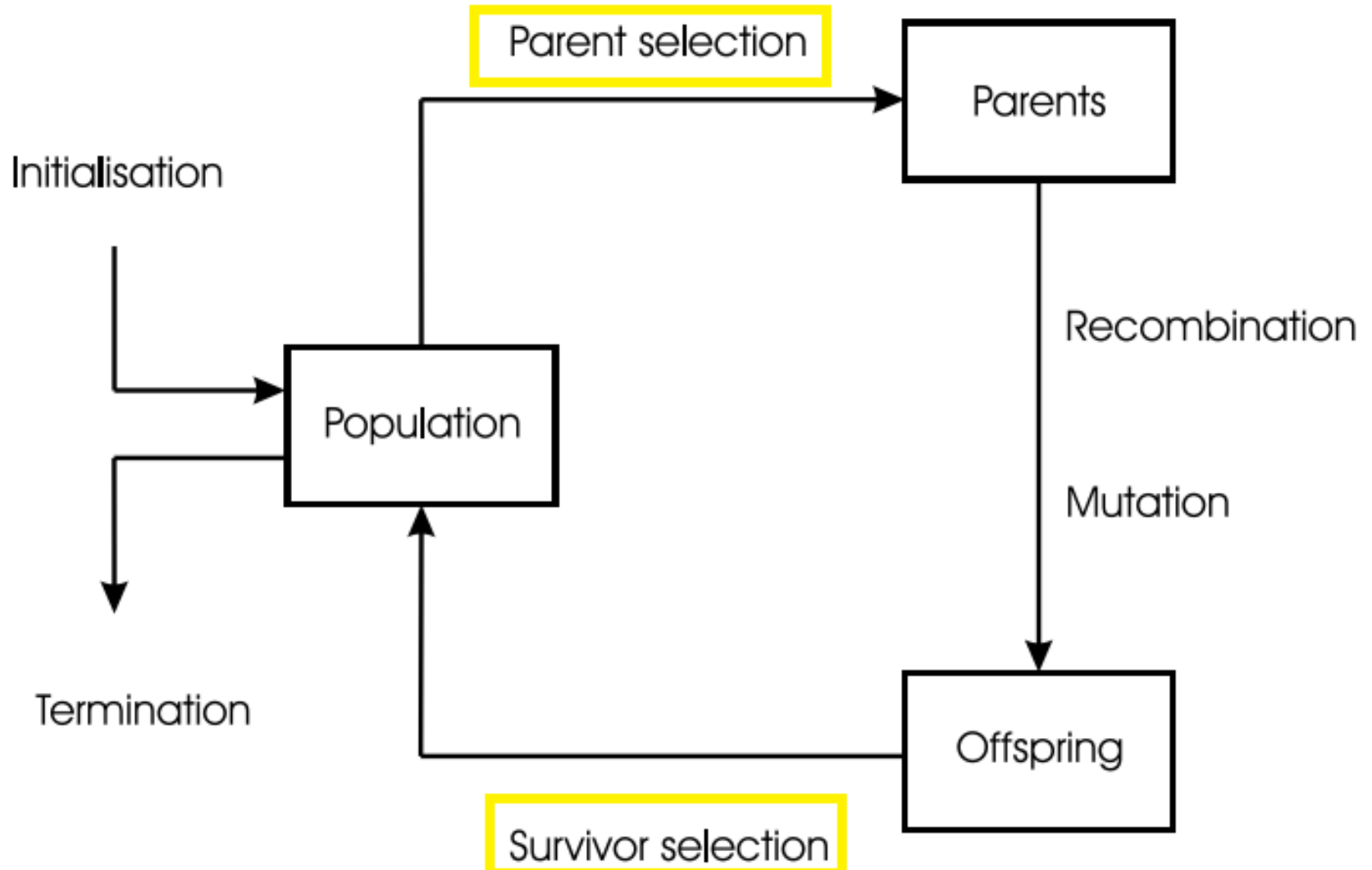
Riyad Parvez

1. Geospiza magnirostris.
2. Geospiza fortis.
3. Geospiza parvula.
4. Certhidea olivasca.

Selections in E.A.

Parent Selection

- An individual is parent if it has been selected to undergo variation in order to create offspring.
- High quality individuals get a higher chance to become parents than those with low quality.
- Low quality individuals are given small positive chance.
- Otherwise, the whole search could become too greedy and stuck in local optima.
- It's usually stochastic.
- With survivor selection mechanism, parent selection is responsible for pushing quality improvement.

Steady State Selection

- In every generation a few high fitness chromosomes for are selected creating a new offspring.
- Low fitness chromosomes are removed and the new offspring is placed in their place.
- The rest of population survives to new generation.

Selection Climbing

- To increase the average fitness of the population, the strength of selective pressure also increases and the fitness function becomes more discriminating.
- When all individuals have a relatively high fitness and only small differences distinguish them in fitness.
- More exertion of selection pressure rapidly eliminates individuals with lower fitness value.
- Useful when immediate solution is necessary.

Tournament Selection

- A group of n individuals is chosen randomly uniformly from the current population, and the one with the best fitness is selected.
- *n* is called the tournament size and can be used to vary the selection pressure.
- The higher n the higher the pressure to select above average quality individuals.
- Several tournaments are held between the individuals of current generation.
- Process is repeated as often as desired (usually until the mating pool is filled).

Tournament Selection

- Tournaments held between pairs are binary tournament selection.
- Deterministic tournament selection selects the best individual (when p=1) in any tournament.
- 1-way tournament (k=1) selection is equivalent to random selection.
- The takeover time for tournament selection is logarithmic to the population size.
- Most all the niches in a population will tend to disappear relatively rapidly.
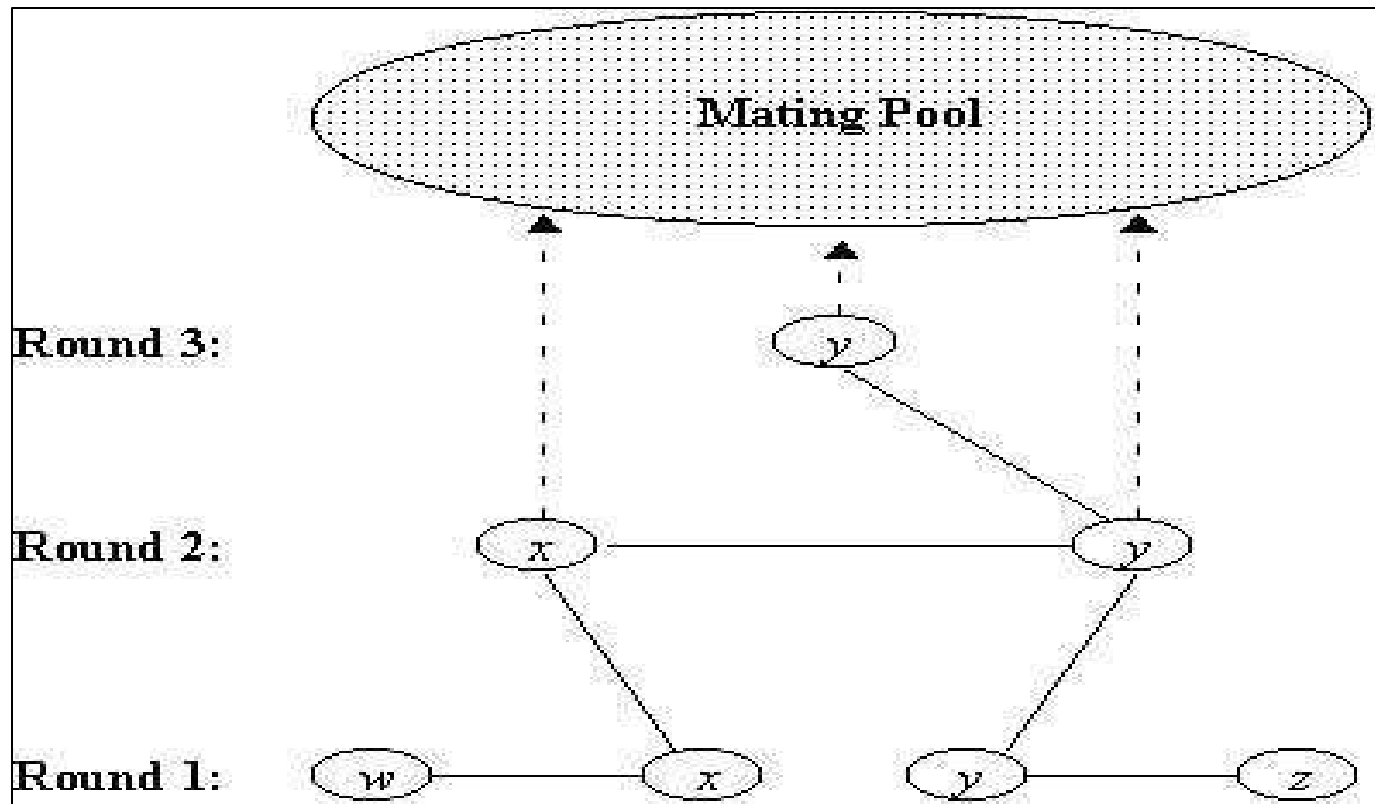
Tournament Selection

Fig: Tournament Selection (Knock-out)

Tournament Selection Variants

- Fixed Tournament Size
- Self-adaptive Tournament Size
- Fuzzy Tournament Selection
- Deterministic Tournament Size Control
- Boltzmann Selection with Annealing

Restricted Tournament Selection

- Needed when we want best n solutions for multimodal problems.
- Unlike TS, RTS preserves the niches.
- It prevents members of one niche competing with members of other niche.
- Strength of RTS is ability to foster different species for finding multiple peaks simultaneously regardless of the inter-peak distances.
- Performs worse when the number of peaks is too large compared to population size because it cannot recover diversity once converged.

Roulette Wheel Selection

- Also known as fitness proportionate selection.
- Every individual is allocated a section in the roulette wheel.
- Sections are different sizes, proportional to the individual's fitness.
- The fittest candidate has the biggest slice of the wheel and the weakest candidate has the smallest.
- The wheel is then spun and the individual associated with the winning section is selected.
- Wheel is spun as many times as is necessary to select the full set of parents for the next generation.
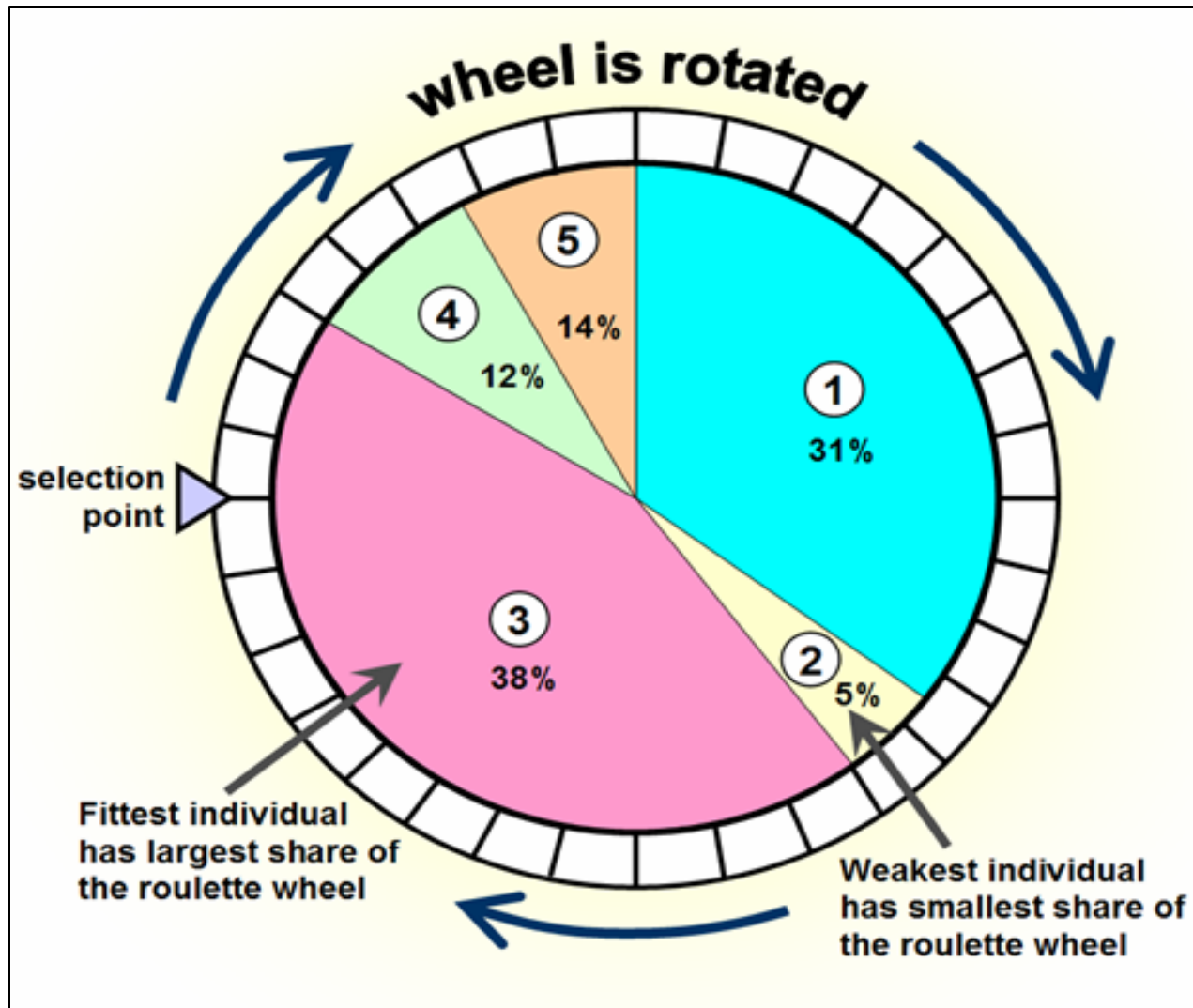
Roulette Wheel Selection
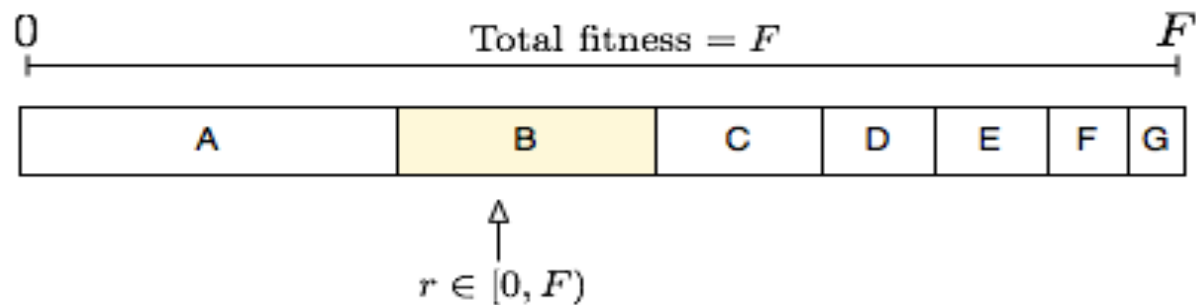
Roulette Wheel Selection

- Individuals with better fitness have higher probability of being selected than the individuals with lower fitness value.
- Weaker individual still get chance to breed offspring.
- Better individuals are usually selected more than one times.
- RWS doesn't guarantee number of times selection in mating pool will be proportional to it's fitness ratio.
- An individual with biggest section may not be chosen.
- Individual with smallest section may be chosen several times (very unlikely) in mating pool.

Roulette Wheel Selection

Stochastic Universal Sampling

- A variation of Roulette Wheel Selection.
- Unlike RWS, to make N selection it only takes a single spin.
- Selection process proceeds by advancing all the way around the wheel in equal sized steps.
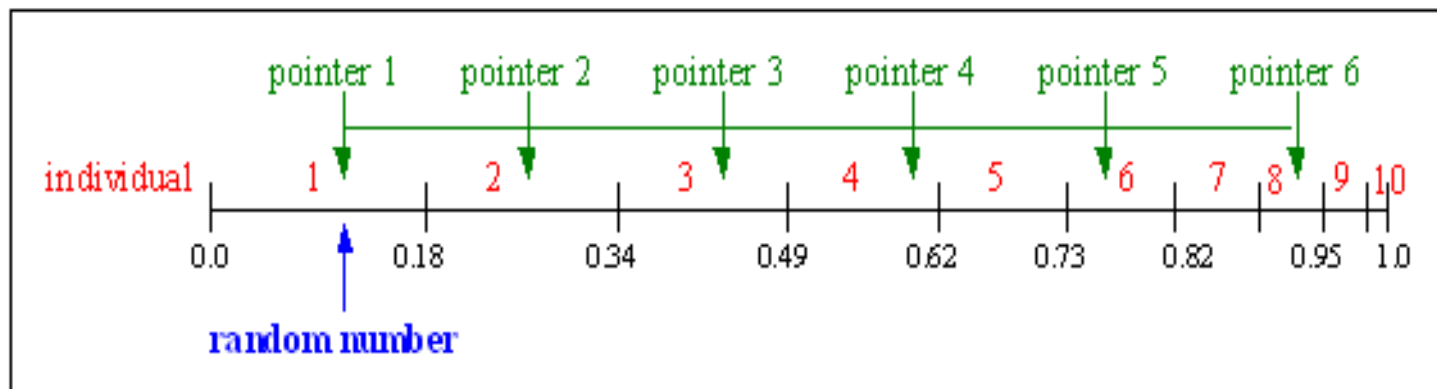
Stochastic Universal Sampling

- Step size is determined by the number of individuals to be selected.
- If an individual that occupies 4.5% of the wheel and select 100 individuals, on average for that individual to be selected between four and five times.
- The individual will be selected either four times or five times, not three times, not zero times and not 100 times.
- Unlike RWS, SUS guarantees that.

Stochastic Sampling Without Replacement

- Variation of RWS.
- For each element i, we calculate the ratio r_i of its fitness on the average fitness.
- Let $e(r_i)$ the integer part of r_i, each element is reproduced exactly $e(r_i)$ times.
- Roulette wheel selection described above is applied to the affected individuals of fitness $r_i - e(r_i)$.
- RWS is then repeated until the mating pool is full.
- Remainder stochastic sampling with replacement has a greater probability of diversity in the population than RWS.
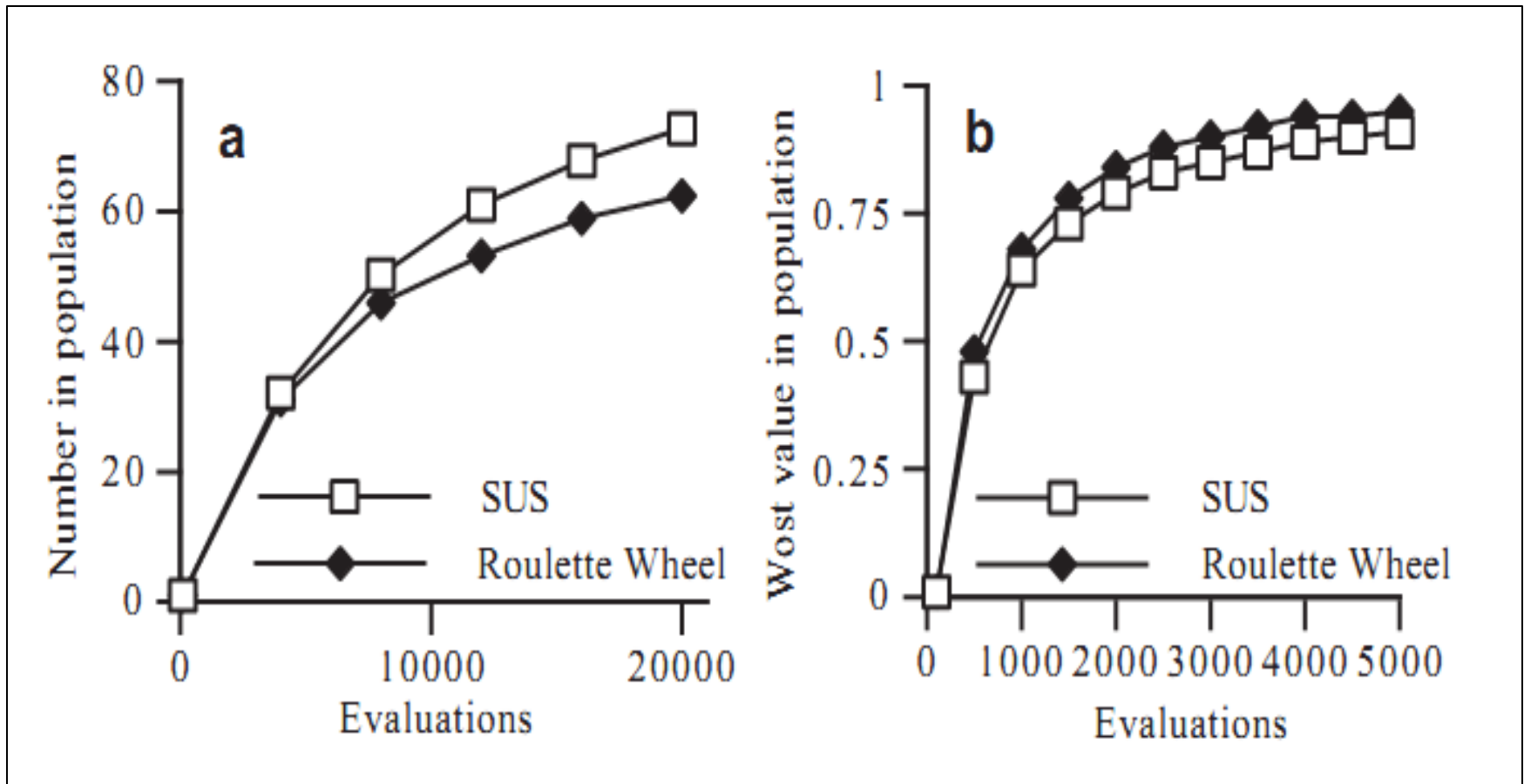
RWS vs. SUS

Fig: Takeover Time

Top Percent Selection

- X% best individuals from population are selected according to fitness value.

- Among them, individuals are selected into mating pool randomly.

Fitness Uniform Selection

- It focuses the selection pressure towards fitness levels which have relatively few individuals rather than on the highest fitness levels.
- Parts of the fitness space which are difficult to reach, are focused on, rather than easy to reach areas which are already well represented in the population.
- FUSS preserves genetic diversity more actively than the standard selection schemes which tend to drive the populations on lower fitness levels to zero.
- Greater diversity of population comes at the cost of a potential loss of performance due to the large number of selections from low fitness levels.

Fitness Uniform Selection
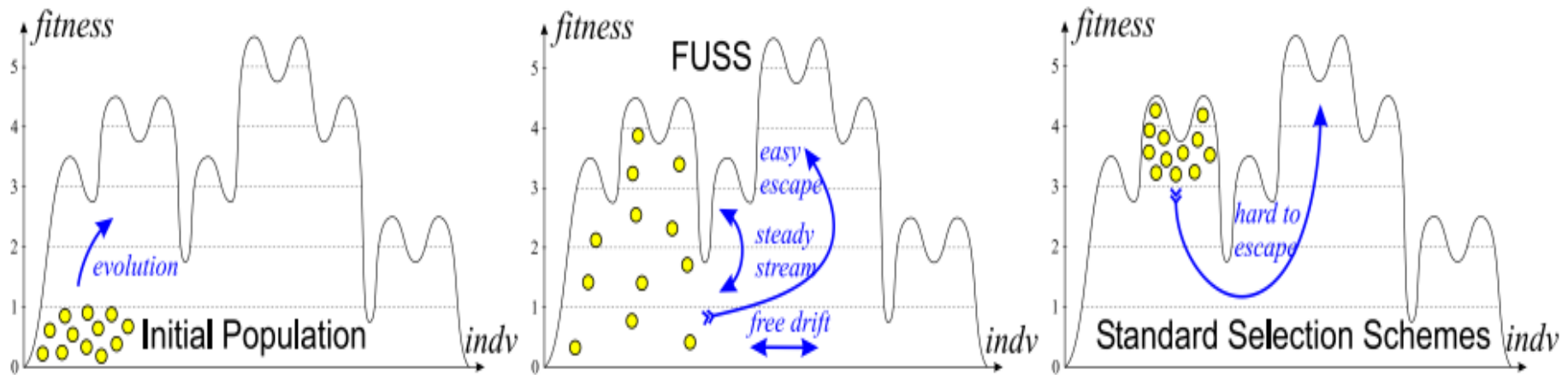
- Currently highest parts of the fitness space are now searched more slowly than under a standard selection scheme.
- When fittest solutions takeover the population, FUSS will automatically move selection pressure away from these highly fit individuals and focus its search energy on lower fitness levels that have fewer individuals.
- In this way the selection intensity varies dynamically with the evolution of the population.

Drawbacks of Fitness Function

- Leads to the appearance of few super individuals who dominates the selection process and slow it down.
- Results is premature convergence.
- If diversity still persists, all individuals (better or avg.) get nearly the same number of copies in the future generation.
- At the end, competition becomes absent, and the whole process becomes a random walk in individual solutions.
- **Answer:** Fitness Scaling.

Fitness Scaling

- Purpose is to reduce the affect of fitness outliers in the population.
- Process of scaling is intended "to maintain selective pressure as the overall performance rises within the population".
- Scaling mechanism is used to distinguish the performance of individuals when improvement among the population is minute.
- Instead of selecting parents based on raw fitness value X, we select parents based on some function of raw fitness value $f(X)$.

Fitness Scaling

- **Low Scaling Factor:** Individuals won't converge to the real optima because they are not attractive.

- **High Scaling Factor:** "Super individuals" in initial populations may dominate the population quickly. The evolution may not be able to locate all peaks.

Linear Scaling

- Linear scaling adjust the fitness values of all the strings such that the best individual gets a fixed number of expected offspring.
- Other values are altered so as to ensure that the correct total number of new strings are produced: an average individual will still expect one offspring.
- Exceptionally fit individual are prevented to takeover quickly.
- A linear relation between the scaled fitness, F and the raw fitness, f, something of the form: $F = af + b$, where, a and b are scaling coefficients.

Sigma Scaling

- Helps to overcome a potential problem with particularly poor individuals which with windowing keep baseline very low.
- Sigma scaling keeps the baseline near average.
- Setting the baseline s standard deviation (sd) below the mean where s is the scaling factor
- Individuals below this score assigned fitness 0.

Windowing

- Windowing introduces a moving baseline.
- Worst value observed in the w most generations is subtracted from all the fitness value.
- FPS initially converges rapidly, but slows down on the verge of convergence.
- It maintains much more constant selection pressure, stronger for small window size.

FPS, Linear Scaling, Sigma Scaling, Windowing
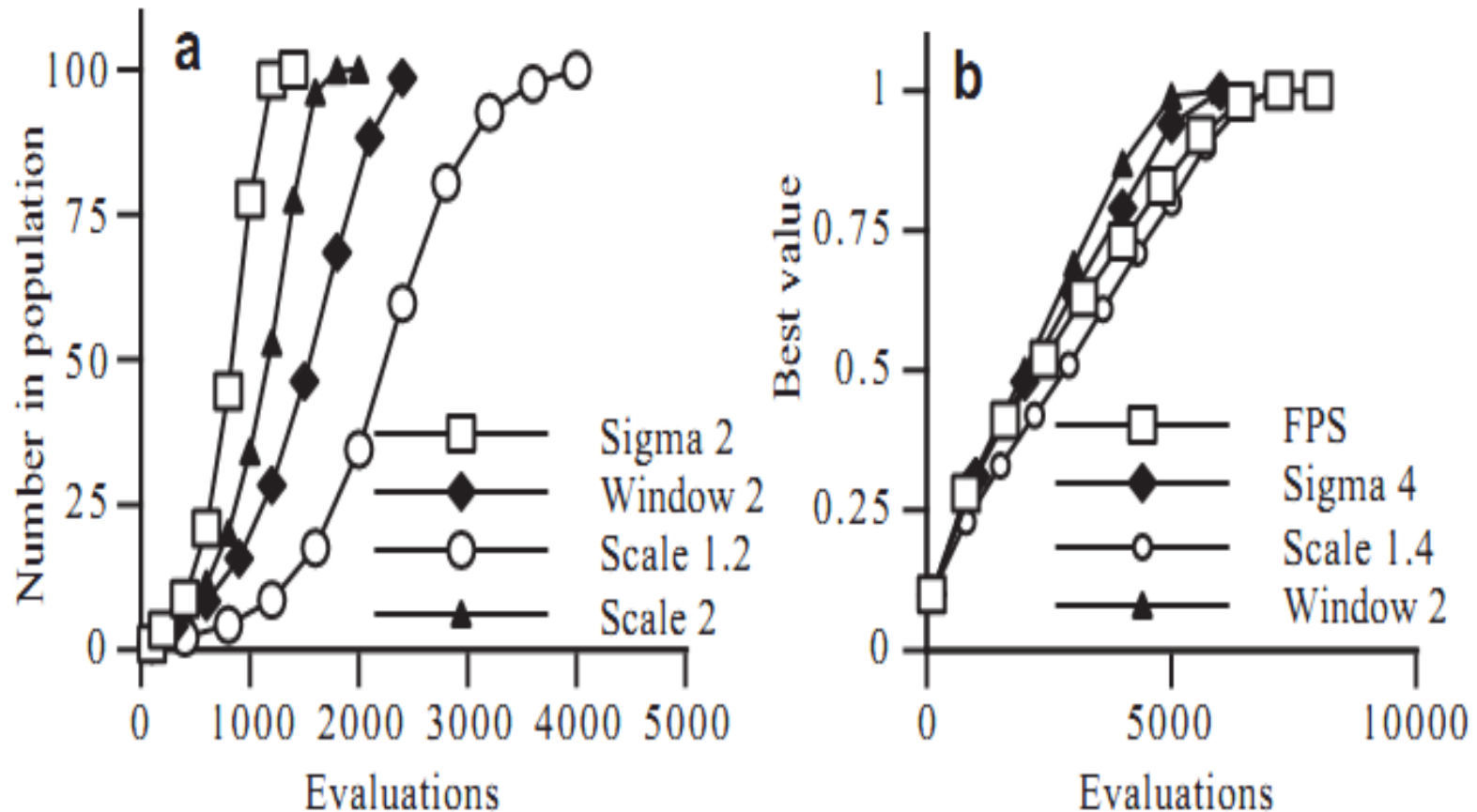

Fig: Takeover Time

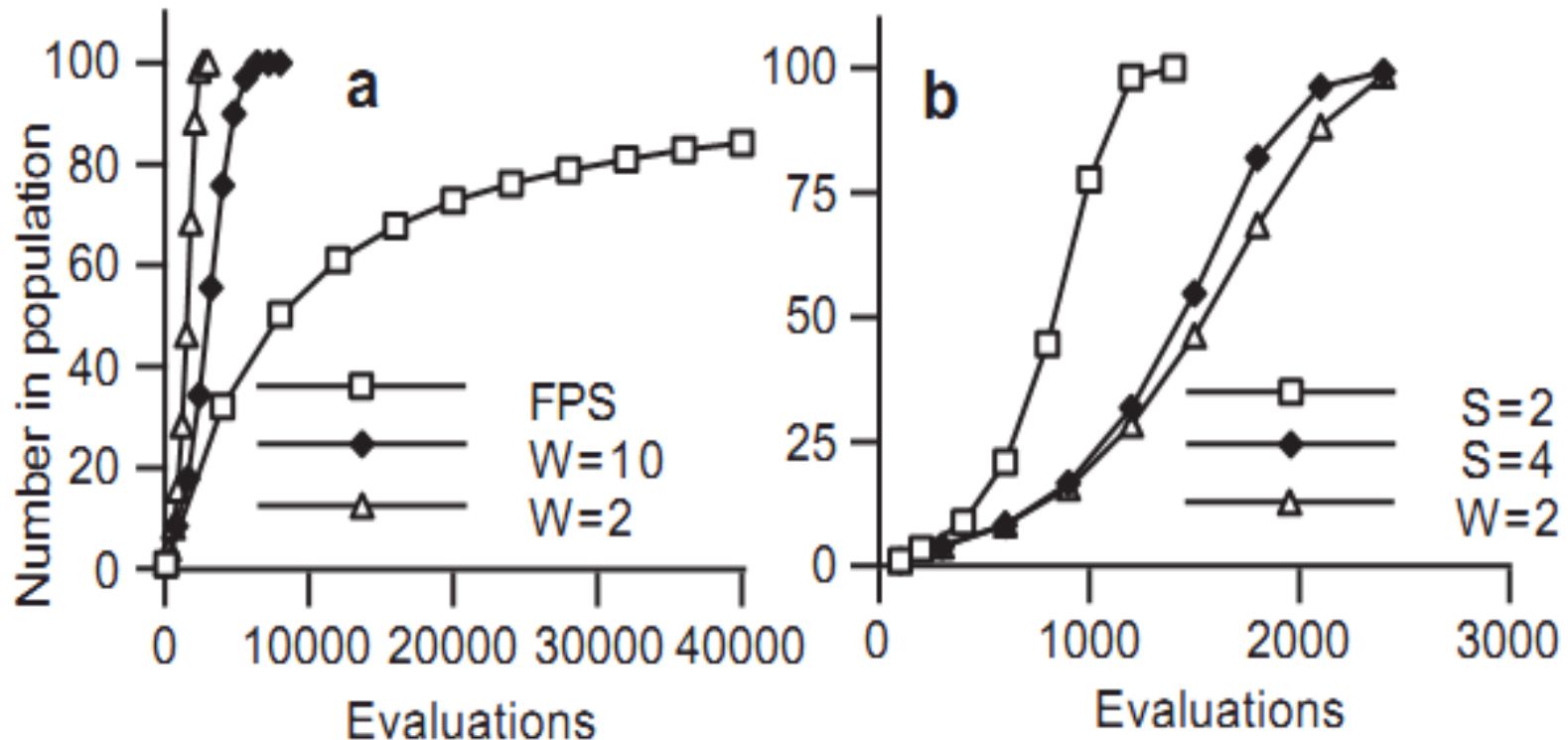# FPS, Sigma Scaling, Windowing

Fig: Takeover Time

Survivor Selection

- Role of survival selection is to distinguish among individuals based on their quality.
- Survivors replace individuals with low fitness value.
- Unlike parent selection, it's usually deterministic.
- Usually, offspring that survive replace weak or similar individuals of previous population.
- It is often called replacement.

Selection Generation

- Offspring of selected individuals in each generation becomes the next generation.
- Individuals are not retained between generations.
- Simple but very inefficient.
- Results in premature convergence, early loss of diversity.
- Once population is trapped in local optima, almost impossible to recover.

Elitist Selection

- Creating new population by crossover and mutation, we have a big chance, that we will loose the best chromosome.
- It first copies the best chromosome (or a few best chromosomes) to next generation.
- Rest is done in classical way.
- Elitism can very rapidly increase performance of GA, because it prevents losing the best found solution.

Fitness Proportionate Reproduction

- Reproduction in EA, an identical copy of the individual selected is carried over into the next generation: survival of the fittest.
- Fitness proportionate reproduction is the asexual reproduction of chromosomes selected stochastically from the population.
- Selection of an individual chromosome is based on a probability that is relative to that chromosomes relative fitness within its population.

Rank Selection

- Another variant of RWS
- Selection probability is proportional to relative fitness rather than absolute fitness.
- Doesn't make any difference whether the fittest candidate is ten times fitter than the next fittest or 0.001% fitter. Both cases the selection probabilities would be the same.
- All that matters is the ranking relative to other individuals.
- Selection probabilities are calculated and successive pairs of individuals are drawn using RWS.
- After drawing a pair, the string with the highest fitness is inserted into the mating pool, and another pair is drawn. The process continues until the mating pool is full.

Truncation Selection

- Truncation selection simply retains the fittest x% of the population.
- These fittest individuals are duplicated into the next generation, so that the population size is maintained.
- Less fit candidates are culled without being given the opportunity to evolve into something better.
- Often results in premature convergence.
- Advantage is rapid convergence.

Truncation Selection
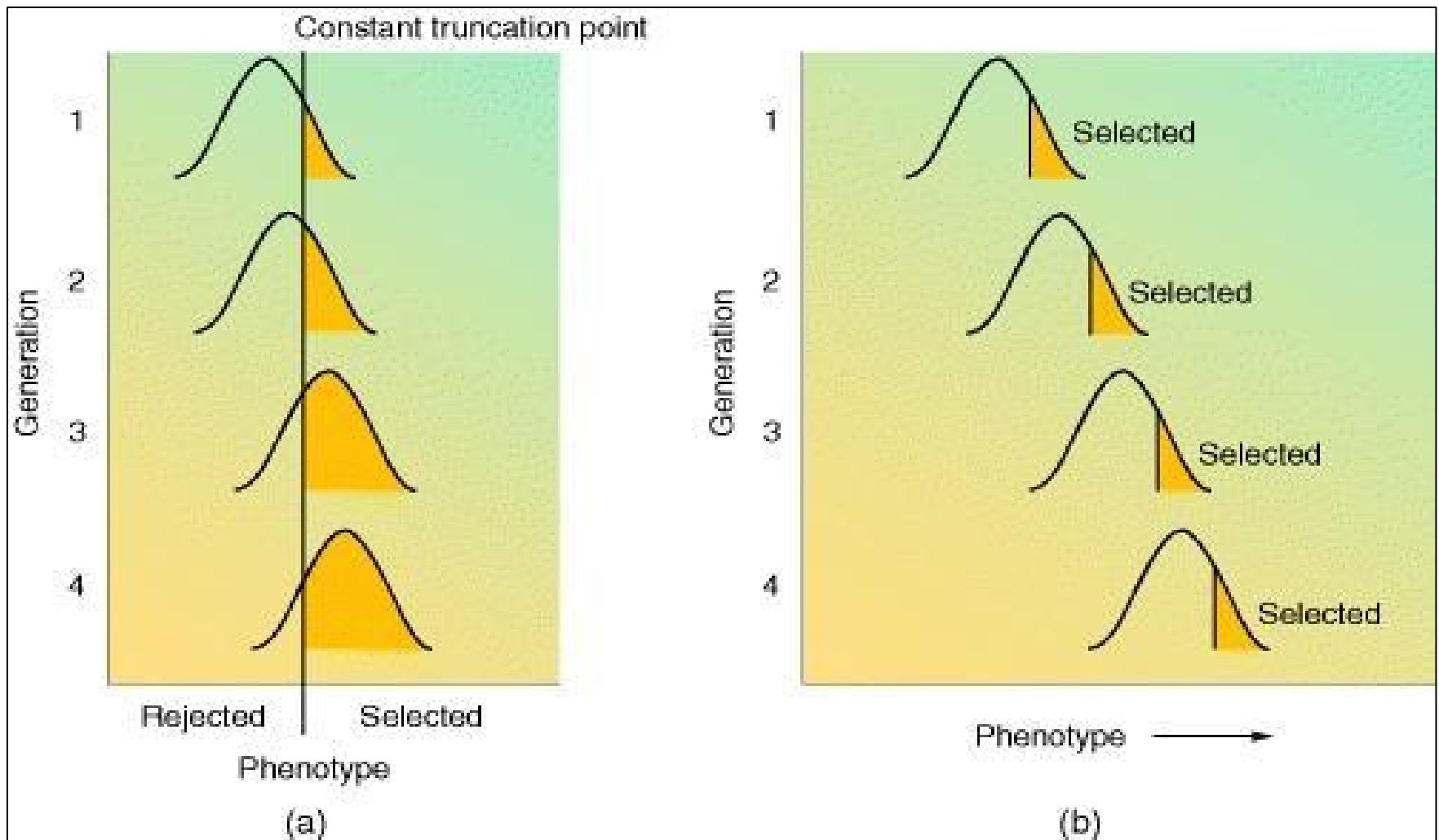
Pareto Ranking

Pareto Rankings

- **Belegundu's ranking:**
  1. All non dominated individuals are assigned rank 0 (or 1) and the dominated ones rank 1 (or 2).
- **Goldberg's ranking:**
  1. It assigns equal probability to all non dominated individuals.
  2. Assign rank 1 to the non dominated individuals and removing them from contention, then finding a new set of non dominated individuals, ranked 2, and so forth.
- **Fonseca and Fleming's ranking:**
  1. An individual's rank corresponds to the number of individuals in the current population by which it is dominated.
  2. Non dominated individuals are all assigned the same rank equal to zero, the dominated ones have values between 1 and k-1, where k is the population size.

Niching Methods

- Useful for multimodal problems.
- Prevent the convergence of the whole population to just one of the peaks.
- Penalizes individuals that are "close" to other individuals in the population.
- Two techniques are used:
  1. Fitness Sharing
  2. Crowding

Fitness Sharing

- Fitness is a shared resource of the population.
- Population is first divided into niches.
- Shared fitness of any individual is computed only with respect to the individuals that are in its niche.
- Sharing Radius defines the niche size.
- Individuals within this radius will be regarded as being similar to each other and thus need to share fitness
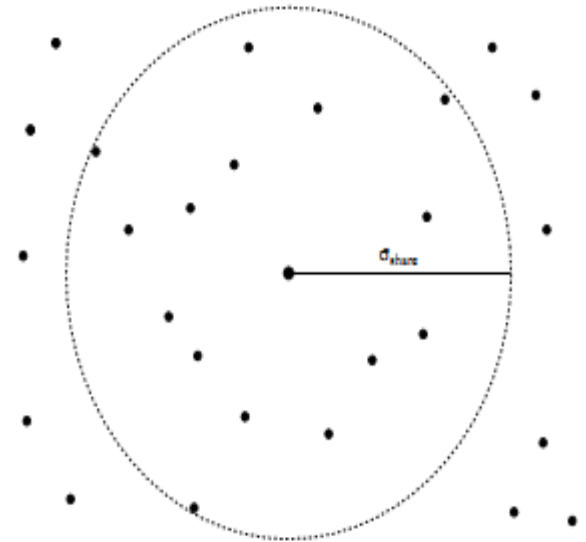
Fitness is shared between all the individuals in the circle

Fitness Sharing

Sharing Radius

- Fixed amount of fitness is allocated or all the individuals staying in the sharing radius.
- If too small, practically no effect on the process.
- If too large, several peaks will 'melt' individual peaks into one.

Implicit Fitness Sharing

- Idea comes from the immune system.
- Antibody which best matches an invading antigen receives the payoff for that antigen.
- A strategy that scores the best against a test case receives payoff.
- Implicit fitness sharing has often been used in machine learning.

Implicit Fitness Sharing: Algorithm

- For each test case i to be solved, do the following C times:
  - Select a sample of ¾ individuals from the population.
  - Find the individual in the sample that achieves the best performance for solving test case i.
  - This best individual (and this one only) receives the payoff.
  - Ties are broken evenly, i.e., payoff will be shared evenly among all best individuals if they have the same best performance.
- A larger C often leads to better sharing performance, but more time-consuming.

Implicit vs. Explicit Fitness Sharing

- Implicit fitness sharing and (explicit) fitness sharing both have the same theoretical basis.
- Implicit fitness sharing covers optima more comprehensively even when those optima have small basins of attraction, if population size is large enough for a species to form at each optimum.
- Explicit fitness sharing can find the optima with larger basins of attraction and ignore the ones with smaller bases, if population size is not large enough to cover all optima.

Crowding

- Similar individuals in natural population, often of the same species, compete against each other for limited resources.
- Dissimilar individuals tend to occupy different niches, they typically don't compete.
- Crowding uses individuals newly entering in a population to replace similar individuals.
- Random sample of *CF* (Crowding Factor) individuals, is taken from the population.
- Larger crowding factor indicates less tolerance for the similar solutions, smaller values indicate similar solutions are more welcome.

Crowding
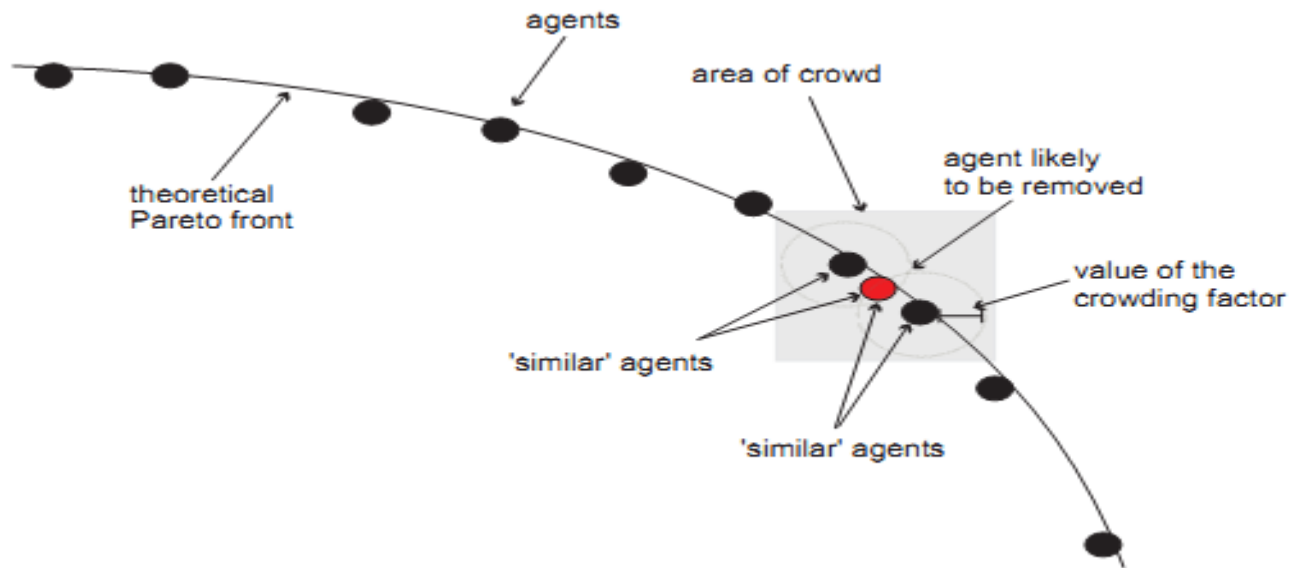
- The one most similar to element being inserted gets replaced.

Crowding

- New members of particular species replace older members of that species, not replacing members of other species.
- Crowding doesn't increase the diversity of population, rather it strives to maintain the pre-existing diversity.
- It's not directly influenced by fitness value.

Probabilistic Crowding

- Primarily a distance based niching method.
- Main difference is the use of a probabilistic rather than deterministic acceptance function.
- No longer do stronger individuals win over weaker individuals, they win proportionally according to their fitness, thus we get restorative pressure.
- Two core ideas of probabilistic crowding are
  1. To hold tournament between similar individuals
  2. To let tournaments be probabilistic

Deterministic Crowding

- There is a lack of analysis of convergence; as a result it is not entirely clear what deterministic crowding computes.
- Considering the internal working of the algorithm - the main problem appears to be there is no restoration pressure.
- Species of higher fitness tend to win over species of lower fitness.
- Niches may get lost even though they should not be according to their fitness value.

Niching vs. Speciation

- Two terms are used interchangeably although some people distinguish between them.
- Niching is concerned more with locating peaks (locating basins of attraction).
- Speciation is more focused on converging to the actual peaks.

Preselection

- It's a parental replacement technique.
- Like in crowding the goal of these schemes is to preserve diversity.
- Computing a difference measure for each new population element would be too expensive.
- One could estimate that a parent would be one of the members of the population closest to the new element if a child has higher fitness than the worse parent, it replaces that parent.
- Preselection algorithms are faster than crowding even with low *CF*.

Deterministic Sampling

- Average fitness of the population is calculated
- Fitness associated to each individual is divided by the average fitness, but only the integer part of this operation is stored.
- If the value is equal or higher than one, the individual is copied to the next generation.
- Remaining free places in the new population is fulfilled with individuals with the greatest fraction.

Genitor Algorithm

- It is less vulnerable to some of the biases that degrade performance in standard genetic algorithms.
- Differences between Genitor and simple EA
  1. Reproduction produces one offspring at a time. Two parents are selected for reproduction and produce an offspring that is immediately placed back into the population.
  2. Offspring do not replace parents, but rather the least t (or some relatively less t) member of the population. In Genitor, the worst individual in the population is replaced.
  3. Fitness is assigned according to rank rather than by fitness proportionate reproduction. Ranking helps to maintain a more constant selective pressure over the course of search.